Optimal Performance of a Tontine Overlay Subject to Withdrawal Constraints

Peter A. Forsyth[a] Kenneth R. Vetzal[b] Graham Westmacott[c]

October 20, 2023

Abstract

We consider the holder of an individual tontine retirement account, with maximum and minimum withdrawal amounts (per year) specified. The tontine account holder initiates the account at age 65 and earns mortality credits while alive, but forfeits all wealth in the account upon death. The holder wants to maximize total withdrawals and minimize expected shortfall at the end of the retirement horizon of 30 years (i.e. it is assumed that the holder survives to age 95). The holder controls the amount withdrawn each year and the fraction of the retirement portfolio invested in stocks and bonds. The optimal controls are determined based on a parametric model fitted to almost a century of market data. The optimal control algorithm is based on dynamic programming and the solution of a partial integro differential equation (PIDE) using Fourier methods. The optimal strategy (based on the parametric model) is tested out of sample using stationary block bootstrap resampling of the historical data. In terms of an expected total withdrawal, expected shortfall (EW-ES) efficient frontier, the tontine overlay dramatically outperforms an optimal strategy (without the tontine overlay), which in turn outperforms a constant weight strategy with withdrawals based on the ubiquitous four per cent rule.

Keywords: tontine, decumulation, expected shortfall, optimal stochastic control
JEL codes: G11, G22
AMS codes: 91G, 65N06, 65N12, 35Q93

1 Introduction

It is now commonplace to observe that defined benefit (DB) plans are disappearing. A recent OECD study (OECD, 2019) observes that less than 50% of pension assets in 2018 were held in DB plans in over 80% of countries reporting. Of course, the level of assets in defined contribution (DC) plans is a lagging indicator, since historically many employees were covered by traditional DB plans. These traditional DB plans still have a sizeable share of pension assets, simply because these plans have accumulated contributions over a longer period of time.

Consider the typical case of a DC plan investor upon retirement. Assuming that the investor has managed to accumulate a reasonable amount in her DC plan, the investor now faces the problem

[a]David R. Cheriton School of Computer Science, University of Waterloo, Waterloo ON, Canada N2L 3G1, paforsyt@uwaterloo.ca, +1 519 888 4567 ext. 34415.

[b]School of Accounting and Finance, University of Waterloo, Waterloo ON, Canada N2L 3G1, kvetzal@uwaterloo.ca, +1 519 888 4567 ext. 46518.

[c]Richardson Wealth, 540 Bingemans Centre Drive, Suite 100, Kitchener, ON, Canada N2B 3X9, Graham.Westmacott@RichardsonWealth.com, +1 800 451 4748.

of determining a decumulation strategy, i.e. how to invest and spend during retirement. It is often suggested that retirees should purchase annuities, but this is quite unpopular (Peijnenburg et al., 2016). MacDonald et al. (2013) note that this avoidance of annuities can be entirely rational.

A major concern of DC plan investors during the decumulation phase is running out of savings. Possibly the most widely cited benchmark strategy is the *4% rule* (Bengen, 1994). This rule posits a retiree who invests in a portfolio of 50% stocks and 50% bonds, rebalanced annually, and withdraws 4% of the original portfolio value each year (adjusted for inflation). This strategy would have never depleted the portfolio over any rolling 30-year historical period tested by Bengen on US data. This rule has been revisited many times. For example, Guyton and Klinger (2006) suggest several heuristic modifications involving withdrawal amounts and investment strategies.

Another approach has been suggested by Waring and Siegel (2015), which they term an Annually Recalculated Virtual Annuity (ARVA) strategy. The idea is that the amount withdrawn in any given year should be based on the cash flows from a virtual (i.e. theoretical) fixed term annuity that could be purchased using the existing value of the portfolio. In this case the DC plan can never run out of cash, but the withdrawal amounts can become arbitrarily small.

Turning to asset allocation strategies, Irlam (2014) used dynamic programming methods to conclude that deterministic (i.e. glide path) allocation strategies are sub-optimal. Of course, the asset allocation strategy and the withdrawal strategy are intimately linked. A more systematic approach to the decumulation problem involves formulating decumulation strategies as a problem in optimal stochastic control. The objective function for this problem involves a measure of risk and reward, which are, of course, conflicting measures. Forsyth (2022b) uses the withdrawal amount and the asset allocation (fraction in stocks and bonds) as controls. The measure of reward is the total (real) accumulated withdrawal amounts over a 30-year period. The withdrawal amounts have minimum and maximum constraints, hence there is a risk of depleting the portfolio. The measure of risk is the expected shortfall at the 5% level, of the (real) value of the portfolio at the 30-year mark. Utilizing both withdrawal amounts and asset allocation as controls considerably reduces the risk of portfolio depletion compared to fixed allocation or fixed withdrawal strategies.

A recent innovation in retirement planning involves the use of modern tontines (see, e.g. Donnelly et al., 2014; Donnelly, 2015; Milevsky and Salisbury, 2015; Fullmer, 2019; Weinert and Gründl, 2021; Winter and Planchet, 2022; Milevsky, 2022). In a tontine, the investor makes an irrevocable investment in a pooled fund for a fixed time frame. If the investor dies during the time horizon of the investment, the investor's portfolio is divided amongst the remaining (living) members of the fund. If the investor survives until the end of the time horizon, then she will earn mortality credits from those members who have passed away, in addition to the return on her investment portfolio. Note that in some tontine formulations (e.g. Donnelly et al., 2014)) there is a final mortality credit paid to the estate of the deceased, while in other formulations (e.g. Fullmer, 2019)) there is no final payment. Unlike an annuity, there are no guaranteed cash flows, since the funds are typically invested in risky assets. Moreover, the mortality credits received are stochastic, depending on the realized mortality of investors in the pool. Since there are no guarantees, the expected cash flows from a tontine are larger than for an annuity with the same initial investment. Some authors have argued that the *annuity puzzle* should be replaced by the *tontine puzzle*, i.e. since tontines seem to very efficient products for pooling longevity risk, it is puzzling that the tontine market is still in its infancy (Chen and Rach, 2022). However, as noted by sources such as Milevsky and Salisbury (2015), it is important to distinguish between two components of mortality risk. *Idiosyncratic* mortality risk is related to the probability of death in a period for any individual plan member in accordance with a specified mortality table, while *systematic* mortality risk considers the mortality experience of the pool as a whole, i.e. whether or not the aggregate number of deaths in a period is roughly equal to that predicted by the mortality table. The potential issue for a tontine is that if longevity improves for the pool as a whole beyond that projected in the mortality table, then the mortality credits received will be lower (or received later) than anticipated. Tontines offer insurance only against the idiosyncratic component of mortality risk. This is in contrast to annuities, which offer protection against both components. However, as noted by Milevsky and Salisbury (2015), the extra insurance provided by annuities makes them more costly than tontines, and investors choosing between tontines and annuities would have to judge whether the additional longevity protection of an annuity is worth the associated higher cost.

Pooled funds with tontine characteristics have been in use for some time. The variable annuity funds offered by TIAA[1], the University of British Columbia pension plan[2], and the Australian Q-super fund[3] can all be viewed as having tontine characteristics. In the Canadian context, the Purpose Longevity Plan[4] was launched in 2021 and the Guardian Capital Modern Tontine a year later. Key similarities are that both use a mutual fund structure, restrict participation into age cohorts to be actuarially fair with a small pool size, are non-transferable and have redemption risk. The Purpose plan is more like an annuity that offers income for life that is enhanced by mortality credits, but without any guarantees. This contrasts with the Guardian Capital product[5] which accumulates mortality credits over 20 years and then pays out a lump sum to hedge against the risk of investors outliving their capital. This has more in common with term insurance but benefits the living.

We should also mention that retail investors may find the concept of a tontine appealing, simply due to the peer-to-peer model for managing longevity risk, which is also consistent with the trend towards financial disintermediation.[6] However, tontines may also require changes to existing legislation in some jurisdictions (MacDonald et al., 2021). There have also been suggestions for government management of tontine accounts (Fullmer and Forman, 2022; Fuentes et al., 2022). The attractiveness of tontines from a behavioral finance perspective is discussed in Chen et al. (2021). See Bär and Gatzert (2023) for an overview comparison of modern tontines with existing decumulation products.

Our focus in this article is on individual tontine accounts (Fullmer, 2019), whereby the investor has full control over the asset allocation in her account. We also allow the investor to control the withdrawal amount from the account, subject to maximum and minimum constraints. Usually it is suggested that withdrawal amounts from a tontine account cannot be increased to avoid moral hazard issues.[7] However, we view the maximum withdrawal as the desired withdrawal, allowing temporary reductions in withdrawals to minimize sequence of return risk and probability of ruin.

Consider an investor whose objective function uses reward as measured by total expected accumulated (real) withdrawals (EW) over a 30-year period. As a measure of risk, the investor uses the expected shortfall (ES) of the portfolio at the 30-year point. We define the expected shortfall to be the mean of the worst 5% of the outcomes after 30 years. The investor's controls are the amount

[1] `https://www.tiaa.org/public/`

[2] `https://faculty.pensions.ubc.ca/`

[3] `https://qsuper.qld.gov.au/`. However, the Q-super fund takes the approach of averaging mortality credits over the entire pool, giving age-independent mortality credits. This appears to violate actuarial fairness `https://i3-invest.com/2021/04/behind-qsupers-retirement-design/`. The Q-super fund is perhaps more properly termed a collective defined contribution (CDC) fund. CDCs (`https://www.ft.com/content/10448b2c-1141-4d2e-943c-70cce2caec52`) have been criticized for lack of transparency and fairness.

[4] `https://www.retirewithlongevity.com/fund`

[5] `https://www.guardiancapital.com/investmentsolutions/guardpath-modern-tontine-trust/`

[6] See van Benthem et al. (2018) for an experiment with setting up a tontine using blockchain techniques.

[7] An obvious case would be if an investor was given a medical diagnosis with a high probability of a poor outcome, at which point the investor would withdraw all remaining funds in her account.

withdrawn each year and the allocations to stocks and bonds. The investor follows an optimal strategy to maximize this objective function.

Alternatively, the investor can use the same objective function with the same controls, but this time add a tontine overlay (i.e. the investor is part of a pooled tontine). The investor retains control over the withdrawals (subject of course to the same maximum and minimum constraints) and the investment allocation strategy. Of course, we expect that the investor who uses the tontine overlay would achieve a better result than without the overlay, due to the mortality credits earned (we assume that the investor does not pass away during the 30-year horizon). However, this does not come without a cost. If the investor passes away during the horizon, then her portfolio is forfeited. Therefore, the investor must be compensated with a sizeable reduction in the risk of portfolio depletion, compared to the no-tontine overlay case. The objective of this article is to quantify this reduction, assuming optimal policies are followed in each case.

More precisely, we consider a 65-year old retiree who can invest in a portfolio consisting of a stock index and a bond index, with yearly withdrawals and rebalancing. The investor seeks to maximize the multi-objective function in terms of the risk and reward measures described above, evaluated at the 30-year horizon (i.e. when the investor is 95).

We calibrate a parametric stochastic model for real (i.e. inflation-adjusted) stock and bond returns to almost a century of market data. We then solve the optimal stochastic control problem numerically, using dynamic programming. Robustness of the controls is then tested using block bootstrap resampling of the historical data.

Our main conclusion is that for a reasonable specification of acceptable tail risk (i.e. expected shortfall), the expected total cumulative withdrawals (EW) are considerably larger with the tontine overlay, compared to without the overlay. This conclusion holds even if the tontine overlay has fees of the order of 50-100 basis points (bps) per year. Consequently, if the retiree has no bequest motive, and is primarily concerned with the risk of depleting her account, then a tontine overlay is an attractive solution.

It is also interesting to note that the optimal control for the withdrawal amount is (to a good approximation) a bang-bang control, i.e. it is only optimal to withdraw either the maximum or minimum amount in any year. The allocation control essentially starts off with 40-50% allocation to stocks. The median allocation control then rapidly reduces the fraction in equities to a very small amount after $5 - 10$ years. The median withdrawal control starts off at the minimum withdrawal amount, and then rapidly increases withdrawals to the maximum after $2 - 5$ years. The precise timing of the switch from minimum withdrawal to maximum withdrawal depends on how much depletion risk (ES) the investor is prepared to take.

