Pension Decumulation Strategies

A State-of-the-Art Report

Thomas Bernhardt* and Catherine Donnelly†

https://risk-insight-lab.com/

Risk Insight Lab,
Department of Actuarial Mathematics and Statistics,
Heriot-Watt University, Edinburgh, Scotland EH14 4AS

30 April 2018

This work was funded by the Actuarial Research Centre (ARC) of The Institute and Faculty of Actuaries (IFoA), UK.

Please cite this report as: Bernhardt, T. and Donnelly, C. (2018). Pension decumulation strategies: A State-of-the-Art Report. Technical report number 1, Risk Insight Lab, Heriot-Watt University, UK. https://risk-insight-lab.com/outputs/

*T.Bernhardt@hw.ac.uk
†C.Donnelly@hw.ac.uk

Foreword

The authors are part of the ARC research project “Minimising Longevity and Investment Risk while Optimising Future Pension Plans”. The goal is to develop new pension product designs that keep the customers’ needs at the forefront. As a first step, this report was written to familiarize the project team with the existing knowledge on decumulation strategies. There is a UK-focus in the report.

The report is aimed at the reader who is familiar with the actuarial world, but are not necessarily familiar with stochastic control techniques. For this reason, technical descriptions of the mathematical methods used to solve many of the problems studied in the cited literature are not included.

Questions relating to property, taxes, regulations and solvency have been ignored in order to keep the focus on what is the state-of-the-art on the fundamental question of “how should I decumulate in retirement”?

The literature on decumulation strategies is vast and only a fraction of it is included. Omission of relevant material is unintentional.

Contents

1 Executive summary

This report is focused on income withdrawal and investment strategies that are proposed for decumulation in retirement. Popular strategies which are used in practice and specifications of how to decide the income and investment strategy are reviewed.

One of the most popular rules, the SWR or "4% rule", is very simple and easy to implement. The retiree invests in a constant-mix strategy and withdraws an inflation-indexed income for 30 years. There is intended to be a 90% chance of the strategy being successful.

However the constant-mix investment strategy is criticized by Scott et al. (2009) as being too expensive. In many possible future scenarios, there is an excess of assets at the end of 30 years. That means that the retiree has followed a wasteful strategy. They have paid for surpluses in excess of their stated goal. Moreover, there is a chance that the retiree could exhaust their funds before 30 years are up.

Scott et al. (2009) show that options or a different dynamic strategy could be employed to give either better outcomes for the same cost or the same outcomes for a lower cost.

As shown repeatedly in the literature, optimal investment strategies are dynamic strategies. They are usually not constant-mix strategies, although their exact dynamics depend on the optimization problem. In contrast, many drawdown products in the market are constant-mix strategies.

Optimal withdrawal rates, when allowed to be a decision variable, also vary over time. How much they vary will depend on the problem specification. However, if the problem set-up allows for risk aversion, then the withdrawal rate will vary less as the investor becomes more risk averse.

Annuities or modern tontines should be part of any decumulation strategy, particular at older ages. They are a cost-efficient strategy which either eliminate or significantly reduce the chance of running out of money. In general, the annuity or modern tontine should be incorporated into the retiree's assets when their mortality is the main driver of the decumulation strategy. This is the case when buying these products has a higher value than trading solely in the financial market.

There are many approaches to specifying the decumulation problem. Utility theory-based methods, such as Expected Utility Theory and Cumulative Prospect Theory, allow for investor risk aversion. This latter property can give intuitively appealing results. However, the problem specification is not easy to communicate.

Minimizing the difference between the actual consumption and a desired consumption via the minimization of a quadratic function has also been studied. Closed-form solutions are difficult to obtain.

Probabilistic methods, such as minimizing the probability of ruin, appear more objective. However, as they don't allow for investor risk aversion, the results can potentially be unattractive. The problems have to be carefully specified in order to obtain an agreeable solution.

Habit formation tries to match actual consumption behaviour closely. The investor seeks to keep their consumption above a weighted average of their past consumption. It may be difficult to justify such a formulation to the investor. Similarly, a problem specification (confusingly called drawdown) where the probability of the investor's wealth falling below some fraction of their running maximum wealth is minimized, may also be hard to explain as a decumulation goal.

2 Introduction

Decumulation is the process of converting pension savings into an income for retirement. The process involves investing the pension savings in the financial market, as it is expected to increase the value of the pension fund. We do not consider the accumulation phase. Our report is focused on withdrawal and investment strategies. We use the words withdrawal, income and consumption interchangeably. Although they do not have identical meanings, we do this for the sake of word variety over correctness.

Decumulation can be a one-off decision made at the point of retirement, for example to use all of one's wealth at retirement to buy a life annuity, or it could be a sequence of decisions. The choice of how to decumulate the funds is considered as a decision made by the customer.

For customers who manage their own assets, industry advice to consumers is focused on rule of thumbs that are easy to communicate (O'Connell et al., 2017). The basis is that the customer decides periodically how much to withdraw, for example during an annual review of their investments, and needs guidance on how much to withdraw over the next year.

A variety of decumulation products have emerged in the market following pension freedoms (Hyams et al., 2017), although the innovation for which the UK Government hoped has not yet materialized[1]. Most of these products are investment-only products and the investment strategy does not adjust to changing market conditions. They may come with broad guidance on how much to withdraw. In the absence of guarantees, there is generally no dynamic interaction between the remaining fund value, the desired future withdrawal amount and the investment strategy.

There are two main issues when deciding on a decumulation strategy.

- How much to withdraw periodically to live on. The exact lifespan of anyone is unknown, which makes it difficult to choose a sustainable level of spending.
- How to invest the remaining pension savings. There is no obvious choice how to invest the pension fund. Losing money is undesirable for a single person who has only one pension pot. However, in the current low interest environment and to maintain the purchasing power of savings, it seems that investment into the financial markets is unavoidable.

These two issues should be considered together. It is not enough to consider an investment strategy in isolation from the withdrawal strategy. An investor can obtain benchmark-beating returns. However, if they decumulate too quickly they can still end up with no savings before they die. They can withdraw too much to allow their savings to last for their lifetime.

Efficiently managing a retiree's portfolio for a lifetime of sustainable income is a difficult problem, when few people voluntarily buy life annuities (Mitchell et al., 2011). Fullmer (2009) states that wealth managers have widely varying solutions to this problem, even allowing for the individual investor's tolerance for longevity risk. Yet the same wealth managers have solutions based on an identical foundation, that of modern portfolio theory, for managing an investment portfolio for growth. Fullmer (2009) asks why we have generally accepted principles for answering the accumulation problem of investment growth, but not for the decumulation problem of a sustainable income.

Next we introduce some of the topics that are considered in the report.

[1] Pension freedoms (Paragraph 12), *Ninth Report of Session 2017-19*, Paragraph 12, UK House of Commons, Work and Pensions Committee, HC917, Published 5 April 2018, `https://publications.parliament.uk/pa/cm201719/cmselect/cmworpen/917/917.pdf`.

2.1 Investment strategies

An investment strategy is a rule of how to invest money in the financial market over time. For example, what proportion of their retirement savings should a pensioner invest in equities, bonds and property? How should the proportions vary over time? Should they be periodically re-balanced to maintain a constant proportion in each asset class? Should they vary according to the returns achieved, expectations of future market returns or the goals of the investor?

Dynamic investment strategies are common in the academic literature, when solving and analyzing drawdown problems. However, a wide range of dynamic strategies are rarely see in practice. A typical drawdown product that can be bought by a pensioner is based on a constant-mix portfolio. In a constant-mix portfolio, the fund is periodically re-balanced to specific proportions in certain assets.

2.2 Drawdown

Drawdown is a word that means withdrawing an income from pension savings which are invested by the retiree. The investments can be chosen directly by the retiree. Alternatively, an investment product which follows a specified investment strategy could be purchased. The underlying investments are likely to be mainly liquid assets like equities, bonds and cash.

2.3 Bequest

In drawdown, bequest refers to the savings of a person that remain after their death. The bequest is distributed to their beneficiaries.

Some people may adjust their consumption downwards in order to preserve assets for their heirs, while others pay scant regard to their bequest (Fullmer, 2008).

2.4 Financial market models

The most common continuous-time model applied in the literature, the Black-Scholes model, models the price B of a risk-free bond as

$$B(t) = B(0)e^{rt} \tag{1}$$

and the price S of a risky stock as

$$S(t) = S(0)e^{(\mu-\sigma^2/2)t+\sigma W(t)}, \tag{2}$$

in which the constant r is the risk-free rate of interest, the constant μ is the mean rate of return on the risky stock, the constant $\sigma > 0$ is the volatility of return on the risky stock and W is a standard Brownian motion. The initial prices $B(0)$ and $S(0)$ are strictly positive constants.

The Black-Scholes model can easily be extended to include more assets and it is usually a straightforward extension of a derived result to allow for more assets. It is also possible to allow the parameters r, μ and σ to become deterministic functions or non-trivial stochastic processes.

Due to its structure and properties, the Black-Scholes model is one of the simplest models under which closed-form solutions can be found. In stochastic control problems, it is extremely difficult to find closed-form solutions. Indeed, the existence of such solutions is an area of mathematical study.

Usually, a paper is written showing how to solve a particular problem within the Black-Scholes model. Depending on the aims of the authors, the focus of the paper may be more on the mathematics or it may be more on the broad guidance for investment or consumption strategies that the results imply. Or it may be a combination of both.

Later papers may extend the result to more complicated models. For example, it may be possible to get an analytic result when modelling the risk-free interest rate using the Vašíček model of interest rates.

The Black-Scholes model is often criticized for not capturing well enough some observed features of actual stockmarket returns. For example, the probability of extreme returns under the Black-Scholes model is too low, compared to the observed chance of very high or very low returns. It assumes a constant volatility of returns, which is not observed in practice.

The use of the Black-Scholes model facilitates closed-form solutions which are very useful. The sensitivity of the solution to the various parameters can generally be much better understood, and understood more quickly, when dealing with a closed-form solution. Moreover, the solution in a more complicated model has often the same broad structure.

Contrast this to trying to understand a solution that is the numerical output of a very complicated econometric model, without having the guidance of a solution derived in a simple model like the Black-Scholes model. There is no help as to what are the important parameters for the solution. Doing sensitivity testing becomes an extremely difficult task.

3 Decumulation strategies

In this section we review the proposed approaches to decumulation. Both pure decumulation and annuitization strategies are covered.

The first section outlines the main approaches that are widely used in practice in a drawdown environment. The most popular of these appears to be the "4% rule".

The subsequent sections review the literature on decumulation and annuitization. They are not mutually exclusive categorizations as many of the papers reviewed could go in several sections.

3.1 Drawdown rules derived from experience and judgment

Common advice on drawdown is often based on simple rules and assumptions. Calculations are often based on a fixed lifespan which is considered unlikely to be exceeded. For example, the assumption may be 30 years of retirement, based on a retirement age of 65 years. They are often tested by an analysis using historical data or simulation.

One drawdown strategy is to follow a 1/30 rule, which means to split the fund in equal portions and pay one portion annually out to the retiree. The fund itself is invested in the financial market in order to maintain each portion's value over time. Following the 1/30 rule means that, if you are alive after 30 years, you will have exhausted your funds. In addition, the amount that you withdraw – namely the appropriate invested portion – will fluctuate directly with investment returns.

Along similar lines, there is the concept of a safe withdrawal rate (SWR). Sometimes this is known as the "4% rule", when the SWR is 4%. The idea is to maintain the real value of the income withdrawn in the first year of retirement. The SWR rule equates to an inflation-increasing income that is expected to last for at least 30 years in 90% of all market scenarios. The probability may come from a model of future investment returns and inflation, or it may be a frequentist result derived from historical data over rolling 30 year time periods. An advantage of the SWR over the 1/30 rule is that the SWR pays a stable income to retiree, irrespective of market returns. A disadvantage is that the income may not last for 30 years.