2 Problem Setting

In order to be consistent with practitioner literature, we will consider the scenario set out in Bengen (1994). This scenario posits that the retiree desires fixed (real) minimum cash flows and that the cash flows are desired over a fixed planning horizon. The investor's primary concern is that of exhausting savings during the planning horizon. We conjecture that the reason that the Bengen rule continues to be very popular in practice is that it directly addresses the typical concerns of retirees (see, e.g. Ameriks et al., 2001; Scott et al., 2009; Pfau, 2015; Ruthbah, 2022; Daily et al., 2023).

It may seem counterintuitive to use a fixed, relatively long term planning horizon (usually 30 years for a 65 year old retiree). Based on current mortality tables, the probability of a 65-year old Canadian male reaching the age of 95 is about 0.13. A 95-year old male has only a one in six chance of reaching his 100^{th} birthday.

However, surveys show that retirees fear exhausting their savings more than death (Hill, 2016). Along these lines Pfau (2018) writes:

> *"Play the long game. A retirement income plan should be based on planning to live, not planning to die. A long life will be expensive to support, and it should take precedence over death planning."*

Therefore, use of a long, fixed planning horizon has become the default test of the risk of running out of funds.

It is also of interest to not only estimate the probability of ruin but also the size of the shortfall. To this end, we allow the retiree to continue to withdraw the minimum desired cash flows under a stochastic scenario where savings are exhausted. This debt accumulates at the borrowing rate. This allows us to measure the size of the shortfall for this scenario. We use the mean of the worst 5% of the outcomes as a quantitative measure of shortfall. A negative shortfall signals that the retiree has run out of cash and has an accumulated debt. An accumulated debt of one dollar at age 95 is certainly less concerning than a debt of $100,000 at that time. Although we are measuring the shortfall at the same point in time, the case with a more negative shortfall is likely due to having run out of money earlier and having debt accumulate. This can occur due to the forced minimum annual withdrawals. Effectively, the accumulated debt in this case penalizes strategies which run out of cash early during the planning horizon.

How would this work in practice? Basically, we assume that the investor divides his wealth into *mental accounts*, containing funds intended for different purposes (e.g. current spending or future needs). The standard life cycle model assumes that all wealth is completely fungible. In contrast, the behavioral approach posits that all wealth is not fungible, and that *mental bucketing* is commonplace (Shefrin and Thaler, 1988). In particular, we will assume that the investor has mortgage-free residential real estate, which is in a separate mental account. This real estate is to be considered a hedge of last resort, if needed.[8] If the investments work out well, or if the retiree passes away, this real estate can be considered as a bequest.

It is commonplace in actuarial applications to mortality-weight cash flows. While this is clearly appropriate for annuity providers, it does not seem to be very informative for an individual retiree. Consider the perspective of a 65-year old male with median life expectancy of about 87. The standard mortality-weighting approach would weight the minimum cash flows at 22 years after retirement by one half. However, if the retiree is *planning to live* rather than planning to die, he needs the entire minimum cash flow at age 87, not half of it.

As noted above, we assume that the investor trades off the reward of total real withdrawals over the 30-year horizon with the risk of expected shortfall at the end of the horizon. This risk/reward tradeoff is reminiscent of the tradeoff between expected return and standard deviation from traditional portfolio theory, but with different measures of risk and reward. An obvious alternative would be to specify a utility function. Most prior academic studies involving various forms of tontines have done so, typically assuming constant relative risk aversion (CRRA) utility (see, e.g. Milevsky and Salisbury, 2015; 2016; Bernhardt and Donnelly, 2019; Chen et al., 2019; 2021). However, we believe it is useful to consider an objective function based on the expected withdrawal, expected shortfall criteria. First, utility functions in principle should include all of the investor's wealth, but this would be incompatible with the mental accounts framework discussed above. Second, related to the inclusion of the entire amount of the investor's wealth, utility functions often have infinite marginal

[8] Pfeiffer et al. (2013) discuss how a reverse mortgage can be used to hedge the risk of exhausting savings.

utility of wealth at zero. This means that the investor would always avoid reducing wealth to zero.[9] However, this is incompatible with a minimum withdrawal constraint: if the investor must withdraw some funds each year, there is inevitably some chance of insolvency if the investor survives long enough. Third, we believe that in practice it is easier to communicate with retired clients if the discussion is framed in terms of monetary amounts, which can be directly compared to the value of a residential real estate hedge.

3 Overview of Individual Tontine Accounts

3.1 Intuition

We give a brief overview of modern tontines in this section. We restrict attention to the case of an individual tontine account (Fullmer, 2019), which is a constituent of a perpetual tontine pool. Consider a pool of m investors, who are alive at time t_{i-1}. Let v_i^j be the balance in the portfolio of investor j at time t_i. We assume that $v_i^j \geq 0$. In a tontine, if investor j participates in a tontine pool in time interval (t_{i-1}, t_i), and investor j dies in that interval, then her portfolio v_i^j is forfeited and given to the surviving members of the pool in the form of mortality credits (gains). Suppose that the probability that j dies in (t_{i-1}, t_i) is q_{i-1}^j.

Consider tontine members $j = 1, \ldots, m$ who are alive at t_{i-1}. Let

$$
\begin{aligned}
\mathbf{1}_i^j &= \begin{cases} 1 & \text{Investor } j \text{ is alive at } t_{i-1} \text{ and alive at } t_i \\ 0 & \text{Investor } j \text{ is alive at } t_{i-1} \text{ and dead at } t_i \end{cases} \\
E_{i-1}[\mathbf{1}_i^j] &= 1 - q_{i-1}^j \ ,
\end{aligned}
\tag{3.1}
$$

where $E_{i-1}[\cdot]$ denotes an expectation operator conditional on mortality information known at t_{i-1}.

Let the tontine gain (mortality credit) for investor j, conditional on $\mathbf{1}_i^j = 1$, for the period (t_{i-1}, t_i), paid out at time t_i, be denoted by c_i^j. The tontine will be a fair game if, for each player j, the expected gain from participating in the tontine is zero[10],

$$
-v_i^j E_{i-1}[1 - \mathbf{1}_i^j] + E_{i-1}[\mathbf{1}_i^j] \cdot \mathbb{E}_{i-1}^j[c_i^j] = 0 \ , \tag{3.2}
$$

where

$$
\mathbb{E}_{i-1}^j[\cdot] \equiv E_{i-1}[\ \cdot \ \big| \mathbf{1}_i^j = 1;\ \{v_i^j\}_{j=1}^m] \tag{3.3}
$$

i.e. conditional on member j being alive at t_i, with known (realized) values of investment accounts $\{v_i^j\}_{j=1}^m$. This is because the member enters into the pool at t_{i-1}, with unknown mortality states of the other members of the pool at t_i. However, the mortality credits are divided up based on the realized investment accounts at t_i. From equation (3.2) this gives

$$
\mathbb{E}_{i-1}^j[c_i^j] = \overbrace{\left(\frac{q_{i-1}^j}{1 - q_{i-1}^j} \right)}^{\textit{Gain rate}} v_i^j \ . \tag{3.4}
$$

[9]Of course, the origin of the utility function can be shifted so that infinite marginal utility is obtained at a finite negative value of wealth. However, this just shifts the problem, unless the negative wealth value is set to the total maximum amount of withdrawals, which will underpenalize running out of savings.

[10]This assumes no fees or transaction costs. In practice, such expenses will result in an expected gain of less than zero.

We emphasize that v_i^j are unknown at t_{i-1} since we allow investment in risky assets. However, in terms of fairness, we only require that the realized investment proceeds are reallocated fairly, taking into account random mortality.

Note that the right hand side of equation (3.4) is independent of the other members' accounts in the pool, their investment strategies, and their mortality status. This is surprising and counter-intuitive, as the expectation operator on the left hand side as defined in equation (3.3) does depend on the values of other members' accounts and implicitly their investment strategies, since these determine the account values at the end of the period. In fact, it is easy to construct somewhat pathological cases where equation (3.4) will fail. For example, suppose that every investor other than j invests their entire account value in losing lottery tickets, driving their account values to zero. If every other investor's account value is zero, that would include the accounts of investors who pass away during the period and there would be no tontine gains to be distributed to investor j. Underlying the apparent independence of the right hand side of equation (3.4) from the values of other members' accounts are implicit assumptions that any member's gain is small compared to the overall expected amount of mortality credits available and that the pool is sufficiently large so that the realized amount of mortality credits is approximately equal to the expected amount. This *small bias condition* is stated more precisely below (see Condition 3.1) with additional discussion in Remark 3.4.

We also have the budget rule that the total of the tontine gains distributed is equal to the total amounts forfeited

$$\sum_{j=1}^{m} \mathbf{1}_i^j c_i^j = \sum_{j=1}^{m} (1 - \mathbf{1}_i^j) v_i^j \ . \tag{3.5}$$

Let

$$\Omega_i = \{\{\mathbf{1}_i^j\}_{j=1}^m, \{v_i^j\}_{j=1}^m\} \ . \tag{3.6}$$

In general $c_i^j = c_i^j(\Omega_i \big| \mathbf{1}_i^j = 1)$.

Remark 3.1 (No tontine gains to members who have died)**.** *Note that equation (3.2) assumes that no tontine gains accrue to members who have just died. This will maximize tontine gains of survivors, which is the focus of the current study. Several previous studies have specified payments to the estates of members who die in* (t_{i-1}, t_i) *(see, e.g. Bernhardt and Donnelly, 2018; Hieber and Lucas, 2022; Denuit et al., 2022).*

We can always write c_i^j as

$$c_i^j = \left(\frac{q_{i-1}^j}{1 - q_{i-1}^j} \right) v_i^j \ H_i^j(\Omega_i \big| \mathbf{1}_i^j = 1) \ , \tag{3.7}$$

for some function $H_i^j(\Omega_i \big| \mathbf{1}_i^j = 1)$. Then the fairness condition (3.4) becomes

$$\mathbb{E}_{i-1}^j [H_i^j] = 1 \ . \tag{3.8}$$

It is also desirable to impose the condition that in each period a surviving investor is never made worse off by participating in a tontine pool, i.e. the tontine gain is non-negative

$$H_i^j(\Omega_i \big| \mathbf{1}_i^j = 1) \geq 0 \ . \tag{3.9}$$

It is convenient to summarize the essential tontine properties. These properties are largely the same as in Hieber and Lucas (2022), with the exception that there are no payments to the estates of deceased members.

Property 3.1 (Tontine Properties)**.** *The desirable properties of a tontine are*

(i) Fairness: $\mathbb{E}^j_{i-1}[H^j_i] = 1$; $j = 1,\ldots,m$ *(see equation (3.8)).*

(ii) Budget constraint (3.5).

(iii) Tontine gain non-negativity: $H^j_i \geq 0$; $j = 1,\ldots,m$.

Sharing rules which satisfy Property 3.1 are discussed in Sabin (2010; 2011), and sharing rules for the case where payments are made to just deceased members are described in, for example Bernhardt and Donnelly (2018), Hieber and Lucas (2022), and Denuit et al. (2022), and the references therein.

3.2 A Simplified Approach

One of the downsides of sharing rules which exactly satisfy Property 3.1 for finite-sized pools is that these rules are somewhat complex. Many of these exact rules require processing deaths one at a time. It is argued in Sabin and Forman (2016) and Fullmer and Sabin (2019) that complex sharing rules can impede consumer acceptance. These authors argue that it is sufficient to have simple rules which satisfy Property 3.1(i)-(iii) in the limit of large, perpetual pools.

3.2.1 Group Gain

In the terminology of Sabin and Forman (2016), we define the group gain G_i as

$$G_i = \frac{\sum_k (1 - \mathbf{1}^k_i) v^k_i}{\sum_k \mathbf{1}^k_i \left(\frac{q^k_{i-1}}{1-q^k_{i-1}}\right) v^k_i} \ . \tag{3.10}$$

Note that the G_i is the same for all members j. G_i has the convenient interpretation as being the ratio of the total realized mortality credits to the total expected credits for survivors.

Sabin and Forman (2016) suggest the following simplified sharing rule

$$c^j_i = \left(\frac{q^j_{i-1}}{1 - q^j_{i-1}}\right) v^j_i \ G_i \ , \tag{3.11}$$

which uses the group gain G_i in place of the function H^j_i in equation (3.7). By assumption, $v^j_i \geq 0$, hence Property 3.1(iii) is satisfied.

The total tontine gains are

$$\begin{aligned} \text{Total Tontine Gains} &= \sum_{j=1}^m \mathbf{1}^j_i c^j_i = G_i \sum_{j=1}^m \mathbf{1}^j_i \left(\frac{q^j_{i-1}}{1 - q^j_{i-1}}\right) v^j_i \\ &= \sum_{j=1}^m (1 - \mathbf{1}^j_i) v^j_i \ = \ \text{Total Forfeited} \ , \end{aligned} \tag{3.12}$$

hence Property 3.1(ii) is satisfied.

In essence, the group gain G_i in equation (3.10) is a scaling factor to adjust for actual deaths compared to expected deaths, as suggested in Piggott et al. (2005) and Qiao and Sherris (2013).

Remark 3.2 ($\sum_{j=1}^m \mathbf{1}^j_i c^j_i = 0$)**.** *Note that equation (3.10) is undefined if all members die in* (t_{i-1},t_i). *We assume that the tontine is large (in terms of members) and perpetual (i.e. open to new members), so that the probability of all members dying in a period is negligible. For mathematical completeness, we can suppose that if all members die in* (t_{i-1},t_i), *we collapse the tontine, and distribute all remaining account values* v^j_i *to the estates of members* j.

However, in general Property 3.1(i) will not be satisfied for a finite-sized pool, i.e.

$$\exists j \ s.t. \ \mathbb{E}^j_{i-1}[G_i \big| \mathbf{1}^j_i = 1] < 1, \quad j \in \{1, \cdots, m\} \ . \tag{3.13}$$

This means that there is a bias that favors some members over others, i.e. some members have a negative expected gain, implying that other members must have a positive expected gain. This is illustrated in Winter and Planchet (2022), using an example with a pool consisting of a large number of young investors (with small individual portfolios), and a single elderly member with a large portfolio. The elderly member effectively subsidizes the younger members. Sabin and Forman (2016) show that the bias is negligible under the following conditions:

Condition 3.1 (Small bias condition)**.** *Suppose that:*

(a) *the pool of participants in the tontine is sufficiently large; and*

(b) *the expected amount forfeited by all members is large compared to any member's nominal gain, i.e.*

$$\left(v^j_i \frac{q^j_{i-1}}{1 - q^j_{i-1}} \right) \ll \sum_k q^k_{i-1} v^k_i \quad ; \quad j = 1, \ldots, m \ . \tag{3.14}$$

Then the bias is negligibly small (Sabin and Forman, 2016).

Equation (3.14) is essentially a diversification requirement: no member of the pool has an abnormally large share of the total pool capital. In addition, of course, if the pool is sufficiently large, then the actual number of deaths in (t_{i-1}, t_i) will converge to the expected number of deaths.

In Fullmer and Sabin (2019), simulations were carried out to determine the magnitude of the volatility of G_i under practical sizes of tontine pools. Given a tontine pool of 15,000 members, with varying ages, initial capital, and randomly assigned investment policies (i.e. the bond/stock split), the simulations showed that $E[G_i] \simeq 1$ and that the standard deviation was about 0.1. This standard deviation at each t_i actually resulted in a smaller effect over a long term (assuming that the tontine member lived long enough). This is simply because everybody dies eventually, so that if fewer deaths than expected are observed in a year, then more deaths will be observed in later years, and vice versa. More detailed analysis of the probability density of G_i is given in Denuit and Vernic (2018), with a slightly different use of the factor q^j_i.

Henceforth we will assume that the pool is sufficiently large and that it satisfies the diversity condition (3.14), so that there is no significant error in assuming that $G_i \equiv 1$. More precisely, our optimal control problem will be formulated assuming that

$$c^j_i = \left(\frac{q^j_{i-1}}{1 - q^j_{i-1}} \right) v^j_i \ . \tag{3.15}$$

As a sanity check, we also carry out a test where we simulate the effect of randomly varying G, based on the statistics of the simulations in Fullmer and Sabin (2019). Our results show that the effect of randomness in G can be safely ignored for a reasonably sized tontine pool. To be more precise, we will modify equation (3.15) so that

$$c^j_i = \left(\frac{q^j_{i-1}}{1 - q^j_{i-1}} \right) v^j_i \hat{G}_i \ . \tag{3.16}$$

where $\hat{G}_i$ is a random variable. We will give a numerical example showing the effects of randomness of $\hat{G}_i$ in Monte Carlo simulations. However, our computation of the optimal strategy will always assume $G_i \equiv 1$.

For notational convenience, we define the tontine gain rate at t_i for investor j as

$$(\mathbb{T}_i^g)^j = \left(\frac{q_{i-1}^j}{1 - q_{i-1}^j} \right) . \tag{3.17}$$

In our optimal control formulation, we will typically drop the superscript j from equation (3.17),

$$\mathbb{T}_i^g = \left(\frac{q_{i-1}}{1 - q_{i-1}} \right) , \tag{3.18}$$

since we will consider a given investor j with conditional mortality probability of q_{i-1} in (t_{i-1}, t_i). Using this notation, equation (3.15) becomes (for a fixed investor j)

$$c_i = \mathbb{T}_i^g v_i . \tag{3.19}$$

Remark 3.3 (Other sharing rules which satisfy Property 3.1(i)-(iii))**.** *We have used sharing rule (3.11) as an example of a practical scheme, which only approximately satisfies Property 3.1(i)-(iii). However, our optimal control formulation will also apply to any sharing rule which satisfies Property 3.1(i)-(iii), provided that the pool is large enough so that var*$[H_i^j]$ *is small, where var*$[\cdot]$ *denotes variance.*

Remark 3.4 (Effect of investment decisions of other members in the pool)**.** *Provided the small bias condition (Condition 3.1) holds, it does not matter what investment strategy is followed by any given investor in period* (t_{i-1}, t_i)*. Each investor can choose whatever policy they like, since only the observed final portfolio value at* t_i *matters.*[11] *At first glance, the idea that the investment strategies of other pool members do not affect the strategy of the individual investor seems counter-intuitive. However, it is important to realize that it is the reallocation of the realized forfeiture amounts that matters. We provide some brief intuition here; for further discussion, see Fullmer and Sabin (2019) and Winter and Planchet (2022) and references cited therein.*

Assuming that equation (3.8) holds with sufficient accuracy (and with the group gain G_i *in place of the function* H_i^j*), then each member's expected gain will be given by equation (3.4). Consider an aggressive investor in a pool dominated by conservative investors. Assume that the aggressive investor has a larger account balance than the conservative investors. Since the aggressive investor's stake is larger, she will get a larger share of the (smaller) forfeitures. In addition, if var*(G_i) *is small, then the realized mortality credits will be close to the expected value (3.4). Moreover, we will show that the optimal strategy for an individual investor (computed assuming* $G_i \equiv 1$*) is robust to volatile* G_i*.*

[11] Observe that our earlier pathological lottery example would violate the small bias condition: investor j (the only investor who doesn't invest entirely in losing lottery tickets) has the entire pool capital after the lottery since the account value for every other member of the pool goes to zero.

3.3 Systematic Risk

It is worth emphasizing again the distinction made in the Introduction between the idiosyncratic and systematic components of mortality risk. The idiosyncratic mortality risk can be made small if the pool of investors is sufficiently large. The same cannot be said for systematic mortality risk, e.g. unexpected mortality improvement. To be precise, the members of the pool bear the systematic risk that $\mathbb{E}[G_i]$ will be significantly less than one, perhaps due to medical advances. As noted above, traditional annuities provide protection against both idiosyncratic and systematic, but at higher cost compared to tontines (Milevsky and Salisbury, 2015).

In principle, it is possible to assume a stochastic process for mortality improvement (see, e.g. Gemmo et al., 2020), and then solve the optimal control problem with this additional risk factor. However, this would be computationally infeasible for our current approach based on partial differential equations. Alternatively, machine learning techniques appear to be a promising method for solving high dimensional control problems in finance (see, e.g. Li and Forsyth, 2019; Ni et al., 2022; van Staden et al., 2023; Chen et al., 2023), and this might be an approach for including systematic mortality risk in this case. However, this is beyond the scope of the current work.

3.4 Variable Withdrawals

We will allow the individual tontine member to withdraw variable amounts, subject to minimum and maximum constraints. We remind the reader that if a tontine pool is strictly actuarially fair, then in theory there are no constraints on withdrawals and injections of cash (Bräutigam et al., 2017).

However, in practice, since pools are finite sized, heterogeneous, and mortality credits are not distributed at infinitesimal intervals, we do not allow arbitrarily large withdrawals. This avoids moral hazard issues.

Since we have a minimum withdrawal amount in each time period, there is a risk of running out of cash. We assume that if the minimum withdrawal exceeds the available amount in the tontine account, the tontine account goes to zero, all trading in this account ceases, and the remaining part of the withdrawal (and any subsequent withdrawals) is funded by debt, which accumulates at the borrowing rate. Of course, insolvent investors will not receive any mortality credits.

In practice, if the tontine account becomes zero, the retiree has to fund expenses from another source. We implicitly assume that the tontine member has other assets which can be used to fund this minimum consumption level (e.g. real estate). Of course, we aim to make this a very improbable event. In fact, this is the reason why we allow variable withdrawals. We can regard the upper bound on the withdrawals as the desired consumption level, but we allow the tontine member to reduce (hopefully only temporarily) their withdrawals, to minimize risk of depletion of their tontine account.

3.5 Money Back Guarantees

In practice, we observe that many tontine funds offer a money back guarantee.[12] This is usually specified as a return of the initial (nominal) investment less any withdrawals (if the sum is non-negative) at the time of death. We do not consider such guarantees in this work, focusing on the pure tontine aspect, which has no guarantees and presumably the highest possible expected total withdrawals. A money back guarantee would have to be hedged, which would reduce returns. In

[12] `https://qsuper.qld.gov.au/`.

practice, this guarantee could be priced separately, and added as an overlay to the tontine investment if desired.

3.6 Survivor Benefits

Many DB plans have survivor benefits which are received by a surviving spouse. A typical case would involve the surviving spouse receiving $60\% - 75\%$ of the yearly pension after the DB plan holder dies.

Consider the following case of a male, same-sex couple, both of whom are exactly the same age. As an extreme case, suppose the survivor benefit is 100% of the tontine cash flows, which continue until the survivor dies. From the CPM2014 table from the Canadian Institute of Actuaries[13], the probability that an 85-year old Canadian male dies before reaching the age of 86 is about .076. Assuming that the mortality probabilities are independent for both spouses, then the probability that both 85-year old spouses die before reaching age 86, conditional on both living to age 85 is $(.076)^2 \simeq .0053$. From equation (3.17), the tontine gain rate per year is

$$\text{tontine gain rate } = \frac{.0053}{1 - .0053} \simeq .0053 \ . \tag{3.20}$$

We will assume in our numerical examples that the base case fee charged for managing the tontine is 50 bps per year. This means that net of fees, there are essentially no tontine gains for our hypothetical couple for the first 20 years of retirement, which is surely undesirable. Once one of the partners passes away, the tontine gain rate will, of course, take a jump in value.

As another extreme case, suppose that the surviving spouse receives 50% of the tontine cash flows. In this case, the total cash flows accruing to this couple are exactly the same as those that would have resulted from dividing the original total wealth in half and then having each spouse invest in their own individual tontine.

It is possible to determine the distribution of the cash flows for a survivor benefit which is intermediate to these edge cases. However, this requires additional state variables in our optimal control problem, and is probably best tackled using a machine learning approach (Li and Forsyth, 2019; Ni et al., 2022). We will leave this case for future work, and focus attention on the individual tontine case with no survivor benefit. Note that in the tontine context, survivor benefits are typically provided by a separate insurance overlay.[14]

4 Formulation

We assume that the investor has access to two funds: a broad market stock index fund and a constant maturity bond index fund. The investment horizon is T. Let S_t and B_t respectively denote the real (inflation adjusted) *amounts* invested in the stock index and the bond index respectively. In general, these amounts will depend on the investor's strategy over time, as well as changes in the real unit prices of the assets. In the absence of an investor determined control (i.e. cash withdrawals or rebalancing), all changes in S_t and B_t result from changes in asset prices. We model the stock index as following a jump diffusion.