The numerical value of the SWR is the first year's calculated income divided by the pension fund at retirement. Each subsequent year's income is the appropriately inflation-adjusted value of the first year's income. For example, if the SWR is 4% and the pension fund at retirement is

£100 000, then the retiree withdraws £4 000 as income in the first year. Supposing that inflation is 5% over the first year, then the second year's income is £4 200 (= £4 000 $\times$ 1.05).

Some studies (Bengen, 1994; Ameriks and Warshawsky, 2001) have shown that the SWR lies from 3% to 4.5%, basing their analysis on historical data. Another finds a lower rate of 2.5% would have applied for UK investors compared to a SWR of 3.5% for US investors (Blanchett et al., 2016).

Another drawdown rule is to make the income withdrawn depend on the investment returns. For example, the investor withdraws more income when last year's investment return has been positive, and less when it has been negative. This type of rule can be more prescriptive, such as specifying which assets to sell to fund withdrawals, or the maximum inflation adjustment to apply to the income (Bengen, 2001; Guyton, 2004; Guyton and Klinger, 2006). The main idea seems to be that the investor "cashes in" on part of their investment gain when returns have been good, particularly the gains on equity. They reduce the impact of crystallizing losses by lowering their income when returns have been negative. As an aside, these general principles are naturally followed by income strategies which withdraw a constant percentage of the fund's value.

The latter type of rules result in dynamic income strategies: they adjust the withdrawn income according to what investment returns and inflation have recently been experienced. They do not adjust the investment strategy. They typically use a constant-mix strategy, that is a constant proportion of the fund value is invested in specified assets.

The above rules are based on the intuition and the financial planning experience of those who state them. From their experience of giving individual financial advice, Guyton and Klinger (2006, page 56) state that most retirees have the following four ideals for income withdrawal from their investments:

- Maximize withdrawals (and withdrawal rates), especially early in retirement;
- Eliminate the possibility of running out of money;
- Avoid undesired changes to their income stream (that is, reductions or freezes);
- Maintain purchasing power.

Their analysis focuses on probabilities as a measure of success. For example, Guyton and Klinger (2006) defined the probability of success as the chance that at least $1 remained in the retiree's account after 40 years of withdrawals. The rule-makers of these prototype dynamic income strategies often use a financial market model to test their suppositions, typically a lognormal model for returns. They show that, by adjusting the income withdrawn over the retirement period in line with their rules, initial withdrawal rates between 4% and 6% may be as safe as the SWR (Bengen, 2001; Guyton and Klinger, 2006).

Pfau and Kitces (2014) find that increasing the equity exposure during retirement can give lower probabilities of the retiree's fund being exhausted and higher fund values at the fifth percentile, at the end of the chosen retirement timeframe. They consider a fixed 4% or 5% SWR in their analysis.

The explanation of Pfau and Kitces (2014) for increasing equity exposure is as follows. The fund value is generally largest in the first half of the retirement period, since fewer withdrawals have been made. During that time, a high investment return will give a large monetary return. If negative returns are subsequently experienced in the second half of retirement, then the overall monetary impact may not be too severe. The monetary gains made in the first half of retirement more than compensate for the negative returns in the second half, when the fund value is much smaller. The retiree is likely to be able to maintain their SWR drawdown rule. However, if the sequence of returns is reversed, with negative returns experienced in the first half of retirement, followed by positive returns in the second half, there are more serious consequences. The retiree may be ruined if the large monetary losses experienced in the first half are not made up in the second half of retirement. To compensate, the investor can try to gain from a rising equity market by increasing their equity exposure in the second half of retirement. Of course, if returns are also negative in the second half of retirement, then this strategy will worsen the financial situation of the retiree.

In contrast, Fullmer (2009) argues that under a dynamic investment strategy, the proportion of a retiree's fund value in equities usually increases as the investor gets older. Longevity risk tends to decrease with age. Decreasing longevity risk, i.e. a decreasing risk of outliving one's savings, means that the investor can afford to take on more investment risk. By increasing their exposure to equities, they will allow their longevity risk to increase back to the desired level. He emphasizes that if the portfolio value declines dramatically, causing longevity risk to increase, then the exposure to equities should decrease.

Scott et al. (2009) criticize the SWR approach as inefficient. It funds surpluses and over-spending. They outline how to produce the desired outcome for a lower cost. Their preferred approach is to maximize a customer's utility of consumption.

Milevsky and Huang (2011) also criticize the SWR approach. They study the impact on the optimal withdrawal policy of uncertainty about an individual's date of death. Allowing for longevity risk aversion, they find that only very high longevity risk aversion leads to stable spending rates. This might explain the 4% SWR as a consequence of some – but not all – investors' personal risk preferences.

3.2 Objective function approach

Rather than come up with rules based on expert judgment and practical experience, another approach is to model explicitly consumers' behaviour, goal or motivation through an objective function. For retirement problems, a typical objective function based on utility theory combines the following elements:

- A utility function u of consumption which represents preferences or risk aversion towards consumption;
- A discounting factor $\mathrm{e}^{-\rho t}$ modelling how much an investor prefers money sooner rather than later. The constant ρ is called the time-preference rate;
- A terminal time T which could be a random time of death or a fixed time, depending on the problem considered;
- An integral $\int_0^T \mathrm{d}t$ that sums the total utility derived from following a particular consumption and investment strategy up to time T in one possible future state of the world;
- A terminal condition v expressing the importance of the remaining funds at the terminal time. The condition could express the desire to purchase an annuity, leave a bequest or to ensure that as little as possible is left.
- An expectation operator $\mathbb{E}$ that averages the total utility and terminal condition over all possible future states of the world. This gives a numerical value which is called the value of the objective function.

The value of the objective function is controlled by decision variables. The decision variables, or controls, can include the consumption rate and the proportion of wealth invested in risky assets like equity. There may be restrictions on the controls, like no short-selling of stocks or bonds, and no more than 100% of the pension fund can be invested in a single asset.

In all situations, the controls have to be self-financing. This means that an admissible investment strategy uses only the initial wealth to support future consumption and the terminal wealth. For example, under the Black-Scholes model (1)-(2), the dynamics of the pension fund's wealth X obey

$$dX(t) = (rX(t) + (\mu - r)\alpha(t)X(t) - c(t))\,dt + \alpha(t)X(t)\sigma dW(t),$$

with $X(0)$ set equal to the initial value of the pension fund. The stochastic process α represents the proportion of the pension fund value invested in the risky stock and the stochastic process c represents the rate at which the pension fund's wealth is withdrawn as income by the retiree.

3.3 Maximizing the expected discounted utility of lifetime consumption

Based on economic theory, one way of determining the income and investment strategy of an individual over their retirement is by maximizing their future lifetime utility of consumption. In this section, we consider only utility functions based on Expected Utility Theory. An alternative approach of using Cumulative Prospect Theory is reviewed in Section 3.6.

Under Expected Utility Theory, this approach corresponds to the maximization problem

$$\sup_{(\alpha,c)} \mathbb{E}\left[\int_0^T e^{-\rho s} u\left(c(s)\right)\,\mathrm{d}s + e^{-\rho T} v\left(X(T)\right)\right],$$

where the maximization is taken over all admissible consumption rates c and equity proportions α.

There are standard choices for the utility function u and terminal condition function v. The choices reflect the investor's personal risk aversion and, more generally, the fundamental assumption that investors prefer more money to less. Some forms of the utility functions make the problem more amenable to a closed form solution.

Most prominently, there is the class of constant relative risk aversion (CRRA) utility functions which is often applied in the literature. For example, for a constant $\gamma \in (-\infty, 1) \setminus 0$, the power function

$$u_(c) = \frac{1}{\gamma} c^{\gamma}, \qquad \text{for } c > 0,$$

is a CRRA utility function. The choice of γ affects the risk aversion of the individual. The investor becomes more risk averse as the value of γ becomes more negative.

Richard (1975) seeks the optimal consumption, investment and life insurance strategy for an investor who can die and who wishes to maximize the above type of objective function. He considers a rich financial market with different equities and bonds. He shows that the optimal investment strategy only considers two mutual funds with fixed proportions of equities and bonds. In the literature, such a property is called Mutual Fund Separation. He shows that this happens regardless of an uncertain lifespan or the chosen utility function. Moreover, the mutual funds coincide with the ones originally obtained by Merton (1971) for a fixed time frame. In particular, there is no difference between the optimal strategy using the mutual funds and the optimal strategy obtained within the Black-Scholes framework.

Using the technical approach of dynamic programming, Richard (1975) finds explicit solutions for CRRA utility functions. The optimal investment and consumption are linear in the investment fund's value. In particular, the consumption varies in the same way as the mutual funds.

Milevsky and Huang (2011) try to illustrate what a lifecycle model says about optimal consumption rates. They contrast recommendations like the 4% (SWR) rule with the optimal consumption rates that result from an investor with no bequest motive. They find that very high aversion to longevity risk can explain the 4% rule as an optimal strategy.

They consider equities with constant returns in a simplified setting. Moreover, it is assumed that the pensioner obtains a fixed income during retirement through a state pension or annuity. Given this additional income, the optimal consumption depletes the fund's wealth before the assumed maximal age. More generally, the authors show that it is favourable to consume more today than tomorrow.

Habib et al. (2017) extend the setting of Milevsky and Huang (2011) to the case of stochastic returns and a random lifetime modelled by a Gompertz-Makeham law of mortality. The fund is assumed to have a constant proportion of wealth in equities. They allow for life annuities in the investment universe.

Similar to the result of Milevsky and Huang (2011), the expected depletion of the fund happens before the maximal age. The optimal consumption tends to decrease with age but is fluctuating in a similar way as the equities. Unable to find an explicit description for the optimal consumption, the authors consider first order approximations. Numerically, they show that the optimal consumption and its approximation are almost the same.

Andréasson et al. (2017) derive similar qualitative results in a different setting. As before, the pensioner obtains a fixed income during retirement through a state pension. Since it is an Australian study, they allow for a means-tested state pension.

In contrast to before, the proportion in equities can be dynamically adjusted. Wealth, family status, mortality risk and health status are all modelled. They maximize the expected value of utility associated with consumption, housing, and bequest for an individual. The control variables are the fraction of wealth invested in the risky stock, the fraction of wealth withdrawn as income to live on and the amount of wealth invested in housing.

In general, Andréasson et al. (2017) find that the proportion allocated to the risky stock increases as wealth decreases, due to the state pension acting as a guarantee. The retiree can afford to take more investment risk knowing that if returns are poor, they will qualify for the means-tested state pension. Thus the state pension works as a buffer against investment losses. They find that couples can take more investment risk due to their longer life expectancy compared to a single pensioner.

Steffensen (2011) allows relative risk aversion to vary over time. Applying a power utility function, he maximizes the expected value of the utility of consumption over a fixed time horizon. He determines expressions for the optimal investment and consumption strategies. The analysis of a situation of increasing relative risk aversion is complicated, and there are no easy messages arising from it.

3.4 Minimizing the distance from a target

Wealth can be conserved by adjusting withdrawal rates in line with investment outcomes. This was observed by Bengen (2001) and Guyton and Klinger (2006). It can be seen in the optimal solutions of maximizing the expected utility of lifetime consumption. However, this leads to a constantly varying income during retirement.

Addressing this issue, one approach is to introduce a target for the desired level of consumption. The problem becomes to minimize the distance of the actual consumption from the target consumption over the retiree's lifetime. A quadratic function is usually used to measure the distance from the target. This corresponds to the minimization problem

$$\min_{(\alpha,c)} \mathbb{E}\left[\int_0^T e^{-\rho s}\left(c(s)-b(s)\right)^2 \mathrm{d}s + \theta e^{-\rho T}\left(X(T)-F(T)\right)^2\right],$$

where the minimization is taken over all admissible equity proportions α and consumption rates c. The function b represents a desired level of intermediate consumption and the function F represents the desired target for the retiree's final wealth. The constant θ can be varied to reflect the importance of the consumption target compared to the wealth target.

Gerrard et al. (2006) consider a fixed time period after retirement until full annuitization takes place. The authors imagine a retiree who has insufficient funds to purchase a desired annuity but has the time to gain the necessary funds via investment returns. During the time period, the target is the continuous payments that would have been paid if full annuitization had taken place at retirement. Additionally, there is a terminal target that would be able to purchase the annuity of the desired income at the end of the time period. The targets work as a mechanism to track the performance of the fund.

The main conclusion of Gerrard et al. (2006) is that moderate risk-averse retirees should delay full annuitization as long as possible until they can purchase their desired annuity. The authors are even able to provide an explicit description for the optimal investment strategy under the simplification of a constant consumption during the fixed time period. However, in the general setting as well as in the simplified one, the optimal investment and consumption strategy is unable to prevent ruin fully. This is a surprising result since a fixed time period is considered.

Di Giacinto et al. (2014) apply the simplified setting of Gerrard et al. (2004) and impose restrictions on the fund. For example, the fund must stay above a minimal fund value, so that ruin never appears. The authors are able to find an explicit investment strategy when they remove the running target but keep the terminal one. This corresponds to the situation of a retiree who purchases an annuity at retirement and aims for a better one in the future. Here, the force of mortality is assumed to be constant, no short-selling can be imposed, but borrowing money to invest more than 100% in equities is allowed. It turns out that the wealth of the fund stays between the minimal value and the target. Excess above the target would yield a complete shift towards risk-free bonds.

He and Liang (2013) consider an equity-linked annuity fund with infinitely many participants. The fund value develops like a theoretical annuity reserve (Thiele's differential equation), with the investment strategy allowed to invest in a risky asset rather than being restricted to the risk-free asset. However, the income withdrawn is based on the fund value divided by age-appropriate life annuity cost calculated using the risk-free interest rate. This ensures that ruin will never occur, since the money withdrawn is always a fraction less than one of the remaining fund value.

The goal of the fund is to minimize the distance of the withdrawn income from a fixed target, until the highest attainable age of an individual. The decision variable is the investment strategy, namely how much to invest in the risky asset. To obtain an explicit solution for the optimal consumption and investment strategy, He and Liang (2013) assume De Moivre's Law for mortality.

They find that the proportion in equity decreases when the fund value approaches the amount needed to pay the target consumption in the future. In He and Liang (2016), these results have been extended to the case for a stochastic target representing the development of an average salary.

Vigna (2014) argues for the approach of minimizing a quadratic function about a target, for members of a defined contribution pension plan. It allows the investor's risk aversion to be indirectly determined, and the resultant portfolios are efficient from a mean-variance point of view. It is more intuitive for individuals.

Vigna (2014) proves that the mean-variance inefficiency decreases with the individual's risk aversion of the individual. It increases with the time horizon and the Sharpe ratio of the risky asset. She proves that the constant absolute risk aversion (CARA) and constant relative risk aversion (CRRA) optimal portfolios are not mean-variance efficient.

3.5 Habit formation

Constantinides (1990) introduces habit formation in order to solve the equity premium puzzle. For a stochastic process c representing the consumption stream and $\rho \geq 0$ the subjective rate of time preference, the lifetime utility associated with the consumption stream is

$$\mathbb{E} \int_0^\infty e^{-\rho t} u\left(c(t) - h(t)\right) \, \mathrm{d}t,$$

in which the utility function $u(x) := x^{1-\gamma}/(1-\gamma)$ for all $x > 0$, for a positive constant γ. For non-negative constants a, b, h_0, the habit level

$$h(t) := e^{-at} h_0 + b \int_0^t e^{a(s-t)} c(s) ds.$$

The habit level is a minimum or subsistence consumption rate with the choice of the power utility function. It is a exponentially-weighted function of the past consumption levels, with greater weight given to the most recent consumption.

Munk (2008) measures the expected power utility of consumption above the habit level, over a fixed time horizon. He determines the optimal investment strategy that maximizes the expected utility in concrete settings, such as in a stochastic interest rate model.

Due to habit formation, the optimal investment strategy is to ensure that future consumption can meet the habit level. Lower risk assets are more suited to this purpose than higher risk assets. Thus there is a different relative allocation to the asset classes, compared to the classical results with no habit persistence. However, Munk (2008) finds that the quantitative differences in allocation are small.

Bruhn and Steffensen (2013) allow for a different type of habit formation. In their setting, the rate of change of consumption is controlled, in addition to the amount invested in the risky asset. They aim to avoid the change in consumption being proportional to the change in the risky stock price. The latter effect occurs both with a habit formation approach (that is, the results of Munk (2008)) and in the classical setting of controlling directly the consumption amount.

They maximize a time-additive quadratic objective function over a fixed time horizon. Deviations of the change in consumption from a fixed percentage of the current consumption rate are punished via the objective function. They also allow for a bequest motive.

3.6 Cumulative Prospect Theory-based objective function

Cumulative Prospect Theory (CPT) is an alternative theory to Expected Utility Theory, based on experimental evidence. It is a descriptive theory (how people *do* behave), unlike Expected Utility Theory which is a normative theory (how people *should* behave).

Several behavioural elements are modelled by CPT (Tversky and Kahneman, 1992). Based on experimental evidence, CPT models investors who:

- Measure gains and losses relative to a reference point.
- Are more sensitive to losses than to gains. This is called *loss aversion*. A financial gain of £1000 does not compensate the investor for a financial loss of £1000.

- Are risk-seeking with respect to losses. They prefer to risk a large loss – with the possibility of avoiding the large loss – rather than lock into a certain loss.
- Are risk-averse with respect to gains. They prefer to lock into a certain gain rather than gamble on the possibility of a large gain.
- Over-weight events that have a low probability and under-weight events that have a moderate or high probability.

A *value function* is used to assign numerical values to gains and losses relative to the reference point. The value function reflects the investor's attitude to risk, in terms of relative gains or losses. It has the shape of an elongated letter S, being concave above the reference point and convex below it. An example of a value function is Tversky and Kahneman (1992, Equation (5)) which assigns to the number x the value

$$v(x) = \begin{cases} x^\alpha & \text{if } x \geq 0 \\ \lambda(-x)^\beta & \text{if } x < 0, \end{cases}$$

for constants α, β, λ.

The shape of the value function captures risk-seeking behaviour with respect to losses and risk-averse behaviour with respect to gains. It is steeper for losses than gains, to capture loss aversion.

Suppose a gamble is represented by a random variable X. The possible values of X give the gains or losses which may result for the investor, depending on which future state of the world we end up in. The next step is to calculate the average welfare derived from the gamble.

In CPT, the weights that an investor assigns to each gain or loss are used to weight the value function. Calculating the weighted value function applied to X gives the average welfare of the random variable X to the investor.

The weights do not usually have the same value as the probability associated with each gain or loss. They are called *decision weights*, to distinguish them from the probabilities. They are obtained by mapping the probabilities to the desired decision weight value, through a *(probability) weighting function*. Strictly speaking, it is the cumulative probabilities that are adjusted via the decision weights. Weighting functions have been proposed by Tversky and Kahneman (1992); Tversky and Fox (1995); Prelec (1998). For example, Tversky and Kahneman (1992, Equation (6)) is a pair of weighting functions

$$w^+(p) = \frac{p^\gamma}{(p^\gamma + (1-p)^\gamma)^{1/\gamma}} \quad \text{and} \quad w^-(p) = \frac{p^\delta}{(p^\delta + (1-p)^\delta)^{1/\delta}}.$$

The function w^+ is used when calculating decision weights relating to gains and the function w^- is used when calculating decision weights relating to losses. Numerical estimates of weighting function parameters have been done by Wu and Gonzalez (1996); Abdellaoui (2000); Booij et al. (2010).

It is technically much more difficult to derive analytical results for a investor who behaves in line with CPT compared to Expected Utility Theory. The CPT value function is partly convex and partly concave. The decision weights that replace probabilities add yet more complexity. The mathematical tools that can be readily applied to Expected Utility Theory-based problems – convex optimization, stochastic control and dynamic programming – fail to have a ready application for CPT-based problems (Jin and Zhou, 2008).

There are delicate but fundamental questions that have to be resolved for problems studied under CPT. Is a CPT-based problem well-posed? Does a solution exist and is it unique?

Jin and Zhou (2008) establish the answer to some of these questions in a continuous-time setting. In their own words,

- Establish, for the first time, a *bona fide* continuous-time behavioral portfolio selection model *à la* cumulative prospect theory, featuring very general S-shaped utility functions and probability distortions;
- Demonstrate that the well-posedness becomes an eminent issue for the behavioral model, and identify several ill-posed problems;
- Develop an approach, fundamentally different from the existing ones for the expected utility model, to overcome the immense difficulties arising from the analytically ill-behaved utility functions and probability distortions. Some of the sub-problems solvable by this approach, such as constrained maximization and minimization of Choquet integrals, are interesting, in both theory and applications, in their own rights; and
- Obtain fairly explicit solutions to a general model, and closed-form solutions for an important special case, based on which they are able to examine how the allocations to equity are influenced by behavioural criteria.

Rásonyi and Rodrigues (2013) provide necessary and sufficient conditions for well-posedness in a general, semimartingale market. Assuming suitable conditions, they prove the existence of a solution but they do not consider uniqueness. They state that "there seems to be no hope of uniqueness in this problem."

Berkelaar et al. (2004) focus on loss aversion, and use probabilities instead of decision weights. Thus they side-step the challenges thrown up by the decision weights. In this simplified version of CPT, they study the impact of different loss-averse utility specifications on portfolio choice. They use a continuous-time setting in a market where prices are driven by Brownian motions. They are able to derive some analytical expressions.

Their main finding is that loss aversion has a substantial effect on the optimal investment strategy. Under a kinked power utility (value) function, a partial portfolio insurance strategy is followed by the investor as they seek to keep their wealth above the reference point at the end of the investment time horizon.

Further, they find that a loss-averse investor will increase their initial equity holdings as their investment time horizon increases. For longer planning horizons, the initial strategy of an investor with a kinked power utility converges to the optimal portfolio of an investor with a smooth power utility function. With the luxury of time, an investor hopes to make good any losses and so invests more in equities at the outset. As the investment time horizon diminishes, a high-equity strategy can result in large losses. Then a portfolio insurance strategy is used to limit losses below the reference point.

Surprisingly, in CPT a discrete-time setting is as, or if not more, difficult than a continuous-time setting. He and Zhou (2011) discuss the issues that arise in the discrete-time setting. There are currently few analytical results in the literature, due to the challenges of applying CPT.

He and Zhou (2011) study when the maximization of the weighted value function is a well-posed problem in a discrete-time model. It is well-posed when the problem has a finite, optimal solution – where the solution is the amount invested in a risky stock – and the resultant weighted value function has a finite value. In their market, there is one risky stock and one risk-free bond.

They are able to prove conditions under which the problem is well-posed. The limiting behaviour of the ratio of the value function above the reference point to its value below the reference function is important. In simpler terms, the ratio of the unhappiness caused by a large loss to the happiness caused by a gain of the same magnitude determines what He and Zhou (2011) call the *large-loss aversion degree*. The latter's value indicates if the problem is well-posed.

He and Zhou (2011) go on to analyze various value functions, and derive explicit solutions for the optimal amount in the risky stock. Their results, albeit in a one-period model, are quite general; they do not specify a particular form for the weighting function.

There has been empirical success through simulations and numerical calculations for a CPT-based problem. The results are still relatively few and they are most relevant to the pre-retirement, accumulation phase of a pension plan. Nonetheless, the results are of interest to the decumulation phase.