In addition, we follow the usual practitioner approach and directly model the returns of the constant maturity bond index as a stochastic process (see, e.g. Lin et al., 2015; MacMinn et al., 2014). Consistent with the stock index, we will assume that the constant maturity bond index

[13] `www.cia-ica.ca/docs/default-source/2014/214013e.pdf`.
[14] `https://i3-invest.com/2021/04/behind-qsupers-retirement-design/`.

also follows a jump diffusion. Empirical justification for this can be found in Forsyth et al. (2022), Appendix A. This will also be discussed in Section 9.

Let $S_{t^-} = S(t - \epsilon), \epsilon \to 0^+$, i.e. t^- is the instant of time before t, and let ξ^s be a random number representing a jump multiplier. When a jump occurs, $S_t = \xi^s S_{t^-}$. Allowing for jumps permits modelling of non-normal asset returns. We assume that $\log(\xi^s)$ follows a double exponential distribution (Kou, 2002; Kou and Wang, 2004). If a jump occurs, u^s is the probability of an upward jump, while $1 - u^s$ is the chance of a downward jump. The density function for $y = \log(\xi^s)$ is

$$f^s(y) = u^s \eta_1^s e^{-\eta_1^s y} \mathbf{1}_{y \geq 0} + (1 - u^s) \eta_2^s e^{\eta_2^s y} \mathbf{1}_{y < 0} \ . \tag{4.1}$$

We also define

$$\gamma_\xi^s = E[\xi^s - 1] = \frac{u^s \eta_1^s}{\eta_1^s - 1} + \frac{(1 - u^s)\eta_2^s}{\eta_2^s + 1} - 1 \ . \tag{4.2}$$

In the absence of control, S_t evolves according to

$$\frac{dS_t}{S_{t^-}} = \left(\mu^s - \lambda_\xi^s \gamma_\xi^s\right) \, dt + \sigma^s \, dZ^s + d \left(\sum_{i=1}^{\pi_t^s} (\xi_i^s - 1) \right) , \tag{4.3}$$

where μ^s is the (uncompensated) drift rate, σ^s is the volatility, dZ^s is the increment of a Wiener process, π_t^s is a Poisson process with positive intensity parameter λ_ξ^s, and ξ_i^s are i.i.d. positive random variables having distribution (4.1). Moreover, ξ_i^s, π_t^s, and Z^s are assumed to all be mutually independent.

Similarly, let the amount in the bond index be $B_{t^-} = B(t - \epsilon), \epsilon \to 0^+$. In the absence of control, B_t evolves as

$$\frac{dB_t}{B_{t^-}} = \left(\mu^b - \lambda_\xi^b \gamma_\xi^b + \mu_c^b \mathbf{1}_{\{B_{t^-} < 0\}}\right) \, dt + \sigma^b \, dZ^b + d \left(\sum_{i=1}^{\pi_t^b} (\xi_i^b - 1) \right) , \tag{4.4}$$

where the terms in equation (4.4) are defined analogously to equation (4.3). In particular, π_t^b is a Poisson process with positive intensity parameter λ_ξ^b, and ξ_i^b has distribution

$$f^b(y = \log \xi^b) = u^b \eta_1^b e^{-\eta_1^b y} \mathbf{1}_{y \geq 0} + (1 - u^b) \eta_2^b e^{\eta_2^b y} \mathbf{1}_{y < 0} \ , \tag{4.5}$$

and $\gamma_\xi^b = E[\xi^b - 1]$. ξ_i^b, π_t^b, and Z^b are assumed to all be mutually independent. The term $\mu_c^b \mathbf{1}_{\{B_{t^-} < 0\}}$ in equation (4.4) represents the extra cost of borrowing (the spread).

The diffusion processes are correlated, i.e. $dZ^s \cdot dZ^b = \rho_{sb} \, dt$. The stock and bond jump processes are assumed mutually independent. See Forsyth (2020b) for justification of the assumption of stock-bond jump independence.

We define the investor's total wealth at time t as

$$\text{Total wealth } \equiv W_t = S_t + B_t. \tag{4.6}$$

The term "total wealth" refers to the sum of the values of the investor's tontine account plus any accumulated debt arising from insolvency due to the minimum required withdrawals. This is perhaps a bit misleading since it excludes the value of any additional assets that the investor has such as real estate or the value of government benefits, but it simplifies our exposition. We impose the constraints that (assuming solvency) shorting stock and using leverage (i.e. borrowing) are not permitted. As noted above, in case of insolvency, the portfolio is liquidated, trading stops and debt accumulates at the borrowing rate.

5 Notational Conventions

Consider a set of discrete withdrawal/rebalancing times $\mathcal{T}$

$$\mathcal{T} = \{t_0 = 0 < t_1 < t_2 < \ldots < t_M = T\} \tag{5.1}$$

where we assume that $t_i - t_{i-1} = \Delta t = T/M$ is constant for simplicity. To avoid subscript clutter, in the following, we will occasionally use the notation $S_t \equiv S(t), B_t \equiv B(t)$ and $W_t \equiv W(t)$. Let the inception time of the investment be $t_0 = 0$. We let $\mathcal{T}$ be the set of withdrawal/rebalancing times, as defined in equation (5.1). At each rebalancing time t_i, $i = 0, 1, \ldots, M-1$, the investor (i) withdraws an amount of cash $\mathfrak{q}_i$ from the portfolio, and then (ii) rebalances the portfolio. At $t_M = T$, the portfolio is liquidated and no cash flow occurs. For notational completeness, this is enforced by specifying $\mathfrak{q}_M = 0$.

In the following, given a time dependent function $f(t)$, we will use the shorthand notation

$$f(t_i^+) \equiv \lim_{\epsilon \to 0^+} f(t_i + \epsilon) \quad ; \quad f(t_i^-) \equiv \lim_{\epsilon \to 0^+} f(t_i - \epsilon) \quad . \tag{5.2}$$

Let

$$(\Delta t)_i = \begin{cases} \Delta t & i = 1, \ldots M, \\ 0 & i = 0 \ \text{ or } \ W(t_i^-) \leq 0 \end{cases} . \tag{5.3}$$

We assume that a tontine fee of $\mathbb{T}^f$ per unit time is charged at $t \in \mathcal{T}$, based on the total portfolio value at t_i^-, after tontine gains but before withdrawals.[15] Recalling the definition of tontine gain rate $\mathbb{T}_i^g$ in equation (3.18), we modify this definition to enforce no tontine gain at $t = 0$,

$$\mathbb{T}_i^g = \begin{cases} \left(\frac{q_{i-1}}{1-q_{i-1}}\right) & i = 1, \ldots, M \\ 0 & i = 0 \quad \text{or} \quad W(t_i^-) \leq 0 \end{cases} . \tag{5.4}$$

Then $W(t_i^+)$ is given by

$$W(t_i^+) = \left(S(t_i^-) + B(t_i^-)\right)\left(1 + \mathbb{T}_i^g\right) \exp(-(\Delta t)_i \mathbb{T}^f) - \mathfrak{q}_i \ ; \ i \in \mathcal{T} \ , \tag{5.5}$$

where we recall that $\mathfrak{q}_M \equiv 0$ and $(\Delta t)_0 \equiv 0$. With some abuse of notation, we define

$$W(t_i^-) = \left(S(t_i^-) + B(t_i^-)\right)\left(1 + \mathbb{T}_i^g\right) \exp(-(\Delta t)_i \mathbb{T}^f) \tag{5.6}$$

as the total portfolio value, after tontine gains and tontine fees, the instant before withdrawals and rebalancing at t_i.

Typically, DC plan savings are held in a tax-advantaged account, with no taxes triggered by rebalancing. With infrequent (e.g. yearly) rebalancing, we also expect other transaction costs, apart from the tontine fees, to be small, and hence can be ignored. It is possible to include transaction costs, but at the expense of increased computational cost (van Staden et al., 2018).

We denote the multi-dimensional controlled underlying process by $X(t) = (S(t), B(t))$, $t \in [0,T]$ and the realized state of the system by $x = (s, b)$. Let the rebalancing control $\mathfrak{p}_i(\cdot)$ be the fraction invested in the stock index at the rebalancing date t_i, i.e.

$$\mathfrak{p}_i\left(X(t_i^-)\right) = \mathfrak{p}\left(X(t_i^-), t_i\right) = \frac{S(t_i^+)}{S(t_i^+) + B(t_i^+)} \ . \tag{5.7}$$

[15] We are implicitly assuming here that the investor is solvent here and thus remains in the tontine pool, paying fees and receiving mortality credits.

Let the withdrawal control $\mathfrak{q}_i(\cdot)$ be the amount withdrawn at time t_i, i.e. $\mathfrak{q}_i\left(X(t_i^-)\right) = \mathfrak{q}\left(X(t_i^-), t_i\right)$. Formally, the controls depend on the state of the investment portfolio, before the rebalancing occurs, i.e. $\mathfrak{p}_i(\cdot) = \mathfrak{p}\left(X(t_i^-), t_i\right) = \mathfrak{p}\left(X_i^-, t_i\right)$, and $\mathfrak{q}_i(\cdot) = \mathfrak{q}\left(X(t_i^-), t_i\right) = \mathfrak{q}\left(X_i^-, t_i\right)$, $t_i \in \mathcal{T}$, where $\mathcal{T}$ is the set of rebalancing times.

However, it will be convenient to note that in our case, we find the optimal control $\mathfrak{p}_i(\cdot)$ amongst all strategies with constant wealth (after withdrawal of cash). Hence, with some abuse of notation, we will now consider $\mathfrak{p}_i(\cdot)$ to be function of wealth after withdrawal of cash

$$\begin{aligned}
W(t_i^-) &= \begin{cases} \Big(S(t_i^-) + B(t_i^-)\Big)\Big(1 + \mathbb{T}_i^g\Big)\exp\big(-(\Delta t)_i \mathbb{T}^f\big) \text{ if } \Big(S(t_i^-) + B(t_i^-)\Big) > 0 \\ \Big(S(t_i^-) + B(t_i^-)\Big) \text{ otherwise} \end{cases} \\
W(t_i^+) &= W(t_i^-) - \mathfrak{q}_i(\cdot) \\
\mathfrak{p}_i(\cdot) &= \mathfrak{p}(W(t_i^+), t_i) \\
S(t_i^+) &= S_i^+ = \mathfrak{p}_i(W_i^+)\ W_i^+ \\
B(t_i^+) &= B_i^+ = (1 - \mathfrak{p}_i(W_i^+))\ W_i^+ \ .
\end{aligned} \tag{5.8}$$

Note that the control for $\mathfrak{p}_i(\cdot)$ depends only W_i^+. Since $\mathfrak{p}_i(\cdot) = \mathfrak{p}_i(W_i^- - \mathfrak{q}_i)$, then it follows that

$$\mathfrak{q}_i(\cdot) = \mathfrak{q}_i(W_i^-) \ , \tag{5.9}$$

which we discuss further in Section 8.

A control at time t_i, is then given by the pair $(\mathfrak{q}_i(\cdot), \mathfrak{p}_i(\cdot))$ where the notation $(\cdot)$ denotes that the control is a function of the state. Let $\mathcal{Z}$ represent the set of admissible values of the controls $(\mathfrak{q}_i(\cdot), \mathfrak{p}_i(\cdot))$. We impose no-shorting, no-leverage constraints (assuming solvency). We also impose maximum and minimum values for the withdrawals. We apply the constraint that in the event of insolvency due to withdrawals ($W(t_i^+) < 0$), trading ceases and debt (negative wealth) accumulates at the appropriate borrowing rate of return (i.e. a spread over the bond rate). We also specify that the stock assets are liquidated at $t = t_M$.