Blake et al. (2013) apply the loss aversion aspect of CPT to the investment problem of the accumulation phase. Like Berkelaar et al. (2004), they do not use decision weights. Their setting is a combination of minimizing the distance from a target and maximizing expected utility from consumption. Here, only below the target, minimizing the distance from the target is favoured. Whereas, above the target, maximizing expected utility is preferred. This is modelled as discounted distance from the target under an S-shaped utility function over time.

The authors choose a target that coincides with the expected outcome of contributions and investments until retirement. The S-shaped value function leads to a V-shaped proportion in equities with respect to the fund's wealth. This shows that the investment choice is driven by the likelihood of a change in achieving the target. The higher the expected fund value, the better. Here 100% in equity is optimal.

Close to the target – when falling away from or below the target is likely – risk aversion is preferred. Moreover, the authors verify that the probability failing the target is lower than for maximizing consumption over lifetime. Due to the complexity of the problem, this is a numerical study only using a discrete model with log-normal returns and a stochastic salary (implying that future contributions are random).

Investor preferences for investment guarantees can be explained by CPT but not by Expected Utility Theory, as shown by Døskeland and Nordahl (2008); Dichtl and Drobetz (2011); Chen et al. (2015a).

In a related work, Dierkes et al. (2010) analyzes the interaction of CPT-preferences with investment time horizon. They find that a CPT investor prefers a pure bond strategy over short time horizons. Investors prefer right-skewed payoffs for most time horizons, under various value functions. Such payoffs can be generated by insurance strategies such as the purchase of a protective put or a constant proportion portfolio insurance (CPPI) strategy. They identify the probability weighting function as the main driver of the attractiveness of such insurance strategies, rather than the loss aversion. Interestingly, they find that a constant-mix strategy is inferior to buy-and-hold strategies for all preference parameters and time horizons.

CPT in its original form fails to capture investor predilections for guarantees on interim capital gains (Dichtl and Drobetz, 2011). This is not surprising as the CPT value function captures only payoffs at the maturity date. The downside terminal payoff is likely to attain higher values with an interim guarantee compared to a single, terminal guarantee. However, as guarantees cost money, the cost of securing interim gains will be to reduce the upside terminal payoff profile compared to that obtained with a single, terminal guarantee. It is easy to be convinced that under CPT, a terminal guarantee on the payoff may be more attractive than interim guarantees.

Ruß and Schelling (2017) propose an extension of CPT under which interim lock-ins of capital gains are more attractive than terminal guarantees. They call it Multi Cumulative Prospect Theory. They sum the weighted annual values of the CPT value function, where the weights are a time-dependent discount factor, to get the interim value sum. Then they find the weighted average of the interim value sum and the terminal value function, where the weights sum to one and are both positive.

Van Bilsen et al. (2017) examine consumption in a lifetime CPT model. Lifetime utility is the sum of discounted utility (value) of future consumption over a dynamic reference consumption level. The reference consumption changes over time and evolves in line with consumption based on habit formation (see Section 3.5). The amount by which actual consumption can fall below the reference consumption level is capped. The utility function is a power version of the CPT value function.

To maximize the expected lifetime utility, Van Bilsen et al. (2017) solve for the optimal dynamic investment and consumption strategies. Thus they use probabilities to weight the average lifetime utility rather than decision weights. They consider both a constant reference consumption and an endogenous one. For the latter, the reference consumption is a weighted average of past consumption with the last five years contributing at least 80% of the weight.

Van Bilsen et al. (2017) find that a loss-averse investor will protect current consumption after a decrease in their pension savings. They adopt a conservative investment strategy when their consumption is close to the reference consumption level, and invest more in the risky stock when consumption is above or below the reference level. In contrast, an investor with CRRA preferences consumes a constant percentage of their wealth and invests a constant percentage of their wealth in the risky stock. Furthermore, an endogenous reference consumption level leads to the investor adopting a lifestyle strategy, where less investment risk is taken as the investor ages.

Curatola (2017) investigates a different kind of habit formation in consumption. He assumes that the reference consumption level increases when current consumption exceeds the initial consumption, and decreases otherwise. In a typical habit formation setting, the reference level increases when the reference level is exceeded by current consumption, and not when the initial consumption is exceeded.

With the choice of reference consumption level in Curatola (2017), a CPT investor acts in a binary fashion. The optimal investment strategy has two components: a mean-variance portfolio and a "gambling" portfolio that is aggressively invested in risky stocks. How much weight is given to each component depends on the value of the state-price density process.

The investor will consume their wealth above the reference level and invest mostly in the mean-variance portfolio element when in "good times", i.e. when the state-price density process is above a certain threshold. The investment strategy means that the investor holds enough bonds to guarantee the consumption above the reference level.

When there are "bad times" (namely, the state-price density process is below the threshold), the investor reduces their consumption to a minimum subsistence level and mostly invests in the "gambling" portfolio element. This is to maximize the probability of the investor's consumption exceeding the reference level.

Curatola (2017) says that this consumption-investment behavior is consistent with the recent empirical evidence on investors' consumption-smoothing and trading behavior, especially during financial crises.

3.7 Minimizing the probability of ruin

Another way of specifying an objective function is by using probabilities. A number of papers have considered the problem of minimizing the probability of ruin in retirement. That is, the probability of the investor's fund value falling below a certain level by a defined time. The motivation is that communicating the chance of being ruined is easier than communicating the idea of lifetime utility to a customer.

Mathematically, if the continuous stochastic process X represents the investor's fund value and

w_a is the ruin level, then the first time at which the fund value hits the ruin level is

$$\tau_a := \inf\left\{t \geq 0 : X(t) = w_a\right\}.$$

To minimize the probability of ruin, the function

$$\psi\left(x,t\right) := \mathbb{P}\left[\tau_a < \infty \,\middle|\, X(t) = x\right]$$

is minimized over admissible investment strategies. To incorporate mortality, let τ_d represent the random time of death of an individual. Then the probability of ruin before death, starting at time t with wealth x, is

$$\psi\left(x,t\right) := \mathbb{P}\left[\tau_a < \tau_d \,\middle|\, X(t) = x, \tau_d > t\right].$$

Closed-form solutions are difficult to obtain for ruin problems and often simplified assumptions must be made in order to obtain analytical results. For example, a simple market consisting of only one risky asset following geometric Brownian motion and a bond accumulating at a constant rate is assumed. The retiree's force of mortality is deterministic. The papers on this topic tend to focus more on the mathematics rather than an extensive numerical study. This is justified. Due to the simplified assumptions, qualitative results are more important and these generally don't change much in a more complicated setting.

Setting more complicated assumptions would mean losing analytical results and resorting to numerical studies. Comonotonic approximations could be used in the numerical studies of problems that have more complicated assumptions (Dhaene et al., 2005; Vanduffel et al., 2005).

Milevsky et al. (1997) calculate the probability of outliving one's wealth, in a numerical study. They assume a constant-mix investment strategy and a constant, monthly, withdrawn amount. Investment returns follow geometric Brownian motion. For Canadians, they determine that the optimal equity allocation is in the range $70 - 100\%$. Another result is that for the younger retiree, the risk of them outliving their assets far outweighs the risk of losing money on higher investments. It is not until past age 75 years that the latter risk becomes more prominent.

Milevsky and Robinson (2000) calculate the probability of outliving one's wealth in a similar set-up to Milevsky et al. (1997). The focus is on approximations to the ruin probability. They find that the probability of lifetime ruin is lowest with a diversified portfolio, i.e. with at least 50% in equities, when the consumption rate is set equal to the life annuity payout.

Some authors have constructed a control problem to minimize the probability of being ruined. They determine the investment and income strategy that minimizes the chance of being ruined. This is in contrast to the approach of Milevsky et al. (1997); Milevsky and Robinson (2000) who calculate the ruin probability for a collection of constant-mix investment strategies. In particular, the latter approach does not consider dynamic strategies.

On minimizing the probability of ruin, an early work is Browne (1997), who allowed for consumption at a constant rate but he did not allow for mortality. He maximizes the probability of an investor being able to secure the consumption for perpetuity before their wealth reaches a bankruptcy level. However, he finds that no optimal investment strategy exists. It is impossible to secure the desired consumption through the formulation of this particular maximization problem.

For the investor who is able to secure the consumption for perpetuity, Browne (1997) minimizes the expected time until a fixed wealth goal is reached (where the wealth goal is above the amount required to secure the constant consumption in perpetuity). The latter problem is a growth problem and an optimal investment strategy exists. He also considers a linear consumption rate.

The first paper to determine analytically the optimal investment strategy that minimizes the probability of ruin is Young (2004). The optimal investment strategy exists in her setting because the investor's mortality is included in the objective function, in contrast to the setup of Browne (1997). Young (2004) defined a ruin event as one in which the investor's wealth drops below a constant, non-negative amount x. She considered two possible withdrawal strategies: a constant amount and a constant proportion of the investor's wealth. Assuming a constant force of mortality, she derived an expression for the optimal, dynamic investment strategy under both withdrawal strategies. The financial market consisted of a risky asset, the latter following geometric Brownian motion, and a risk-free asset whose price accumulates at a constant rate. The optimal investment amount in the risky asset is a positive, decreasing, linear function of wealth, that is independent of the ruin wealth level x.

A constant force of mortality is unrealistic. However, it is often essential in order to obtain a tractable problem. Moreover, as detailed in Cohen and Young (2016, Remark 2.1), by updating the results periodically, the results obtained under the assumption of a constant force of mortality can be adapted to a time-varying one. The idea is given in Moore and Young (2006), in which the probability of lifetime ruin is minimized under a (time-varying) Gompertz force of mortality. Moore and Young (2006) calculate the individual's future life expectancy based on their Gompertz force of mortality. Then, supposing that the future life expectancy is exponentially distributed with a constant hazard rate, they determine the implied constant force of mortality. It turns out that the problem's solution is well-approximated when using the implied constant force of mortality instead of the time-varying rates.

Young (2004) determines the sensitivity to various parameters of the optimal amount to invest in the risky asset. Her conclusions (Young, 2004, Sections 3.3, 4.3) are slightly paraphrased by us so that they can be read here independently of her paper.

- The higher the mortality rate, the less that individuals invest in the risky asset. They are less likely to outlive their wealth if the probability of dying is greater. That is, because an individual is more likely to die, he or she does not need to invest as aggressively in the risky asset to ensure the consumption rate. Alternatively, one might say that the individual has a shorter expected time horizon, so he or she invests more conservatively.

- As the volatility of the risky asset increases, the amount invested in the risky asset decreases. This makes sense because individuals do not want to risk their wealth dropping to the ruin level.

- Once the excess mean return on the risky asset exceeds a calculated constant, the amount invested in the risky asset decreases with the excess return. Individuals do not need to invest as much in the risky asset to maintain their consumption rate. The calculated constant's value depends on whether a constant consumption rate or a constant proportional rate of consumption is applied.

- As individuals consume more, the more they invest in the risky asset. They need to try to get higher returns to support the consumption rate.

Following the seminal paper of Young (2004), there has been a profusion of papers analyzing more complicated versions of the problem she solved. Bayraktar and Young (2007b) observe that the optimal strategy derived by Young (2004) results in borrowing money when the investor is close to be ruined, in order to invest in the risky asset. The investor ends up in a highly leveraged position in an attempt to escape ruin. To avoid this leveraged situation, Bayraktar and Young (2007b) impose borrowing and short-selling constraints on the investor. Using otherwise the same set-up as Young (2004), they solve the adjusted problem and find, unsurprisingly, that the constrained investor who is close to ruin will invest exactly all their wealth in the risky asset.

Bayraktar and Young (2009) try alternative objective functions in an effort to avoid the leveraging effect observed in the probability of lifetime ruin problem. They use a probability of ruin and the shortfall at death as the objective functions. In Bayraktar and Young (2009), the probability of ruin is the probability of: (i) the wealth at death being less than the ruin wealth level x; or (ii) the minimum wealth at death being strictly below the ruin wealth level x. Using M to represent the minimum wealth attained over an individual's lifetime, the shortfall at death is the expected value of $(x - M)_+$. Unfortunately, Bayraktar and Young (2009) find that leveraging is increased by considering these two objective functions.

Bayraktar and Young (2007a) determine a correspondence between two apparently different problem specifications. The first problem is to minimize the probability of ruin over an investor's lifetime, by controlling the investment strategy. It is assumed that the investor consumes a constant proportion of their wealth. The second problem is to maximize the expected utility of discounted lifetime consumption, by controlling both the investment strategy and the consumption. They characterize the utility function under which the solution to both problems is identical. It must exhibit hyperbolic absolute risk aversion. This class includes the power utility function. Both the investment strategy and consumption strategy are linear in wealth. They find that whether someone is ruined is a function of the minimum wealth.

Interestingly, Bayraktar and Young (2007a) find that the optimal investment strategy for minimizing the probability of hitting a particular ruin level is independent of the ruin level itself. More generally, for an investor who minimizes the expectation of a non-increasing function of minimum wealth, the optimal investment strategy is independent of the function.

Young and Zhang (2016) allow for ambiguity in the force of mortality function only. Thus the distribution of the force of mortality is unknown. They assume a constant consumption rate. Ruin is defined as the event that the investor's wealth becomes negative. They minimize the probability that the ruin event occurs before the investor's death. Similarly to Young (2004), they find that the longer someone is likely to live, the more time in which they have to be ruined. Thus their probability of lifetime ruin increases as the force of mortality decreases.

Cohen and Young (2016) consider an investor who seeks to minimize their expected lifetime poverty, with a financial penalty for ruin. Poverty is measured by a non-negative, non-decreasing function of wealth. They allow for both a constant consumption and a consumption that is proportional to wealth. The investor has an exponentially distributed future lifetime, with a constant force of mortality. Cohen and Young (2016) find that the optimal amount invested in the risky asset increases with the poverty level (i.e. the amount of wealth below which an individual is deemed to be in poverty) and decreases with the penalty for ruin. Furthermore, they show that when the investor's wealth is above the given poverty level, the individual invests as if they were minimizing their probability of lifetime ruin. This is an example of myopic investment; the individual appears to ignore the constraints. Drawing on several examples in the literature, the authors conjecture that *...myopic investment concerning constraints and opportunities is the rule, rather than the exception, in goal-seeking problems.*

Milevsky et al. (2006) add life annuities to the asset universe to extend the results of Young (2004); they consider the same problem as her, of minimizing the probability of ruin occurring before death. The investor wishes to avoid their wealth falling below zero before their death. The investor's time-varying force of mortality is not necessarily the same as that used to price life annuities. It is assumed that the investor withdraws a constant amount from their wealth. The investor does not have enough money to purchase their desired annuity.

Milevsky et al. (2006) determine the optimal investment strategy, including the optimal time to buy a life annuity. Unsurprisingly, they find that the probability of ruin is significantly reduced by the purchase of life annuities. In their analysis, the authors show that the amount invested in the risky asset increases with wealth if the investor's force of mortality exceeds the risk-free interest rate. It decreases with wealth if the investor's force of mortality is less than the risk-free interest rate. However, the optimal amount in the risky asset often exceeds the investor's wealth which is an unrealistic result for an individual investor.

Bayraktar and Young (2010) minimize the expected time that the investor's wealth stays below zero, i.e. the occupation time. They allow the investor to consume their wealth at a constant rate and assume that their future lifetime is exponentially distributed with a constant hazard rate. They compare the resultant optimal investment strategy to that for the problem of minimizing the probability of lifetime ruin. The occupation time strategy invests more in the risky asset when the wealth is negative. It invests the same amount when the wealth is positive. Thus this problem formulation does not avoid the leveraging issue of the optimal strategy for the problem of minimizing the probability of lifetime ruin. They note that the problems are identical when the investment in the risky asset is restricted to be zero in the occupation time problem when the wealth becomes negative.

3.8 Another meaning of drawdown

A suite of papers look at what is called drawdown, but this word has an entirely different meaning to ours. In the papers' nomenclature, a drawdown event occurs when the wealth falls below the drawdown level, which is a specified fraction of the running maximum of the wealth. More precisely, if the stochastic process X represents the investor's fund value then the running maximum wealth at time t is

$$M(t) := \max\left\{\sup_{0\le s\le t} X(s), m\right\},$$

in which $m > 0$ allows the investor to have a financial history before time 0. An event in which the investor's wealth hits $\alpha \in [0,1]$ times the running maximum wealth $M(t)$ is called *drawdown*. Its hitting time is

$$\tau_a := \inf\left\{t \ge 0 \,\middle|\, X(t) \le \alpha M(t)\right\}.$$

Again the objective function is a probability. By minimizing the probability of a drawdown event, they are attempting to minimize the volatility of the wealth, with the volatility being measured as the difference between the current wealth and a specified fraction of the running maximum of the wealth. The three papers cited below on this topic are focused on solving the latter (rather difficult) problem and not on the outcomes and implications for the investor.

Angoshtari and Young (2016a) minimize the probability of a drawdown event occurring before the investor is dead. They do this by controlling the investment strategy without portfolio constraints. A constant amount is withdrawn continuously from the investor's wealth. Again, we paraphrase slightly what the authors have written about their findings (Angoshtari and Young, 2016a, Section 6).

- In a specified region, the optimal investment strategy is identical to that for the problem of minimizing lifetime ruin (the problem solved by Young 2004). The specified region is when the investor's maximum running wealth is at or above the safe-level, i.e. the level of wealth that secures the consumption rate for perpetuity, and the investor's current wealth lies between the drawdown level and the safe-level.
- If the investor's initial maximum wealth is below the safe-level and also below a specified constant ($m^\star$), then the optimal investment strategy ensures that the running maximum

wealth remains unchanged from its current value. Angoshtari and Young (2016a) suggest that if the individual were to allow the running maximum wealth to increase, then the drawdown level would be too large given the consumption rate. The authors call the constant $m^\star$ the critical high-water mark (Angoshtari and Young, 2016a, Section 5). It is key in the behaviour of the optimal investment strategy.

- Otherwise, if the investor's initial maximum wealth is below the safe-level but above the critical high-water mark, then the optimal investment strategy allows the running maximum wealth to increase. The individual wishes to increase their wealth in order to fund their consumption.

Angoshtari and Young (2016b) allow for a wider range of possible withdrawal strategies (as long as it is a non-decreasing function of the wealth level) but solve a simpler problem. They minimize the probability of a drawdown event occurring in a finite time (rather than before the death of the investor).

Another contribution on this theme is Chen et al. (2015b), which allows for a cost proportional to the wealth to be deducted from the wealth. The cost can be re-interpreted as the amount withdrawn as income from the investor's wealth. Bäuerle and Bayraktar (2014) give a general result which can be applied to ruin problems. They consider controlled diffusions where the ratio of the drift to the square of the volatility is maximized. The resultant controlled process minimizes the drawdown probability.

3.9 Maximizing the bequest

Maximizing a bequest in isolation is related to optimization problems which seek to reach a specified goal. Bayraktar and Young (2016) maximize the probability of a constant bequest goal being met. The investor consumes their wealth at a constant rate and their future lifetime is exponentially distributed with a constant hazard rate.

When the investor's wealth is positive, Bayraktar and Young (2016) find that the optimal investment strategy is independent of the level of the bequest goal. Compared to the minimization of the probability of lifetime ruin problem – which has no bequest goal – it is optimal to invest more in the risky asset. The reason is the investor with a bequest goal wants to have more money for their bequest, so they take more investment risk in order to achieve this.

3.10 Annuitization

Buying a life annuity contract means that the retiree is guaranteed to receive a specified income stream for life. It eliminates longevity risk for the customer. Economic theory suggests that people with no bequest motive should buy life annuities Yaari (1965). Yet there are few countries where people buy them voluntarily Mitchell et al. (2011).

The result of Battocchio et al. (2007) indicates that the value from investing in equities is far bigger than the actuarially fair value of an annuity at retirement. They consider an insurance company that provides annuities. The annuities are priced in an actuarially fair manner, meaning that the price of each annuities is equal to the value of the annuity income. Moreover, the annuities can be purchased at any time before retirement with constant contributions made until retirement. Under the assumption that the expected return on equities is higher than the expected return on bonds, there is the possibility to create profit by trading in the market.

The company aims to maximize its expected final CRRA utility derived from the surplus of one customer. The surplus is defined as the difference between the fund value and the retrospective reserve from the one customer. Moreover, the company is willing to share a constant proportion of its profits with the customer. If the customer dies earlier than expected then the remaining retrospective reserve is positive and held in the bond. If the customer lives longer than expected the retrospective reserves in the bond are paid out accordingly. However, the surplus is unaffected by this re-balancing.

Given no further restrictions, Battocchio et al. (2007) find an explicit investment strategy. After re-balancing, the strategy never short-sells and yields a positive surplus for every single customer. Surprisingly, under reasonable market conditions, the same strategy without any re-balancing fails in less than 0.01% of the time.

Di Giacinto and Vigna (2012) ask if it is optimal to annuitize and, if not, then what is the cost to the retiree of being force to annuitize at retirement. The cost is defined as the loss of expected present value of consumption from retirement until death.

The authors find that immediate annuitization is sub-optimal in all cases with relative costs ranging from 6% to 40%. Depending on risk aversion, the individual's estimation of their own remaining lifetime and the Sharpe ratio of the risky asset, the retiree would prefer a different annuitization time.

On average, the size of the annuity increases as the optimal annuitization time increases. It is worth waiting in order to obtain a higher reward. In particular, the optimal annuitization time depends on personal factors such as risk aversion, the individual's estimation of their own remaining lifetime and the Sharpe ratio of the risky asset.

For high risk aversion, short future lifetime or low Sharpe ratio, the optimal annuitization time is a few years after retirement. For low risk aversion, long future lifetime or high Sharpe ratio, the optimal time is many years after retirement.

In an earlier paper, Gerrard et al. (2012) determine the optimal time to annuitize. The investor minimizes the distance from a target consumption and chooses a time of full annuitization. The time is chosen so that the terminal fund wealth minimizes the distance from the cost of a desired annuity.

In their numerical study, the investor has a constant force of mortality. Their simulations show that optimal annuitization occurs within 15 years from retirement in three-quarters of the simulated paths. On average, it should occur a few years after retirement.

Fullmer (2009) recommends the use of life annuities as a decumulation benchmark, as they represent the cost of guaranteeing the desired income stream for life. He argues for using longevity risk – the risk of a retiree outliving their savings – as a measure of risk in decumulation. The retiree's pension fund value should be above the cost of annuitizing their desired income stream. If it falls below the cost, then a life annuity should be purchased.

Milevsky and Young (2007) determine an optimal annuitization strategy in various types of markets. The first scenario considered by Milevsky and Young (2007) supposes that the investor must fully annuitize at some future time or not at all. In other words, they cannot buy annuities with only part of their wealth. This is reminiscent of the situation in the UK before pension freedom. In the highly simplified case of no volatility in the stock market returns, the retiree will fully annuitize their assets once the force of mortality of the retiree exceeds the equity risk premium. Allowing for volatility on investment returns, the investor wishes to maximize the expected discounted value of the CRRA utility of their lifetime consumption. The authors find that the retiree who follows their optimal strategy has a very good chance of consuming an amount that is higher by over 20% than the retiree who immediately annuitizes. The probability of consuming more depends upon the investment risk and mortality characteristics of the individual, both in isolation and relative to the pricing basis of the life annuities. From their analysis, they conclude that the optimal time to annuitize is after age 70.

Milevsky (2005) proposes the purchase of a deferred life annuity over an extended time period, as a means of making life annuities attractive. He suggests that the inflation-indexed income from the deferred annuities should start at age 80 or later. He dubs these *Advanced Life Deferred Annuities (ALDAs).* His motivation is that purchasing life annuities at the retirement date will never be appealing, due to the irreversible nature of the annuity contract in conjunction with the large amount of money required to buy a reasonable income. By staggering the decision through buying a series of ALDAs, the psychological barriers to annuitization may be reduced.