More precisely, let W_i^+ be the wealth after withdrawal of cash, and W_i^- be the total wealth before withdrawals (but after fees and tontine cash flows), then define

$$\mathcal{Z}_{\mathfrak{q}}(W_i^-, t_i) = \begin{cases} [\mathfrak{q}_{\min}, \mathfrak{q}_{\max}] & t_i \in \mathcal{T} \ ; \ t_i \neq t_M \ ; \ W_i^- \geq \mathfrak{q}_{\max} \\ [\mathfrak{q}_{\min}, \max(\mathfrak{q}_{\min}, W_i^-)] & t_i \in \mathcal{T} \ ; \ t \neq t_M \ ; \ W_i^- < \mathfrak{q}_{\max} \\ \{0\} & t_i = t_M \end{cases} , \tag{5.10}$$

$$\mathcal{Z}_{\mathfrak{p}}(W_i^+, t_i) = \begin{cases} [0,1] & W_i^+ > 0 \ ; \ t_i \in \mathcal{T} \ ; \ t_i \neq t_M \\ \{0\} & W_i^+ \leq 0 \ ; \ t_i \in \mathcal{T} \ ; \ t_i \neq t_M \\ \{0\} & t_i = t_M \end{cases} . \tag{5.11}$$

The rather complicated expression in equation (5.10) imposes the assumption that as wealth becomes small, the retiree (i) tries to avoid insolvency as much as possible and (ii) in the event of insolvency, withdraws only $\mathfrak{q}_{\min}$.

The set of admissible values for $(\mathfrak{q}_i, \mathfrak{p}_i)$, $t_i \in \mathcal{T}$, can then be written as

$$(\mathfrak{q}_i, \mathfrak{p}_i) \in \mathcal{Z}(W_i^-, W_i^+, t_i) = \mathcal{Z}_{\mathfrak{q}}(W_i^-, t_i) \times \mathcal{Z}_{\mathfrak{p}}(W_i^+, t_i) \ . \tag{5.12}$$

For implementation purposes, we have written equation (5.12) in terms of the wealth after withdrawal of cash. However, we remind the reader that since $W_i^+ = W_i^- - \mathfrak{q}_i$, the controls are formally a function of the state $X(t_i^-)$ before the control is applied.

The admissible control set $\mathcal{A}$ can then be written as

$$\mathcal{A} = \left\{ (\mathfrak{q}_i, \mathfrak{p}_i)_{0 \le i \le M} : (\mathfrak{p}_i, \mathfrak{q}_i) \in \mathcal{Z}(W_i^-, W_i^+, t_i) \right\} . \tag{5.13}$$

An admissible control $\mathcal{P} \in \mathcal{A}$ can be written as

$$\mathcal{P} = \{(\mathfrak{q}_i(\cdot), \mathfrak{p}_i(\cdot)) \ : \ i = 0, \ldots, M\} . \tag{5.14}$$

We also define $\mathcal{P}_n \equiv \mathcal{P}_{t_n} \subset \mathcal{P}$ as the tail of the set of controls in $[t_n, t_{n+1}, \ldots, t_M]$, i.e.

$$\mathcal{P}_n = \{(\mathfrak{q}_n(\cdot), \mathfrak{p}_n(\cdot)), \ldots, (\mathfrak{q}_M(\cdot), \mathfrak{p}_M(\cdot))\} . \tag{5.15}$$

For notational completeness, we also define the tail of the admissible control set $\mathcal{A}_n$ as

$$\mathcal{A}_n = \left\{ (\mathfrak{q}_i, \mathfrak{p}_i)_{n \le i \le M} : (\mathfrak{q}_i, \mathfrak{p}_i) \in \mathcal{Z}(W_i^-, W_i^+, t_i) \right\} , \tag{5.16}$$

so that $\mathcal{P}_n \in \mathcal{A}_n$.

6 Risk and Reward

6.1 Risk: Definition of Expected Shortfall (ES)

Let $g(W_T)$ be the probability density function of wealth W_T at $t = T$. Suppose

$$\int_{-\infty}^{W_\alpha^*} g(W_T) \, dW_T = \alpha, \tag{6.1}$$

i.e. $Pr[W_T > W_\alpha^*] = 1 - \alpha$. We can interpret W_α^* as the Value at Risk (VAR) at level α. For example, if $\alpha = .05$, then 95% of the outcomes have $W_T > W_\alpha^*$. If W_α^* is sufficiently large and positive, this suggests very low risk of running out of savings.[16] The Expected Shortfall (ES) at level α is then

$$\text{ES}_\alpha = \frac{\int_{-\infty}^{W_\alpha^*} W_T \, g(W_T) \, dW_T}{\alpha} , \tag{6.2}$$

which is the mean of the worst α fraction of outcomes. Typically $\alpha \in \{.01, .05\}$. The definition of ES in equation (6.2) uses the probability density of the final wealth distribution, not the density of *loss*. Hence, in our case, a larger value of ES (i.e. a larger value of average worst case terminal wealth) is desired. The negative of ES is commonly referred to as Conditional Value at Risk (CVAR).

Define $X_0^+ = X(t_0^+), X_0^- = X(t_0^-)$. Given an expectation under control $\mathcal{P}$, $E_{\mathcal{P}}[\cdot]$, as noted by Rockafellar and Uryasev (2000), ES_α can be alternatively written as

$$\text{ES}_\alpha(X_0^-, t_0^-) = \sup_{W^*} E_{\mathcal{P}_0}^{X_0^+, t_0^+} \left[W^* + \frac{1}{\alpha} \min(W_T - W^*, 0) \right] . \tag{6.3}$$

The admissible set for W^* in equation (6.3) is over the set of possible values for W_T.

[16]In practice, the negative of W_α^* is often the reported VAR.

The notation $\mathrm{ES}_\alpha(X_0^-, t_0^-)$ emphasizes that ES_α is as seen at (X_0^-, t_0^-). In other words, this is the pre-commitment ES_α. A strategy based purely on optimizing the pre-commitment value of ES_α at time zero is *time-inconsistent*, hence has been termed by many as *non-implementable*, since the investor has an incentive to deviate from the time zero pre-commitment strategy at $t > 0$. However, in the following, we will consider the pre-commitment strategy merely as a device to determine an appropriate level of W^* in equation (6.3). If we fix W^* $\forall t > 0$, then this strategy is the induced time-consistent strategy (Strub et al., 2019; Forsyth, 2020a; Cui et al., 2022) and hence is implementable. We delay further discussion of this subtle point to Appendix A.

An alternative measure of risk could be based on variability of withdrawals (Forsyth et al., 2020). However, we note that we have constraints on the minimum and maximum withdrawals, so that variability is mitigated. We also assume that given these constraints, the retiree is primarily concerned with the risk of depleting savings, which is well measured by ES.

6.2 A Measure of Reward: Expected Total Withdrawals (EW)

We will use expected total withdrawals as a measure of reward in the following. More precisely, we define EW (expected withdrawals) as

$$\mathrm{EW}(X_0^-, t_0^-) = E_{\mathcal{P}_0}^{X_0^+, t_0^+}\left[\sum_{i=0}^{M} \mathfrak{q}_i\right] , \tag{6.4}$$

where we assume that the investor survives for the entire decumulation period, consistent with the Bengen (1994) scenario.

Note that there is no discounting term in equation (6.4) (recall that all quantities are real, i.e. inflation-adjusted). It is straightforward to introduce discounting, but we view setting the real discount rate to zero to be a reasonable and conservative choice. See Forsyth (2022b) for further comments.

7 Problem EW-ES

Since expected withdrawals (EW) and expected shortfall (ES) are conflicting measures, we use a scalarization technique to find the Pareto points for this multi-objective optimization problem. Informally, for a given scalarization parameter $\kappa > 0$, we seek to find the control $\mathcal{P}_0$ that maximizes

$$\mathrm{EW}(X_0^-, t_0^-) + \kappa \ \mathrm{ES}_\alpha(X_0^-, t_0^-) \ . \tag{7.1}$$

More precisely, we define the pre-commitment EW-ES problem $(PCES_{t_0}(\kappa))$ problem in terms of the value function $J(s,b,t_0^-)$

$$
\begin{aligned}
(PCES_{t_0}(\kappa)):\ J\left(s,b,t_0^-\right) = \sup_{\mathcal{P}_0 \in \mathcal{A}} \sup_{W^*} \Bigg\{ E_{\mathcal{P}_0}^{X_0^+,t_0^+} \Bigg[\sum_{i=0}^{M} \mathfrak{q}_i + \kappa \Bigg(W^* + \frac{1}{\alpha} \min(W_T - W^*, 0) \Bigg) \\
+ \epsilon W_T \Bigg| X(t_0^-) = (s,b) \Bigg] \Bigg\}
\end{aligned} \tag{7.2}
$$

$$
\text{subject to} \begin{cases}
(S_t, B_t) \text{ follow processes (4.3) and (4.4)}; \quad t \notin \mathcal{T} \\
W_\ell^+ = W_\ell^- - \mathfrak{q}_\ell\,;\ X_\ell^+ = (S_\ell^+, B_\ell^+) \\
W_\ell^- = \Big(S(t_i^-) + B(t_i^-) \Big) \Big(1 + \mathbb{T}_i^g \Big) \exp(-(\Delta t)_i \mathbb{T}^f) \\
S_\ell^+ = \mathfrak{p}_\ell(\cdot) W_\ell^+\,;\ B_\ell^+ = (1 - \mathfrak{p}_\ell(\cdot)) W_\ell^+ \\
(\mathfrak{q}_\ell(\cdot), \mathfrak{p}_\ell(\cdot)) \in \mathcal{Z}(W_\ell^-, W_\ell^+, t_\ell) \\
\ell = 0, \ldots, M\ ;\ t_\ell \in \mathcal{T}
\end{cases} . \tag{7.3}
$$

Note that we have added the extra term $E_{\mathcal{P}_0}^{X_0^+,t_0^+}[\epsilon W_T]$ to equation (7.2). If we have a maximum withdrawal constraint, and if $W_t \gg W^*$ as $t \to T$, the controls become ill-posed. In this fortunate state for the investor, we can break investment policy ties either by setting $\epsilon < 0$, which will force investments in bonds, or by setting $\epsilon > 0$, which will force investments into stocks. Choosing $|\epsilon| \ll 1$ ensures that this term only has an effect if $W_t \gg W^*$ and $t \to T$. See Forsyth (2022b) for more discussion of this.

Interchange the $\sup \sup(\cdot)$ in equation (7.2), so that value function $J\left(s,b,t_0^-\right)$ can be written as

$$
\begin{aligned}
J\left(s,b,t_0^-\right) = \sup_{W^*} \sup_{\mathcal{P}_0 \in \mathcal{A}} \Bigg\{ E_{\mathcal{P}_0}^{X_0^+,t_0^+} \Bigg[\sum_{i=0}^{M} \mathfrak{q}_i + \kappa \Bigg(W^* + \frac{1}{\alpha} \min(W_T - W^*, 0) \Bigg) \\
+ \epsilon W_T \Bigg| X(t_0^-) = (s,b) \Bigg] \Bigg\} .
\end{aligned} \tag{7.4}
$$

Noting that the inner supremum in equation (7.4) is a continuous function of W^* and that the optimal value of W^* in equation (7.4) is bounded[17], then define

$$
\begin{aligned}
\mathcal{W}^*(s,b) = \arg\max_{W^*} \Bigg\{ \sup_{\mathcal{P}_0 \in \mathcal{A}} \Bigg\{ E_{\mathcal{P}_0}^{X_0^+,t_0^+} \Bigg[\sum_{i=0}^{M} \mathfrak{q}_i \ + \ \kappa \Bigg(W^* + \frac{1}{\alpha} \min(W_T - W^*, 0) \Bigg) \\
+ \epsilon W_T \Bigg| X(t_0^-) = (s,b) \Bigg] \Bigg\} .
\end{aligned} \tag{7.5}
$$

See Forsyth (2020a) for an extensive discussion concerning pre-commitment and time consistent ES strategies. We summarize the relevant results from that discussion in Appendix A.