Gong and Webb (2010) do an extensive study of ALDAs. They compare the following decumulation strategies:

- At retirement, buying an inflation-indexed ALDA that is deferred until age 70, 75, 80, 85 or 90 years;
- At retirement, buying an inflation-indexed immediate life annuity;
- Postponing the purchase of an immediate life annuity until after the retirement date;
- Following a decumulation investment-only strategy.

In their words, they find that the decumulation strategy of buying a ALDA has several important advantages.

- ALDAs enable households to preserve liquidity, because their purchase cost is a fraction of the cost of immediate annuities, thus overcoming a potentially important psychological barrier to annuitization. They calculate that a household planning to smooth consumption through its retirement would need to allocate only 15% of its age 60 wealth to an ALDA with payments commencing at age 85, holding the remainder of its wealth in unannuitized form to finance consumption from age 60 to 85.
- Although a risk averse household facing an uncertain lifespan would prefer the full longevity insurance provided by an actuarially fair annuity to the partial longevity insurance provided by an actuarially fair ALDA, at plausible projected levels of actuarial unfairness, the household would prefer the ALDA to full annuitization. The intuition is simply that the household is suffering much less actuarial unfairness, but getting almost as much longevity insurance.
- An ALDA dominates an optimal decumulation of unannuitized wealth.
- ALDAs have the potential to improve and simplify the process of retirement wealth decumulation. They show that simple rules-of-thumb that perform almost as well as the optimal can be applied to the management of wealth decumulation over a period ending on the date that the ALDA income commences. In contrast, widely advocated rules for managing the decumulation of unannuitized wealth over an entire lifetime are highly sub-optimal.

Babbel and Merrill (2006) consider partial annuitization instead of full annuitization at retirement. Maximizing the expected discounted lifetime consumption with a CRRA utility function, the authors find an optimal annuitization level of 75% at retirement for an annuity price that is 10% above the actuarially fair price. However, this number is fairly sensitive to the assumptions. For example, the optimal annuitization level drops to 20% when default risk is included – that is, the risk that the company offering the annuity declares bankruptcy. It rises to 60% in the case that the customer can recover 10% of their annuity value when the insurance company goes bankrupt.

If the price of an annuity is 40% above its actuarially fair price, then a risk-averse retiree will still annuitize 50% of their wealth. In contrast, a risk-seeking person would optimally annuitize only 10% of their wealth, for the same annuity price mark-up.

Milevsky and Young (2007) determine an optimal annuitization strategy in various types of markets. In their second scenario, investors are free to annuitize their wealth as they want: using lump sums and continuous purchases. Their investor wishes to maximize the expected discounted value of the CRRA utility of their lifetime consumption. This is similar to the current situation in the UK. The optimal control turns out to be a barrier control. The retiree will annuitize just enough of their wealth in order to stay on one side of the barrier. Retirees should always hold some annuities. Moreover, they should increase their annuity holdings as they become wealthier. Additionally, if they have higher levels of investment risk aversion, if there is greater investment volatility or if the annuitant has lower chance of dying than assumed in the annuity pricing basis, then they should buy more annuity income. This is called *phased or gradual annuitization* in the literature.

Blake et al. (2014) and Horneff et al. (2008) also find that phased annuitization is optimal.

Horneff et al. (2008) have an extended version of the setting of Milevsky and Young (2007). They allow for pre-existing pension income and a bequest motive. They find that partial annuitisation is preferred when there is a bequest motive. Full annuitization is preferred when there is no bequest motive.

Blake et al. (2014) consider the entire pension saving and spending cycle. Future earnings are modelled stochastically and restrictions are imposed regarding short-selling or investing more than 100% of wealth.

The optimal strategy in the accumulation phase is to follow a stochastic lifestyling strategy. This results in equities between 20% to 50% of the portfolio at retirement. The investor should then use their bonds to buy a life annuity at the retirement date. During retirement, they should buy more annuity income. When the mortality premium exceeds the equity risk premium the fund is fully annuitised regardless of risk aversion. This is consistent with one of the results of Milevsky and Young (2007) in a simplified setting and happens at age 75. For high risk-aversion, even after full annuitization, the retiree restricts their consumption to buy even more annuity income.

Blake et al. (2003) do a numerical study of three different strategies in retirement. They calculate the expected discounted utility of lifetime consumption in order to quantify the attractiveness of each strategy. The three strategies are: (i) a purchased life annuity; (ii) an equity-linked annuity followed by a level, life annuity purchased at age 75; and (iii) equity-linked income-drawdown followed by a level, life annuity purchased at age 75. They allow for different amount of equity exposure in programmes (ii) and (iii). It turns out the level of equity investment is more important than which of (i), (ii) and (iii) is selected.

Variable Annuities with Guaranteed Minimum Income Benefits are investment contracts with the option to annuitize the fund value (Ledlie et al., 2008). Charupat and Milevsky (2002) assume that the Variable Annuity policyholder will annuitize fully at retirement. However, they must chose how much fixed annuity income to buy and how much of an equity-linked annuity to buy. In their set-up, this is inextricably linked to the choice of a constant-mix investment strategy. Choosing fixed annuity income means investing in the risk-free asset and choosing the equity-linked annuity income means investing in the risky asset. The choice of how to distribute wealth between the two annuity types is irreversible.

Charupat and Milevsky (2002) find that the optimal mix coincides with the optimal portfolio choice of Merton (1971), assuming that the investor maximizes their expected discounted lifetime consumption with a CRRA utility function.

Horneff et al. (2010) allow for variable payout annuities in their asset universe. The investor has to decide how much to consume and how much to invest in a variable payout annuity, a risky stock and a risk-free bond. The solution is a pair consisting of an optimal withdrawal strategy and an optimal investment strategy. The expectation of the sum of the investor's discounted utility of lifetime consumption and utility due to a bequest are maximized. They find that the retiree never fully annuitizes at retirement, even without any bequest motive. Full annuitization is deferred until age 80, at the latest. The optimal amount invested in the risky stock decreases as the retiree gets older.

3.11 Modern tontines

A recent stream of literature proposes and analyzes modern versions of tontines. Modern tontines allow people to get the benefits of longevity risk pooling without having to buy a life annuity contract. The advantage is that they mitigate the risk of outliving one's assets in the decumulation phase.

The purpose of modern tontines is to pay an income stream to the members. This is a very important point. They are not about taking short-term bets on whether death will occur or not.

Everyone in a modern tontine becomes the beneficiaries of each other. This is how longevity risk is pooled. It means that the risk of outliving one's assets is reduced and an income can be sustained for longer than when relying on investment returns only. The downside for those members with a bequest motive is that the assets of the dead member which are allocated to the tontine are not passed onto their estate.

A key feature of pure tontines is that there are no guarantees. This is unlike life annuity contracts which have implicit longevity and investment guarantees. An individual who is not overly risk averse can accept the volatility of the longevity credits, in the same way that they can accept the volatility of investment returns. The expected return on a pure tontine should be higher than the equivalent return on a life annuity contract as the former does not include guarantees. Of course, a risk averse investor may prefer to buy some guarantees around the income on the tontine. The very risk averse investor may prefer the life annuity, with its investment and longevity guarantees.

Actuarial fairness is often cited in the literature on tontines. Actuarial fairness means that the expected actuarial gain from pooling longevity risk is zero. The actuarial gain for an individual allows for that individual to get longevity credits while they are alive, and also to lose their own assets upon their death. It can be calculated over a fixed time period, for example a day, a month or a year, or over a lifetime.

Actuarial fairness is more important when the Law of Large Numbers does not hold for the group. This is most likely to occur when tontines are being set up or wound down. In that case, the fact that actuarial fairness holds by construction means that the administrator does not have to worry about adjusting payments to ensure fairness. The longevity credits are already automatically fairly calculated.

Modern tontines can be operated either explicitly or implicitly. In the explicit version there is a rule on how to share out the assets of the dead members. Members receive an explicit longevity credit into their individual account. In the implicit version, members receive an income. The income is adjusted as the mortality experience unfolds. The income ceases when a member dies. Real-life versions include TIAA (formerly called Teachers Insurance And Annuity Association-College Retirement Equities Fund or TIAA-CREF) annuities (Forman and Sabin, 2015).

The Group Self-Annuitization (GSA) plan, introduced by Piggott et al. (2005), is an implicit tontine. Members are given an income based on an annuity factor. The income is adjusted to take account of actual versus expected investment returns and mortality experience. The GSA plan is not actuarially fair, as shown by Donnelly (2015), which is only an issue if the Law of Large Numbers does not hold approximately in the plan.

Valdez et al. (2006) compare the demand for a GSA plan to private annuities, and also for adverse selection. Maximizing the utility of consumption for various utility functions, they demonstrate that there should be a demand for GSA plans even in the presence of adverse selection.

Further study on the GSA plan with regards to the numbers required to operate the plan and the impact of systematic longevity has been done by Qiao and Sherris (2013). They use a multi-factor stochastic mortality model.

Hanewald et al. (2013) assess the attractiveness of the GSA plan in relation to various contracts, such as nominal and inflation-indexed life annuities, deferred annuities and phased withdrawals. They include different levels of idiosyncratic and systematic longevity risk in their analysis. They find that, with loadings on guaranteed life annuity products, the GSA plan is more appealing than inflation-indexed annuities. Additionally allowing for a bequest motive, strategies that combine drawdown and GSA plans dominate portfolios with life annuities or deferred annuities.

Another implicit tontine structure is Milevsky and Salisbury (2016). They draw a distinction between (actuarial) fairness and equitability holding in a tontine scheme. A scheme is equitable if the present value of income less contributions is the same for all members. They construct a tontine that pays an income to the members and is equitable over the members' lifetimes.

In the explicit version of a tontine, individuals specify how much wealth they wish to allocate to the tontine. Each member can see the value of their own assets. It is not a notional allocation, but the true value of their own assets. Each individual gives up their allocated assets when they die.

A member's allocated assets are shared out among the group members as a longevity credit when the member dies. It is added to the individual account of each member. The payment is like an investment return, except it is an actuarial return. It depends on the actual mortality experience of the group.

Explicit tontines should be operated to pay an income to the members, as long as they survive. That is, they should be operated like the implicit version. The explicit calculation of the longevity credit is simply a building block, like investment returns. It enables individuals to have a higher income in retirement.

The Fair Tontine Annuity of Sabin (2010) is an explicit scheme that pays out a longevity credit in accordance with the principle of actuarial fairness. It works for heterogeneous groups, but not any heterogeneous group. There are restrictions on the group. For example, no-one should be too rich relative to the others. This can be done by placing restrictions on entry to the tontine.

Sabin (2010) proves necessary and sufficient conditions for the existence of the Fair Tontine Annuity. These are conditions on the wealth and mortality profile of the tontine's members. He develops an algorithm to calculate how the longevity credits are shared out in an actuarially fair way.

Bräutigam et al. (2017) propose an explicit tontine that pays out a longevity credit within a heterogeneous group, as a solution to a set of implicit equations. They do not show existence of the solution.

The Annuity Overlay Fund of Donnelly et al. (2014) is another explicit scheme that is actuarially fair. It works for any heterogeneous group, unlike that of Sabin (2010). It does this by making an adjustment payment to the estate of the recently deceased member.

Comparing the Annuity Overlay Fund to an equity-linked annuity, Donnelly et al. (2014) find that the insurer of the equity-linked annuity has to charge annual costs of less than 0.5% per annum to make the equity-linked annuity more attractive than the Annuity Overlay Fund when there are 600 participants in the latter. The cost is the insurer's charge for removing the volatility of the longevity credits.