8 Formulation as a Dynamic Program

We use the method in Forsyth (2020a) to solve problem We (7.4). write equation (7.4) as

$$
J(s,b,t_0^-) = \sup_{W^*} V(s,b,W^*,0^-) \ , \tag{8.1}
$$

[17] This is the same as noting that a finite value at risk exists. Assuming $0 < \alpha < 1$, this is easily shown since our investment strategy uses no leverage and no-shorting.

where the auxiliary function $V(s, b, W^*, t)$ is defined as

$$V(s,b, W^*, t_n^-) = \sup_{\mathcal{P}_n \in \mathcal{A}_n} \Bigg\{ E_{\mathcal{P}_n}^{\hat{X}_n^+, t_n^+} \Bigg[\sum_{i=n}^{M} \mathfrak{q}_i + \kappa \Bigg(W^* + \frac{1}{\alpha} \min((W_T - W^*),0) \Bigg) + \ \epsilon W_T \Bigg| \hat{X}(t_n^-) = (s,b, W^*) \Bigg] \Bigg\} \ . \tag{8.2}$$

$$\text{subject to} \begin{cases} (S_t, B_t) \text{ follow processes (4.3) and (4.4);} \quad t \notin \mathcal{T} \\ W_\ell^+ = W_\ell^- - \mathfrak{q}_\ell \ ; \ X_\ell^+ = (S_\ell^+, B_\ell^+, W^*) \\ W_\ell^- = \Big(S(t_i^-) + B(t_i^-) \Big) \Big(1 + \mathbb{T}_i^g \Big) \exp(-(\Delta t)_i \mathbb{T}^f) \\ S_\ell^+ = \mathfrak{p}_\ell(\cdot) W_\ell^+ \ ; \ B_\ell^+ = (1 - \mathfrak{p}_\ell(\cdot)) W_\ell^+ \\ (\mathfrak{q}_\ell(\cdot), \mathfrak{p}_\ell(\cdot)) \in \mathcal{Z}(W_\ell^-, W_\ell^+, t_\ell) \\ \ell = n, \ldots, M \ ; \ t_\ell \in \mathcal{T} \end{cases} . \tag{8.3}$$

We have now decomposed the original problem (7.4) into two steps:

- For given initial cash W_0, and a fixed value of W^*, solve problem (8.2) using dynamic programming to determine $V(0,W_0, W^*, 0^-)$.
- Solve problem (7.4) by maximizing over W^*

$$J(0,W_0, 0^-) = \sup_{W^*} V(0,W_0, W^*, 0^-) \ . \tag{8.4}$$

8.1 Dynamic Programming Solution of Problem (8.2)

We give a brief overview of the method described in detail in (Forsyth, 2022b). Apply the dynamic programming principle to $t_n \in \mathcal{T}$

$$\begin{aligned} V(s,b,W^*, t_n^-) &= \sup_{\mathfrak{q} \in \mathcal{Z}_\mathfrak{q}(w^-, t_n)} \Bigg\{ \sup_{\mathfrak{p} \in \mathcal{Z}_\mathfrak{p}(w^- - q, t_n)} \Big[\mathfrak{q} + V((w^- - \mathfrak{q})\mathfrak{p}, (w^- - \mathfrak{q})(1 - \mathfrak{p}), W^*, t_n^+) \Big] \Bigg\} \\ &= \sup_{\mathfrak{q} \in Z_\mathfrak{q}(w^-, t_n)} \Bigg\{ \mathfrak{q} + \Big[\sup_{\mathfrak{p} \in \mathcal{Z}_\mathfrak{p}(w^- - q, t_n)} V((w^- - \mathfrak{q})\mathfrak{p}, (w^- - \mathfrak{q})(1 - \mathfrak{p}), W^*, t_n^+) \Big] \Bigg\} \\ w^- &= (s + b) \Big(1 + \mathbb{T}_i^g \Big) \exp(-(\Delta t)_i \mathbb{T}^f) \ . \end{aligned} \tag{8.5}$$

For computational purposes, we define

$$\tilde{V}(w, t_n, W^*) = \Big[\sup_{\mathfrak{p} \in \mathcal{Z}_\mathfrak{p}(w, t_n)} V(w\mathfrak{p}, w(1 - \mathfrak{p}), W^*, t_n^+) \Big] \ . \tag{8.6}$$

Equation (8.5) now becomes

$$\begin{aligned} V(s,b,W^*, t_n^-) &= \sup_{\mathfrak{q} \in \mathcal{Z}_\mathfrak{q}(w^-, t_n)} \Bigg\{ \mathfrak{q} + \Big[\tilde{V}((w^- - \mathfrak{q}), W^*, t_n^+) \Big] \Bigg\} \\ w^- &= (s + b) \Big(1 + \mathbb{T}_i^g \Big) \exp(-(\Delta t)_i \mathbb{T}^f) \ . \end{aligned} \tag{8.7}$$

This approach effectively replaces a two dimensional optimization for $(\mathfrak{q}_n, \mathfrak{p}_n)$, to two sequential one dimensional optimizations. From equations (8.6-8.7), it is clear that the optimal pair $(\mathfrak{q}_n, \mathfrak{p}_n)$ is such that

$$
\begin{aligned}
\mathfrak{q}_n &= \mathfrak{q}_n(w^-, W^*) \\
&\quad w^- = (s+b)\left(1 + \mathbb{T}_i^g\right) \exp(-(\Delta t)_i \mathbb{T}^f) \\
\mathfrak{p}_n &= \mathfrak{p}_n(w, W^*) \\
&\quad w = w^- - \mathfrak{q}_n \ .
\end{aligned}
\tag{8.8}
$$

In other words, the optimal withdrawal control $\mathfrak{q}_n$ is only a function of total wealth (after tontine gains and fees) before withdrawals. The optimal control $\mathfrak{p}_n$ is a function only of total wealth after withdrawals, tontine gains, and fees.

At $t = T$, we have

$$
V(s, b, W^*, T^+) = \kappa\left(W^* + \frac{\min((s+b-W^*), 0)}{\alpha}\right) + \epsilon(s+b) \ . \tag{8.9}
$$

At points in between rebalancing times, i.e. $t \notin \mathcal{T}$, the usual arguments (from SDEs (4.3-4.4), and Forsyth (2022b)) give

$$
\begin{aligned}
V_t &+ \frac{(\sigma^s)^2 s^2}{2} V_{ss} + (\mu^s - \lambda_\xi^s \gamma_\xi^s) s V_s + \lambda_\xi^s \int_{-\infty}^{+\infty} V(e^y s, b, t) f^s(y)\ dy + \frac{(\sigma^b)^2 b^2}{2} V_{bb} \\
&+ (\mu^b + \mu_c^b \mathbf{1}_{\{b<0\}} - \lambda_\xi^b \gamma_\xi^b) b V_b + \lambda_\xi^b \int_{-\infty}^{+\infty} V(s, e^y b, t) f^b(y)\ dy - (\lambda_\xi^s + \lambda_\xi^b) V + \rho_{sb} \sigma^s \sigma^b s b V_{sb} = 0 \ , \\
&\qquad s \geq 0 \ ; b \geq 0 \ .
\end{aligned}
\tag{8.10}
$$

In case of insolvency[18] $s = 0, b < 0$ and then

$$
\begin{aligned}
V_t &+ \frac{(\sigma^b)^2 b^2}{2} V_{bb} + (\mu^b + \mu_c^b \mathbf{1}_{\{b<0\}} - \lambda_\xi^b \gamma_\xi^b) b V_b + \lambda_\xi^b \int_{-\infty}^{+\infty} V(0, e^y b, t) f^b(y)\ dy - \lambda_\xi^b V = 0 \ , \\
&\qquad s = 0 \ ; b < 0 \ .
\end{aligned}
\tag{8.11}
$$

It will be convenient for computational purposes to re-write equation (8.11) in terms of debt $\hat{b} = -b$ when $b < 0$. Now let $\hat{V}(\hat{b}, t) = V(0, b, t), b < 0, b = -\hat{b}$ in equation (8.11) to give

$$
\begin{aligned}
\hat{V}_t &+ \frac{(\sigma^b)^2 \hat{b}^2}{2} \hat{V}_{\hat{b}\hat{b}} + (\mu^b + \mu_c^b - \lambda_\xi^b \gamma_\xi^b) \hat{b} \hat{V}_{\hat{b}} + \lambda_\xi^b \int_{-\infty}^{+\infty} \hat{V}(e^y \hat{b}, t) f^b(y)\ dy - \lambda_\xi^b \hat{V} = 0 \ , \\
&\qquad s = 0 \ ; \ b < 0 \ ; \ \hat{b} = -b \ .
\end{aligned}
\tag{8.12}
$$

Note that equation (8.12) is now amenable to a transformation of the form $\hat{x} = \log \hat{b}$ since $\hat{b} > 0$, which is required when using a Fourier method (Forsyth and Labahn, 2019; Forsyth, 2022b) to solve equation (8.12).

After rebalancing, if $b \geq 0$, then b cannot become negative, since $b = 0$ is a barrier in equation (8.11). However, b can become negative after withdrawals, in which case b remains in the state

[18] Insolvency can only occur due to the minimum withdrawals specified.

$b < 0$, where equation (8.12) applies, unless there is an injection of cash to move to a state with $b > 0$. The terminal condition for equation (8.12) is

$$\hat{V}(\hat{b}, W^*, T^+) = \kappa\left(W^* + \frac{\min((-\hat{b} - W^*), 0)}{\alpha}\right) + \epsilon(-\hat{b}) \ ; \ \hat{b} > 0 \ . \tag{8.13}$$

A brief overview of the numerical algorithms is given in Appendix B, along with a numerical convergence verification.

9 Data

We use data from the Center for Research in Security Prices (CRSP) on a monthly basis over the 1926:1-2020:12 period.[19] Our base case tests use the CRSP US 30 day T-bill for the bond asset and the CRSP value-weighted total return index for the stock asset. This latter index includes all distributions for all domestic stocks trading on major U.S. exchanges. All of these various indexes are in nominal terms, so we adjust them for inflation by using the U.S. CPI index, also supplied by CRSP. We use real indexes since investors funding retirement spending should be focused on real (not nominal) wealth goals.

We use the threshold technique (Mancini, 2009; Cont and Mancini, 2011; Dang and Forsyth, 2016) to estimate the parameters for the parametric stochastic process models. Since the index data is in real terms, all parameters reflect real returns. Table 9.1 shows the results of calibrating the models to the historical data. The correlation ρ_{sb} is computed by removing any returns which occur at times corresponding to jumps in either series, and then using the sample covariance. Further discussion of the validity of assuming that the stock and bond jumps are independent is given in Forsyth (2020b).

Remark 9.1 (Jump diffusion for 30-day T-bill)**.** *In MacMinn et al. (2014), it is assumed that the corporate constant maturity bond index follows a jump diffusion process, while the three month T-bill index follows a pure diffusion. However, in Forsyth et al. (2022), use of the filtering algorithm (Cont and Mancini, 2011) actually identifies more jumps in the 30 day T-bill index than are observed in the stock index. Furthermore, the empirical return histograms in Forsyth et al. (2022) show the higher peaks and fatter tails characteristic of a jump diffusion. Note that in our case, in contrast to MacMinn et al. (2014), we use real (adjusted for inflation) time series, which may cause greater non-normality of returns. The filtering test described in Forsyth et al. (2022) applied to the inflation adjusted T-bill data over 1926:1-2020:12 (1140 monthly data points) shows 47 events which exceed three standard deviations from the mean. Assuming normality, we would expect to observe at most 4 such events.*

Remark 9.2 (Choice of 30-day T-bill for the bond index)**.** *It might be argued that the bond index should hold longer-dated bonds such as 10-year Treasuries since this would allow the investor to harvest the term premium. Long-term bonds have enjoyed high real returns over the last 30 years*

[19] More specifically, results presented here were calculated based on data from Historical Indexes, ©2020 Center for Research in Security Prices (CRSP), The University of Chicago Booth School of Business. Wharton Research Data Services (WRDS) was used in preparing this article. This service and the data available thereon constitute valuable intellectual property and trade secrets of WRDS and/or its third-party suppliers.

CRSP	μ^s	σ^s	λ^s	u^s	η_1^s	η_2^s	ρ_{sb}
	0.08912	0.1460	0.3263	0.2258	4.3625	5.5335	0.08420
30-day T-bill	μ^b	σ^b	λ^b	u^b	η_1^b	η_2^b	ρ_{sb}
	0.0046	0.0130	0.5053	0.3958	65.801	57.793	0.08420

TABLE 9.1: *Estimated annualized parameters for double exponential jump diffusion model. Value-weighted CRSP index, 30-day T-bill index deflated by the CPI. Sample period 1926:1 to 2020:12.*

due to decreasing real interest rates during that period. However, it is unlikely that this will continue to be true over the next 30 years. Hatch and White (1985) study the real returns of equities, short-term T-bills, and long-term corporate and government bonds, over the period 1950-1983 and conclude that, in both Canada and the US, only equities and short-term T-bills had non-negative real returns. Inflation (both US and Canada) averaged about 4.75% per year over the period 1950-1983. If one imagines that the next 30 years will be a period with inflationary pressures, this suggests that the defensive asset should be short-term T-bills. However, there is nothing in our methodology that prevents us from using other underlying bonds in the bond index. We emphasize that we are considering inflation-adjusted returns here, and that the historical real return of short-term T-bills over 1926:1-2020:12 is approximately zero. Hence our use of T-bills as the defensive asset is a conservative approach going forward.