Stamos (2008) proposed the Pooled Annuity Fund, an explicit tontine, to share longevity credits and pay out an income among a homogeneous group. The group members must follow the same investment strategy and have the same force of mortality function.

In a welfare analysis, Stamos (2008) shows that there should be a demand for his Pooled Annuity Fund, in the presence of level life annuities and access to the financial market. For a pool of 100 homogeneous members, he finds a 45% increase in welfare.

4 Conclusion

There are many ways of solving the problem of how much to withdraw as income and how to invest savings in retirement. There is no solution that is appropriate for everyone and neither is there a single solution for any individual.

It seems inevitable for the retiree to invest in the financial markets for two reasons. The first is to gain investment returns, and hence increase their retirement income. The second is to maintain the real value of their savings. Retirement is likely to last decades and it seems unwise to believe that we will never see periods of high inflation again. The literature that we have reviewed nearly always assumes that a pensioner invests in the financial markets during their retirement.

The spending profile of pensioners is sometimes mentioned as a ideal income target. A spending profile could reflect high spending at the start of retirement as a pensioner pays off their mortgage and other debts, does house renovations and takes a holiday or two. At older ages, health-care and care home costs increase, causing spending to increase.

However, pensioners can and do save from their retirement income (Urzi Brancati et al., 2015, Table 1). They do not have to receive an income that adjusts to match their anticipated spending pattern. They manage this themselves.

Moreover, people prefer money sooner rather than later. An example of this can be seen in the life annuity market. When given the choice, individuals nearly always buy a nominal annuity rather than an inflation-indexed annuity (Wells, 2014, Figure 6).

None of the income strategies reviewed in this report try to match an ideal, non-constant (whether in real or nominal terms) spending profile. This seems reasonable. However, due to a preference for receiving money sooner rather than later, strategies that are constructed to give significantly higher income at later ages are unlikely to be attractive.

Rules such as the SWR, often called the "4% rule" when the SWR is chosen to be 4%, are popular in practice. They give the retiree a good chance of an inflation-indexed income that will last for 30 years. They are simple to understand and apply.

The choice of the initial percentage to withdraw – e.g. 4% of the initial fund with subsequent withdrawals being the initial withdrawal increased with inflation – should be such that the fund is enough to pay for 30 years' worth of withdrawals in 90% of all market scenarios. This expression of risk is relatively simple to understand.

Moreover, from the viewpoint of a 65-year-old, living for 30 years until age 95 years seems unlikely. Fewer than 10% of respondents, who are older than age 45 years, in a biennial Society of Actuaries survey (Greenwald & Associates, 2018, Figures 171–172) thought that they would live past age 90. (Although 37% of pre-retirees and 45% of retirees either did not know or had not thought about their life expectancy. For those who did hazard a guess, the median answer was 85 years; Greenwald & Associates 2018, Page 20). More people under-estimate their own life expectancy than over-estimate it (Helman et al., 2016, page 8).

The SWR rule appears to be deliberately conservative. Its stated goal is to pay a constant real income for 30 years. Nothing more, nothing less. However, that ignores the impact of the investment strategy.

Neither a constant-mix nor a glide-path investment strategy is a good one for the SWR rule, as shown by the 1990 winner of the Nobel Prize in Economic Sciences and his co-authors in Scott et al. (2009). In a constant-mix strategy, the investment volatility is re-balanced regularly to a constant value, while it is decreased over time for a glide-path strategy. We summarize only part of their arguments here.

These two investment strategies are not efficient ones for the SWR rule. In many possible future scenarios, there is an excess of assets at the end of 30 years. That means that the retiree has followed a wasteful strategy. They have paid for surpluses in excess of their stated goal.

At the same time, there are many scenarios in which the investor fails to get the desired constant real income for 30 years. They have run out of money too soon. While the goal of the SWR rule may say something like *an inflation-increasing income that is expected to last for at least 30 years in 90% of all market scenarios*, it is unlikely that the retiree will be happy to be in one of the scenarios in which their savings don't last for 30 years.

Scott et al. (2009) find that the investor should be using either options or a dynamic investment strategy which is not a constant-mix or glide-path strategy when applying the SWR rule. They could meet their goal for a lower cost.

More broadly, Scott et al. (2009) argue that individual risk preferences should be taken into account, and suggest expected utility theory as a starting point.

This report has covered most of the decumulation approaches detailed in the literature. Many of them specify an optimization problem and determine an optimal solution, usually consisting of one or both of a consumption strategy and investment strategy.

As is shown over and over again in the literature, annuities or modern tontines should be part of any decumulation strategy. They are the only way of maintaining a cost-efficient strategy which either completely eliminates (in the case of insured annuities) or severely reduces the chance (in the case of modern tontines) of running out of money at older ages. In general, the annuity or modern tontine should be incorporated into the retiree's assets when their mortality is the main driver of the decumulation strategy. This is the case when buying these products has a higher value than trading solely in the financial market.

It is also shown repeatedly in the literature that optimal investment strategies are dynamic strategies. These are usually not constant-mix strategies, but their exact dynamics depend on the optimization problem. In contrast, many drawdown products in the market are constant-mix strategies.

Optimization problems for drawdown have been studied in the literature with mixed success for finding closed-form solutions. For example, there are explicit results (Gerrard et al., 2004, 2006) when drawdown is considered as transition from retirement to a fixed annuity phase. Here, the optimization criteria were chosen to fit in the known literature, resulting in undesirable features like negative withdrawal amounts. Other studies (Di Giacinto et al., 2010; Babbel and Merrill, 2006) considered partial annuitization and constraints on short-selling and borrowing, resulting in numerical solutions only.

Some optimization problems seek both an optimal consumption and investment strategy. The optimal consumption strategy is generally not constant, but varies with the fund value (although usually it will go up or down by a fraction of the percentage change in the fund value). For example, under a CRRA utility function, a more risk averse investor is expected to have a more stable consumption profile since they invest less in equities. Conversely, a more risk-seeking investor can expect a consumption that fluctuates in line with investment returns.

Other optimization problems assume that the investor draws income at a certain rate from their funds. Then the problem is to find an optimal investment strategy. This may be a particularly appealing formulation when the problem is expressed in probabilistic terms. These latter expressions don't take into account an investor's attitude to risk or a preference for more money over less. In consequence, probabilistic targets can give optimal strategies that look extreme. It is possible to get binary outcomes; either the target is met or the fund value is zero.

Habit formation tries to match actual consumption behaviour closely. It may be difficult to justify such a formulation to the investor. Similarly, the "other meaning of drawdown", where the probability of the investor's wealth falling below some fraction of their running maximum wealth is minimized, may also be hard to explain as a decumulation goal.

A bequest is often considered in drawdown optimization problems. It adds a goal to keep the fund value consistently high. It is contrary to the original goal of decumulating the fund for consumption. In particular, optimal strategies sacrifice even low income or take more risk in favour of the estate. As such strategies are based on idealized models, in reality these strategies might be far from optimal for both goals.

There are drawdown strategies that separate bequest from consumption (provided that they support a sufficient income). For example, the initial amount of the fund can be preserved, if the dividends of the investments are consumed only. Another example is to keep the fund invested in equities and withdraw the payments of an annuity only.

References

Abdellaoui, M. (2000). Parameter-free elicitation of utility and probability weighting functions. *Management Science*, 46:1497–1512.

Ameriks, J. and Warshawsky, M. (2001). Making retirement income last a lifetime. *Journal of Financial Planning*, pages 60–76. December 2001.

Andréasson, G., Shevchenko, P., and Novikov, A. (2017). Optimal consumption, investment and housing with means-tested public pension in retirement. *Insurance: Mathematics and Economics*, 75:32–47.

Angoshtari, B. Bayraktar, E. and Young, V. R. (2016a). Minimizing the probability of lifetime drawdown under constant consumption. *Insurance: Mathematics and Economics*, 69:210–223.

Angoshtari, B. Bayraktar, E. and Young, V. R. (2016b). Optimal investment to minimize the probability of drawdown. *Stochastics*, 88:946–958.

Babbel, D. and Merrill, C. (2006). Rational decumulation. Wharton Financial Institutions Center Working Paper, Number 06-14. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=917223.

Battocchio, P., Menoncin, F., and Scaillet, O. (2007). Optimal asset allocation for pension funds under mortality risk during the accumulation and decumulation phases. *Annals of Operations Research*, 152(1):141–165.

Bäuerle, N. and Bayraktar, E. (2014). A note on applications of stochastic ordering to control problems in insurance and finance. *Stochastics*, 86(2):330–340.

Bayraktar, E. and Young, V. (2007a). Correspondence between lifetime minimum wealth and utility of consumption. *Finance and Stochastics*, 11:213–236.

Bayraktar, E. and Young, V. (2007b). Minimizing the probability of lifetime ruin under borrowing constraints. *Insurance: Mathematics and Economics*, 41(1):196–221.

Bayraktar, E. and Young, V. (2009). Minimizing the lifetime shortfall or shortfall at death. *Insurance: Mathematics and Economics*, 44(3):447–458.

Bayraktar, E. and Young, V. (2010). Optimal investment strategy to minimize occupation time. *Annals of Operations Research*, 176(1):389–408.

Bayraktar, E. and Young, V. (2016). Optimally investing to reach a bequest goal. *Insurance: Mathematics and Economics*, 70:1–10.

Bengen, W. (1994). Determining withdrawal rates using historical data. *Journal of Financial Planning*, pages 171–180. October 1994.

Bengen, W. (2001). Conserving client portfolios during retirement, Part IV. *Journal of Financial Planning*, 5:110–119.

Berkelaar, A., Kouwenberg, R., and Post, T. (2004). Optimal portfolio choice under loss aversion. *Review of Economics and Statistics*, 4:973–987.

Blake, D., Cairns, A., and Dowd, K. (2003). Pensionmetrics 2: stochastic pension plan design during the distribution phase. *Insurance: Mathematics and Economics*, 33(1):29–47.

Blake, D., Wright, D., and Zhang, Y. (2013). Target-driven investing: Optimal investment strategies in defined contribution pension plans under loss aversion. *Journal of Economic Dynamics and Control*, 37(1):195–209.

Blake, D., Wright, D., and Zhang, Y. (2014). Age-dependent investing: Optimal funding and investment strategies in defined contribution pension plans when members are rational life cycle financial planners. *Journal of Economic Dynamics and Control*, 38:105–124.

Blanchett, D., Buffenoir, M., Kemp, D., and Watt, S. (2016). Safe withdrawal rates for retirees in the United Kingdom. Technical report, Morningstar. Morningstar Research, 1 May 2016.

Booij, A. S., van Praag, B. M. S., and van de Kuilen, G. (2010). A parametric analysis of prospect theory's functionals for the general population. *Theory and Decision*, 68:115–148.

Bräutigam, M., Guillén, M., and Nielsen, J. P. (2017). Facing up to longevity with old actuarial methods: A comparison of pooled funds and income tontines. *The Geneva Papers on Risk and Insurance - Issues and Practice*, 42:406–422.

Browne, S. (1997). Survival and growth with a liability: Optimal portfolio strategies in continuous time. *Mathematics of Operations Research*, 22:468–493.

Bruhn, K. and Steffensen, M. (2013). Optimal smooth consumption and annuity design. *Journal of Banking and Finance*, 37:2693–2701.

Charupat, N. and Milevsky, M. (2002). Optimal asset allocation in life annuities: a note. *Insurance: Mathematics and Economics*, 30:199–209.

Chen, A., Hentschel, F., and Klein, J. (2015a). A utility- and CPT-based comparison of life insurance contracts with guarantees. *Journal of Banking and Finance*, 61:327–339.

Chen, X., Landriault, D., Li, B., and Li, D. (2015b). On minimizing drawdown risks of lifetime investments. *Insurance: Mathematics and Economics*, 65:46–54.

Cohen, A. and Young, V. (2016). Minimizing lifetime poverty with a penalty for bankruptcy. *Insurance: Mathematics and Economics*, 69:156–167.