Remark 9.3 (Sensitivity to Calibrated Parameters)**.** *It might be argued that the stochastic processes (4.3-4.4) are simplistic, and perhaps inappropriate. However, we will test the optimal strategies (computed assuming processes (4.3-4.4) with calibrated parameters in Table 9.1) using bootstrap resampled historical data (see Section 10 below). The computed strategy seems surprisingly robust to model misspecification. Similar results have been noted for the case of multi-period mean-variance controls (van Staden et al., 2021).*

10 Historical Market

We compute and store the optimal controls based on the parametric model (4.3-4.4) as for the synthetic market case. However, we compute statistical quantities using the stored controls, but using bootstrapped historical return data directly. In this case, we make no assumptions concerning the stochastic processes followed by the stock and bond indices. We remind the reader that all returns are inflation-adjusted. We use the stationary block bootstrap method (Politis and Romano, 1994; Politis and White, 2004; Patton et al., 2009; Cogneau and Zakalmouline, 2013; Dichtl et al., 2016; Cavaglia et al., 2022; Simonian and Martirosyan, 2022; Anarkulova et al., 2022). A key parameter is the expected blocksize. Sampling the data in blocks accounts for serial correlation in the data series. We use the algorithm in Patton et al. (2009) to determine the optimal blocksize for the bond and stock returns separately, see Table 10.1. We use a paired sampling approach to simultaneously draw returns from both time series. In this case, a reasonable estimate for the

Data series	Optimal expected block size $\hat{b}$ (months)
Real 30-day T-bill index	50.6
Real CRSP value-weighted index	3.42

TABLE 10.1: *Optimal expected blocksize $\hat{b} = 1/v$ when the blocksize follows a geometric distribution $Pr(b = k) = (1 - v)^{k-1}v$. The algorithm in Patton et al. (2009) is used to determine $\hat{b}$. Historical data range 1926:1-2020:12.*

blocksize for the paired resampling algorithm would be about 2.0 years. We will give results for a range of blocksizes as a check on the robustness of the bootstrap results. Detailed pseudo-code for block bootstrap resampling is given in Forsyth and Vetzal (2019).

11 Investment Scenario

Table 11.1 shows our base case investment scenario. We use thousands of dollars as our units of wealth. For example, a withdrawal of 40 per year corresponds to \$40,000 per year (all values are real, i.e. inflation-adjusted), with an initial wealth of 1000 (i.e. \$1,000,000). This would correspond to the use of the four per cent rule (Bengen, 1994). Our base case scenario assumes a fee of 50 bps per year for the tontine overlay. See Chen et al. (2021) for a discussion of tontine fees.

As a motivating example, we consider a 65-year old Canadian retiree with a pre-retirement salary of \$100,000 per year, with \$1,000,000 in a DC savings account. Government benefits are assumed to amount to about \$20,000 per year (real). The retiree needs the DC plan to generate at least \$40,000 per year (real), so that the DC plan and government benefits together replace 60% of pre-retirement income. We assume that the retiree owns mortgage-free real estate worth about \$400,000. In an act of mental accounting, the retiree plans to use the real estate as a longevity hedge, which could be monetized using a reverse mortgage. In the event that the longevity hedge is not needed, the real-estate will be a bequest. Of course, the retiree would like to withdraw more than \$40,000 per year from the DC plan, but has no use for withdrawals greater than \$80,000 per year. We further assume that the real estate holdings can generate \$200,000 through a reverse mortgage. Hence, as a rough rule of thumb any expected shortfall at $T = 30$ years greater than $-\$200{,}000$ is an acceptable level of risk.

Our view that personal real estate is not fungible with investment assets (unless investment assets are depleted) is consistent with the behavioral life cycle approach originally described in Shefrin and Thaler (1988) and Thaler (1990). In this framework, investors divide their wealth into separate "mental accounts" containing funds intended for different purposes such as current spending or future need.

We take the view of a 65-year old retiree, who wants to maximize her total withdrawals and minimize the risk of running out of savings, assuming that she lives to the age of 95. We also assume that the retiree has no bequest motive.

Recall that Bengen (1994) attempted to determine a safe real withdrawal rate, and constant allocation strategy, such that the probability of running out of cash after 30 years of retirement was small. In other words, Bengen (1994) maximized total withdrawals over a 30 year period, assuming that the retiree survived for the entire 30 years. This is, of course, a conservative assumption.

In our case, we are essentially answering the same question. The key difference here is that we allow for (i) dynamic asset allocation, (ii) variable withdrawals (within limits) and (iii) a possible tontine overlay.

Retiree	65-year old Canadian male
Tontine Gain $\mathbb{T}^g$	equation (3.18)
Group Gain G (see equation (3.16))	1.0
Mortality table	CPM 2014
Investment horizon T (years)	30.0
Equity market index	CRSP Cap-weighted index (real)
Bond index	30-day T-bill (US) (real)
Initial portfolio value W_0	1000
Cash withdrawal/rebalancing times	$t = 0,1.0, 2.0, \ldots, 29.0$
Maximum withdrawal (per year)	$q_{\max} = 80$
Minimum withdrawal (per year)	$q_{\min} = 40$
Equity fraction range	[0,1]
Borrowing spread μ_c^b	0.02
Rebalancing interval (years)	1.0
α (EW-ES)	.05
Fees $\mathbb{T}^f$ (see equation (5.5))	50 bps per year
Stabilization ϵ (see equation (7.2))	-10^{-4}
Market parameters	See Table 9.1

TABLE 11.1: *Input data for examples. Monetary units: thousands of dollars. CPM2014 is the mortality table from the Canadian Institute of Actuaries.*

12 Constant Withdrawal, Constant Equity Fraction

As a preliminary example, in this section we present results for the scenario in Table 11.1, except that a constant withdrawal of 40 per year is specified, along with a constant weight in stocks at each rebalancing date.

Table 12.1 gives the results for various values of the constant weight equity fraction in the synthetic market. The best result[20] for ES (the largest value) occurs at the rather low constant equity weight of $p = 0.1$, with ES $= -239$. Table 12.2 gives similar results, this time using bootstrap resampling of the historical data (the historical market). Here the best value of ES $= -305$ occurs for a constant equity fraction of $p = 0.4$. Consequently, in both the historical and synthetic market, the constant weight, constant withdrawal strategy fails to meet our risk criteria of ES > -200.

These simulations indicate that there is significant depletion risk for the constant withdrawal, constant weight strategy suggested in Bengen (1994).

13 Synthetic Market Efficient Frontiers

Figure 13.1 shows the efficient EW-ES frontiers computed in the synthetic market for the following cases:

Tontine: the case in Table 11.1. The control is computed using the algorithm in Section 8 and then stored, and used in Monte Carlo simulations. The detailed frontier is given in Table D.1.

[20] Recall that ES is defined in terms of the left tail mean of final wealth (not losses) hence a larger value is preferred.

Equity fraction p	$E[\sum_i \mathfrak{q}_i]/T$	ES (5%)	$Median[W_T]$
0.0	40	-302.57	-150.56
0.1	40	-238.62	-6.82
0.2	40	-245.48	168.10
0.3	40	-280.27	386.05
0.4	40	-330.37	649.58
0.5	40	-391.61	958.33
0.6	40	-461.54	1312.17
0.7	40	-538.04	1706.49
0.8	40	-619.31	2135.24

Table 12.1: *Constant weight, constant withdrawals, synthetic market results. No tontine gains. Stock index: real capitalization weighted CRSP stocks; bond index: real 30-day T-bills. Parameters from Table 9.1. Scenario in Table 11.1. Units: thousands of dollars. Statistics based on* 2.56×10^6 *Monte Carlo simulation runs.* $T = 30$ *years.*

Equity fraction p	$E[\sum_i \mathfrak{q}_i]/T$	ES (5%)	$Median[W_T]$
0.0	40	-508.44	-155.04
0.1	40	-418.02	-10.98
0.2	40	-350.00	164.75
0.3	40	-312.24	382.16
0.4	40	-305.52	649.04
0.5	40	-326.40	966.61
0.6	40	-370.18	1336.31
0.7	40	-432.55	1759.66
0.8	40	-509.00	2232.29

Table 12.2: *Constant weight, constant withdrawals, historical market. No tontine gains. Historical data range 1926:1-2020:12. Constant withdrawals are 40 per year. Stock index: real capitalization weighted CRSP stocks; bond index: real 30-day T-bills. Scenario in Table 11.1. Units: thousands of dollars. Statistics based on* 10^6 *bootstrap simulation runs. Expected blocksize = 2 years.* $T = 30$ *years.*

No Tontine: the case in Table 11.1, but without any tontine gains The control is computed and stored, and then used in Monte Carlo simulations The detailed frontier is given in Table D.2.

Const q=40, Const p: The best single point from Table 12.1, based on Monte Carlo simulations.

Note that all these strategies produce a minimum withdrawal of 40 per year (i.e. 4% real of the initial investment) for 30 years. However, the best result for the constant weight strategies was (EW,ES) = $(40, -239)$ This can be improved significantly by optimizing over withdrawals and asset allocation, but with no tontine gains. For example, from Table D.2, the nearest point with roughly the same level of risk is (EW,ES) = $(58, -242)$. However, the improvement with optimal controls and tontine gains is remarkable. For example, it seems reasonable to target a value of $ES \simeq 0$. From Table D.1, we note the point (EW,ES) = (69,47), which is dramatically better than any No Tontine Pareto point. This can also be seen from the large outperformance in the EW-ES frontier compared to the No Tontine case in Figure 13.1 .

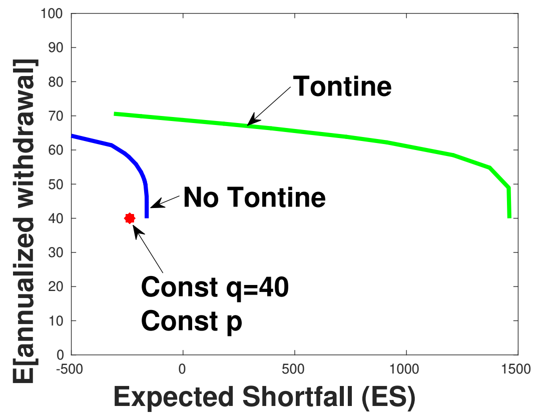


FIGURE 13.1: *Frontiers generated from the synthetic market. Parameters based on real CRSP index, real 30-day T-bills (see Table 9.1). Tontine case is as in Table 11.1. The No Tontine case uses the same scenario, but with no tontine gains, and no fees. The Const q, Const p case has $q = 40$, $p = 0.10$, with no tontine gains, which is the best result from Table 12.1, assuming no tontine gains, and no fees. Units: thousands of dollars.*

13.1 Effect of Fees

Figure 13.2 shows the effect of varying the annual fee in the synthetic market, for the scenario in Table 11.1. Recall that the base case specified a fee of 50 bps per year. Assuming a shortfall target of ES $\simeq 0$, then the effect of fees in the range $0 - 100$ bps is quite modest. Even with annual fees of 100 bps, the Tontine case still significantly outperforms the No Tontine case (which is assumed to have no fees).