Constantinides, G. (1990). Habit formation: A resolution of the equity premium puzzle. *Journal of Political Economy*, 98:519–543.

Curatola, G. (2017). Optimal portfolio choice with loss aversion over consumption. *The Quarterly Review of Economics and Finance*, 66:345–358.

Dhaene, J., Vanduffel, S., Goovaerts, M., Kaas, R., and Vyncke, D. (2005). Comonotonic approximations for optimal portfolio selection problems. *Insurance: Mathematics and Economics*, 72:253–300.

Di Giacinto, M., Federico, S., Gozzi, F., and Vigna, E. (2010). Constrained portfolio choices in the decumulation phase of a pension plan. Number 155, October 2010. `http://www.carloalberto.org/assets/Fworking-papers/no.155.pdf`.

Di Giacinto, M., Federico, S., Gozzi, F., and Vigna, E. (2014). Income drawdown option with minimum guarantee. *European Journal of Operational Research*, 234(3):610–624.

Di Giacinto, M. and Vigna, E. (2012). On the sub-optimality cost of immediate annuitization in DC pension funds. *Central European Journal of Operations Research*, 20(3):497–527.

Dichtl, H. and Drobetz, W. (2011). Portfolio insurance and prospect theory investors: Popularity and optimal design of capital protected financial products. *Journal of Banking and Finance*, 35(7):1683–1697.

Dierkes, M., Erner, C., and Zeisberger, S. (2010). Investment horizon and the attractiveness of investment strategies: A behavioral approach. *Journal of Banking and Finance*, 34(5):1032–1046.

Donnelly, C. (2015). Actuarial fairness and solidarity in pooled annuity funds. *ASTIN Bulletin*, 45(1):49–74.

Donnelly, C., Guillén, M., and Nielsen, J. (2014). Bringing cost transparency to the life annuity market. *Insurance: Mathematics and Economics*, 56:14–27.

Døskeland, T. and Nordahl, H. (2008). Optimal pension insurance design. *Journal of Banking and Finance*, 32(3):382–392.

Forman, J. and Sabin, M. (2015). Tontine pensions. *University of Pennsylvania Law Review*, 163(3):755–831.

Fullmer, R. (2008). The fundamental differences in accumulation and decumulation. *Journal of Investment Consulting*, 9(1):36–40.

Fullmer, R. (2009). A framework for portfolio decumulation. *Journal of Investment Consulting*, 10(1):63–71.

Gerrard, R., Haberman, S., and Vigna, E. (2004). Optimal investment choices post-retirement in a defined contribution pension scheme. *Insurance: Mathematics and Economics*, 35(2):321–342.

Gerrard, R., Haberman, S., and Vigna, E. (2006). The management of decumulation risks in a defined contribution pension plan. *North American Actuarial Journal*, 10(1):84–110.

Gerrard, R., Højgaard, B., and Vigna, E. (2012). Choosing the optimal annuitization time post-retirement. *Quantitative Finance*, 12(7):1143–1159.

Gong, G. and Webb, A. (2010). Evaluating the advanced life deferred annuity – an annuity people might actually buy. *Insurance: Mathematics and Economics*, 46(1):210–221.

Greenwald & Associates (2018). 2017 Risks and process of retirement survey. 2017 Retirement Risk Survey. Sponsored by the Society of Actuaries' Committee on Post Retirement Needs and Risks. `https://www.soa.org/research-reports/2018/retirement-risk-survey/`.

Guyton, J. (2004). Decision rules and portfolio management for retirees: is the 'safe' initial withdrawal rate too safe? *Journal of Financial Planning*, pages 54–62. October 2004.

Guyton, J. and Klinger, W. (2006). Decision rules and maximum initial withdrawal rates. *Journal of Financial Planning*, pages 50–58. March 2006.

Habib, F., Huang, H., and Milevsky, M. (2017). Approximate solutions to retirement spending problems and the optimality of ruin. `http://dx.doi.org/10.2139/ssrn.2944125`.

Hanewald, K., Piggott, J., and Sherris, M. (2013). Individual post-retirement longevity risk management under systematic mortality risk. *Insurance: Mathematics and Economics*, 52(1):87–97.

He, L. and Liang, Z. (2013). Optimal dynamic asset allocation strategy for ELA scheme of DC pension plan during the distribution phase. *Insurance: Mathematics and Economics*, 52(2):404–410.

He, L. and Liang, Z. (2016). Optimal asset allocation and benefit outgo policies of the DC pension plan during the decumulation phase. *3rd International Conference on Economics and Management (ICEM2016)*.

He, X. and Zhou, X. (2011). Portfolio choice under cumulative prospect theory: An analytical treatment. *Management Science*, 57(2):315–331.

Helman, R., Rappaport, A., Levering, C., and Toale, T. a. (2016). Key findings and issues – living longer and impact on planning. 2015 Retirement Risk Survey. Sponsored by the Society of Actuaries' Committee on Post Retirement Needs and Risks. `https://www.soa.org/Files/Research/Projects/research-2016-living-longer-impact.pdf`.

Horneff, W., Maurer, R., Mitchell, O., and Stamos, M. (2010). Variable payout annuities and dynamic portfolio choice in retirement. *Journal of Pension Economics and Finance*, 9:163–183.

Horneff, W., Maurer, R., and Stamos, M. (2008). Optimal gradual annuitization: Quantifying the costs of switching to annuities. *Journal of Risk and Insurance*, 75(4):1019–1038.

Hyams, S., Woodruff, M., Warren, G., Smith, A., Atherton, J., Pickett, E., and Willetts, P. (2017). Pension decumulation market research: Review of UK and worldwide markets. Defined Contributions Participation, Accumulation and Decumulation Working Party, September 2017. `https:\\www.actuaries.org.uk\documents\pension-decumulation-market-research-september-2017`.

Jin, H. and Zhou, X. (2008). Behavioural portfolio selection in continuous time. *Mathematical Finance*, 18(3):385–426.

Ledlie, M., Corry, D., Finkelstein, G., Ritchie, A., Su, K., and Wilson, D. (2008). Variable annuities. *British Actuarial Journal*, 14(2):327–389.

Merton, R. (1971). Optimal consumption and portfolio rules in a continuous-time model. *Journal of Economic Theory*, 3(4):373–413.

Milevsky, M. (2005). Real longevity insurance with a deductible: Introduction to advanced-life delayed annuities (ALDA). *North American Actuarial Journal*, 9(4):109–122.

Milevsky, M., Ho, K., and Robinson, C. (1997). Asset allocation via the conditional first exit time or how to avoid outliving your money. *Review of Quantitative Finance and Accounting*, 9:53–70.

Milevsky, M. and Huang, H. (2011). Spending retirement on Planet Vulcan: The impact of longevity risk aversion on optimal withdrawal rates. *Financial Analysts Journal*, 67:45–58.

Milevsky, M., Moore, K., and Young, V. (2006). Asset allocation and annuity-purchase strategies to minimize the probability of financial ruin. *Mathematical Finance*, 16:647–671.

Milevsky, M. and Robinson, C. (2000). Self-annuitization and ruin in retirement, with discussion. *North American Actuarial Journal*, 4:112–129.

Milevsky, M. and Young, V. (2007). Annuitization and asset allocation. *Journal of Economic Dynamics and Control*, 31(9):3138–3177.

Milevsky, M. A. and Salisbury, T. S. (2016). Equitable retirement income tontines: Mixing cohorts without discriminating. *ASTIN Bulletin*, 46(3):571–604.

Mitchell, O., Piggott, P., and Takayama, N., editors (2011). *Securing Lifelong Retirement Income: Global Annuity Markets and Policy*. Oxford University Press.

Moore, K. and Young, V. (2006). Optimal and simple, nearly-optimal rules for minimizing the probability of financial ruin in retirement. *North American Actuarial Journal*, 10:145–161.

Munk, C. (2008). Portfolio and consumption choice with stochastic investment opportunities and habit formation in preference. *Journal of Economic Dynamics and Control*, 32(11):3560–3589.

O'Connell, A., Edgar, C., Ormrod, C., Mussett, D., Benbow, J., and Prisk, K. (2017). Decumulation options in the New Zealand market: How rules of thumb can help. Technical report, New Zealand Society of Actuaries.

Pfau, W. and Kitces, M. (2014). Reducing retirement risk with a rising equity glide path. *Journal of Financial Planning*, 27:38–45.

Piggott, P., Valdez, E., and Detzel, B. (2005). The simple analytics of a pooled annuity fund. *The Journal of Risk and Insurance*, 72(3):497–520.

Prelec, D. (1998). The probability weighting function. *Econometrica*, 66(3):497–527.

Qiao, C. and Sherris, M. (2013). Managing systematic mortality risk with group self-pooling and annuitization schemes. *Journal of Risk and Insurance*, 80(4):949–974.

Rásonyi, M. and Rodrigues, A. M. (2013). Optimal portfolio choice for a behavioural investor in continuous-time markets. *Annals of Finance*, 9(2):291–318.

Richard, S. (1975). Optimal consumption, portfolio and life insurance rules for an uncertain lived individual in a continuous time model. *Journal of Financial Economics*, 2:187–203.

Ruß, J. and Schelling, S. (2017). Multi cumulative prospect theory and the demand for cliquet-style guarantees. *Journal of Risk and Insurance*. In press.

Sabin, M. (2010). Fair tontine annuity. `http://ssrn.com/abstract=1579932`.

Scott, J., Sharpe, W., and Watson, J. (2009). The 4% rule – at what price? *Journal of Investment Management*, 7:31–48.

Stamos, M. Z. (2008). Optimal consumption and portfolio choice for pooled annuity funds. *Insurance: Mathematics and Economics*, 43:56–68.

Steffensen, M. (2011). Optimal consumption and investment under time-varying relative risk aversion. *Journal of Economic Dynamics and Control*, 35(5):659–667.

Tversky, A. and Fox, C. (1995). Weighing risk and uncertainty. *Psychological Review*, 102:269–283.

Tversky, A. and Kahneman, D. (1992). Advances in prospect theory: Cumulative representation of uncertainty. *Journal of Risk and Uncertainty*, 5:297–323.

Urzi Brancati, C., Brian Beach, B., and Franklin, B.and Jones, M. (2015). Understanding retirement journeys: Expectations vs reality. `http://www.ilcuk.org.uk/index.php/publications/publication_details/understanding_retirement_journeys_expectations_vs_reality`.

Valdez, E. A., Piggott, J., and Wang, L. (2006). Demand and adverse selection in a pooled annuity fund. *Insurance: Mathematics and Economics*, 39(2):251–266.

Van Bilsen, S., Laeven, R., and Nijman, T. (2017). Consumption and portfolio choice under loss aversion and endogenous updating of the reference level. Available at SSRN. `http://dx.doi.org/10.2139/ssrn.2530259`.

Vanduffel, S., Dhaene, J., and Goovaerts, M. (2005). On the evaluation of 'saving-consumption' plans. *Journal of Pension Economics and Finance*, 4:17–30.

Vigna, E. (2014). On efficiency of mean-variance based portfolio selection in defined contribution pension schemes. *Quantitative Finance*, 14(2):237–258.

Wells, J. (2014). Pension annuities: a review of consumer behaviour. Report prepared for the Financial Conduct Authority. `https://www.fca.org.uk/publication/research/annuities-consumer-behaviour-review.pdf`.

Wu, G. and Gonzalez, R. (1996). Curvature of the probability weighting function. *Management Science*, 42:1676–1690.

Yaari, M. (1965). Uncertain lifetime, life insurance, and the theory of the consumer. *Review of Economic Studies*, 32:137–150.

Young, V. (2004). Optimal investment strategy to minimize the probability of lifetime ruin. *North American Actuarial Journal*, 8:106–126.

Young, V. and Zhang, Y. (2016). Lifetime ruin under ambiguous hazard rate. *Insurance: Mathematics and Economics*, 70:125–134.