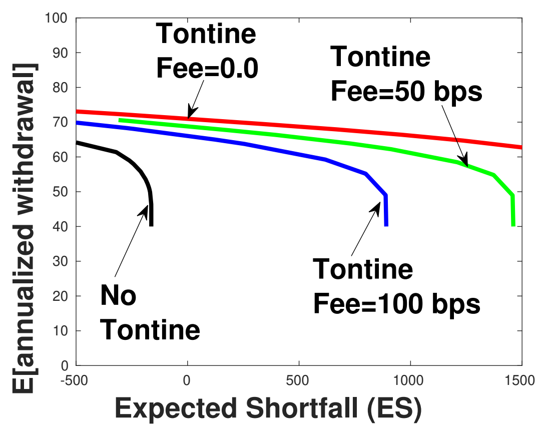


FIGURE 13.2: *Effect of varying fees charged for the Tontine, basis points (bps) per year. Frontiers generated from the synthetic market. Parameters based on real CRSP index, real 30-day T-bills (see Table 9.1). Base case Tontine is as in Table 11.1 (fees 50 bps per year). The No Tontine case uses the same scenario, but with no tontine gains, and no fees. Units: thousands of dollars.*

13.2 Effect of Random G

Recall the definition of the group gain G_i at time t_i in equation (3.10). Basically, the group gain is used to ensure that the total amount of mortality credits disbursed is exactly equal to the total amount forfeited by tontine participants who have died in (t_{i-1}, t_i).

If Condition 3.1 holds, then we expect that randomly varying G_i will have a small cumulative effect. In Fullmer and Sabin (2019) and Winter and Planchet (2022), the authors create synthetic tontine pools where the investors have different initial wealth, ages, genders, and investment strategies. These pools are perpetual, i.e. new members join as original members die. It is assumed that the investors can only select an asset allocation strategy from a stock index and a bond index, both of which follow a geometric Brownian motion (GBM).

The payout rules are different from those suggested in this paper, but it is instructive to observe the following. In Fullmer and Sabin (2019), the perpetual tontine pool has 15,000 investors in steady-state. After the initial start-up period, the expected value of the group gain G_i at each rebalancing time is close to unity, with a standard deviation of about 0.1. Fullmer and Sabin (2019) also note that there is essentially no correlation between investment returns and the group gain.

Further simulations were carried out in Winter and Planchet (2022), using the same sharing rule as in Fullmer and Sabin (2019) for a single period (one year), with $500 - 1{,}000 - 5{,}000$ participants. The investors had different allocation amounts with ages from $40 - 70$, but all participants had the same investment strategy. The variance of G_i was negligible for the pool with 5,000 initial investors. The Fullmer and Sabin (2019) study, with the additional variability of random asset allocation to risky assets, had a low standard deviation for G_i at around 15,000 participants. Consequently, it would appear that the number of participants required to be reasonably sure that the assumption that $var(G_i)$ is small is in the range of 5,000-15,000, depending on restrictions for individual asset allocation.

Figure 13.3 shows the effect of randomly varying G_i. The curve labeled $G = 1.0$ is the base case EW-ES curve from the scenario in Table 11.1, in the synthetic market (parameters in Table 9.1). The controls from this base case are stored, and then used in Monte Carlo simulations, where G is assumed to have a normal distribution with mean one, and standard deviation of 0.1. The EW-ES curves for both cases essentially overlap, except for very large values of ES, which are not of any practical interest. We get essentially the same result if we use a uniform distribution for G with $E[G] = 1$, with the same standard deviation. This is not surprising, since, assuming that the value function is smooth, then a simple Taylor series argument shows that, for any assumed distribution of G with mean one, the effect of randomness of G is a second order effect in the standard deviation.

Of course, we cannot determine the actual distribution of G without a detailed knowledge of the characteristics of the tontine pool. In fact, if we knew the distribution, we could include it in the formulation of the optimal control problem. However, knowledge of the distribution of G is unlikely to be available to pool participants in practice.

Nevertheless, the simulations in Fullmer and Sabin (2019) and Winter and Planchet (2022), coupled with our results as shown in Figure 13.3, suggest that for a sufficiently large, diversified pool of investors the effects of randomly varying G are negligible. Note that we are only considering idiosyncratic mortality risk here, not systematic risk, e.g. unexpected mortality improvement.

14 Bootstrapped Results

As discussed in Section 10, a key parameter in the stationary block bootstrap technique is the expected blocksize. In Figure 14.1(a), we show the results of the following test. We compute and store the optimal controls, based on the synthetic market. Then we use these controls, but carry out tests on bootstrapped historical data. The efficient frontiers in Figure 14.1(a), for ES < 1000 essentially overlap for all expected blocksizes in the range $0.5 - 5.0$ years. Since it is probably not of interest to aim for an ES of 1000 (which is one million dollars) at age 95, this indicates that the

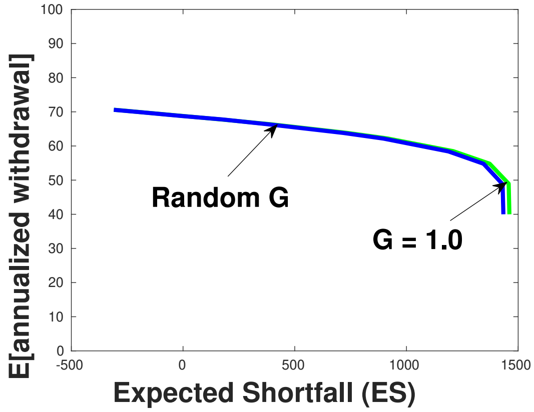


FIGURE 13.3: *Effect of randomly varying group gain G (Section 3.2.1). Frontiers generated from the synthetic market. Parameters based on real CRSP index, real 30-day T-bills (see Table 9.1). Base case Tontine* ($G = 1.0$) *is as in Table 11.1. Random G case uses the control computed for the base case, but in the Monte Carlo simulation, G is normally distributed with mean one and standard deviation* 0.1*. Units: thousands of dollars.*

computed strategy is robust to parameter uncertainty.

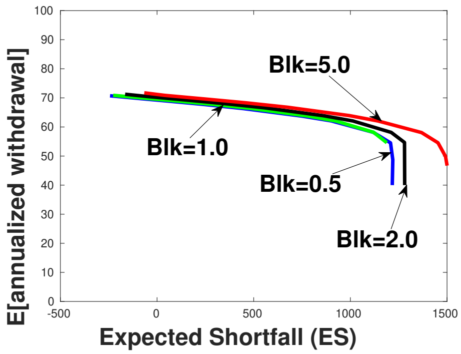
Figure 14.1(b) compares the efficient frontier tested in the historical market (expected blocksize 2 years), with the efficient frontier in the synthetic market. We observe that the synthetic and historical curves overlap for ES < 1000, which again verifies that the controls are robust to data uncertainty. The efficient frontiers/points for the No Tontine case and the constant weight, constant withdrawal strategy (computed in the historical market) are also shown. The Tontine overlay continues to outperform the No Tontine case by a wide margin.

15 Detailed Historical Market Results: EW-ES Controls

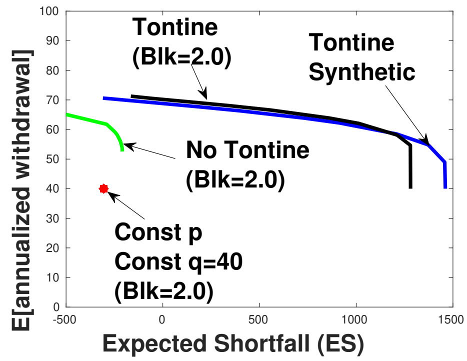
In this section, we examine some detailed characteristics of the optimal EW-ES strategy, tested in the historical market for the scenario in Table 11.1. Figure 15.1 shows the percentiles of fraction in stocks, wealth, and withdrawals versus time for the EW-ES control with $\kappa = 0.18$, with (EW,ES) $=$ $(69, 204)$. To put this in perspective, recall that this strategy never withdraws less than 40 per year. Compare this to the best case for a constant withdrawal, constant weight strategy (no tontine) from Table 12.2, which has (EW,ES) $= (40, -306)$, or to the optimal EW-ES strategy, but with no tontine, from Table E.2, which has (EW,ES) $= (70, -806)$.

Figure 15.1(a) shows that the median optimal fraction in stocks starts at about 0.60, then drops to about 0.20 at 15 years, finally ending up at zero in year 26. Figure 15.1(b) indicates that for the years in the span of $20 - 30$, the median and fifth percentiles of wealth are fairly tightly clustered, with the fifth percentile being well above zero at all times. The optimal withdrawal percentiles are shown in Figure 15.1(c). The median withdrawal starts at 40 per year, then increases to the maximum withdrawal of 80 in years $3 - 4$, and remains at 80 for the remainder of the 30-year time horizon.

Figure 15.2 shows the optimal control heat maps for the fraction in stocks and withdrawal amounts, for the scenario in Table 11.1. Figure 15.2(a) shows a smooth behavior of the optimal fraction in stocks as a function of (W,t). This can be compared with the equivalent heat map for the EW-ES control in Forsyth (2022b) (no tontine gains), which is much more aggressive at changing the asset allocation in response to changing wealth amounts. The smoothness of the controls in



(a) *Efficient frontiers, historical market, effect of varying expected blocksize (Blk).*



(b) *Efficient frontiers, historical market, expected blocksize* = 2.0 *years. Synthetic market frontier (tontine) also shown.*

FIGURE 14.1: *Optimal strategy determined by solving Problem 7.2 in the synthetic market, parameters in Table 9.1. Control stored and then tested in bootstrapped historical market. Inflation adjusted data, 1926:1-2020:12. Non-Pareto points eliminated. Expected blocksize (Blk, years) used in the bootstrap resampling method also shown. Units: thousands of dollars. The const q, const p case had* $(p,q) = (0.4, 40)$ *(no tontine gains). This is the best result for the constant (p,q) case, shown in Table 12.2.*

Figure 15.2(a) appears to be due to the rapid de-risking of a strategy which includes tontine gains, which provides a natural protection against sudden stock index drops. The upper blue zone in Figure 15.2(a) is de-risking due to the fact that, with sufficiently large wealth, there is essentially no probability of running out of cash even at the maximum withdrawal amount. The use of the stabilization factor $\epsilon = -10^{-4}$ forces the strategy to increase the weight in bonds for large values of wealth (see equation (7.2)).[21] The lower red zone is in response to extremely poor wealth outcomes, which means that the optimal strategy is to invest 100% in stocks and hope for the best. However, this is an extremely unlikely outcome, as can be verified from Figure 15.1(b).

From Figure 15.2(b), we can observe that the optimal withdrawal strategy is essentially a bang-bang control, i.e. withdraw at either the maximum or minimum amount per year. This is not unexpected, as discussed in Appendix C. We also note that this type of strategy has been suggested previously, based on heuristic reasoning.[22]

16 Discussion

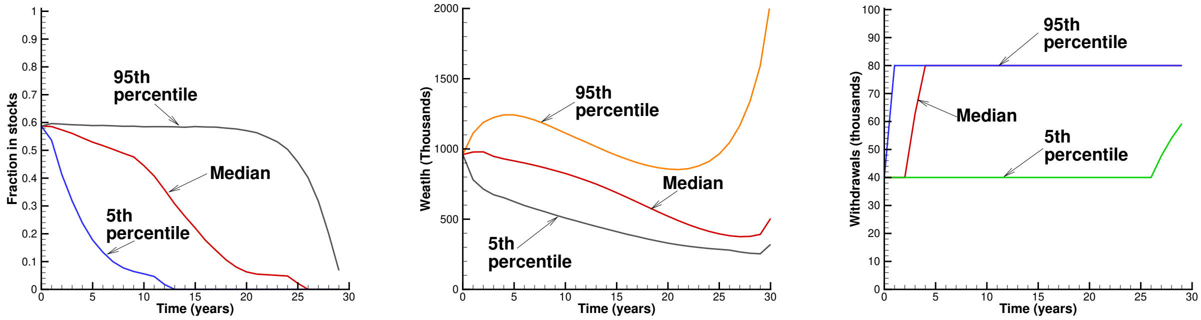
Traditional annuities with true inflation protection are unavailable in Canada.[23] Since inflation is expected to be a major factor in the coming years, inflation protection is a valuable aspect of the optimal EW-ES strategy, with a tontine overlay.[24] This strategy has an expected *real* withdrawal rate, over 30 years, of about 7% of the initial capital (per annum), never withdraws less than 4% of

[21]"When you have won the game, stop playing," – William Bernstein.

[22]*"If we have a good year, we take a trip to China, ...if we have a bad year, we stay home and play canasta,"* retired professor Peter Ponzo, discussing his DC plan withdrawal strategy `https://www.theglobeandmail.com/report-on-business/math-prof-tests-investing-formulas-strategies/article22397218/`.

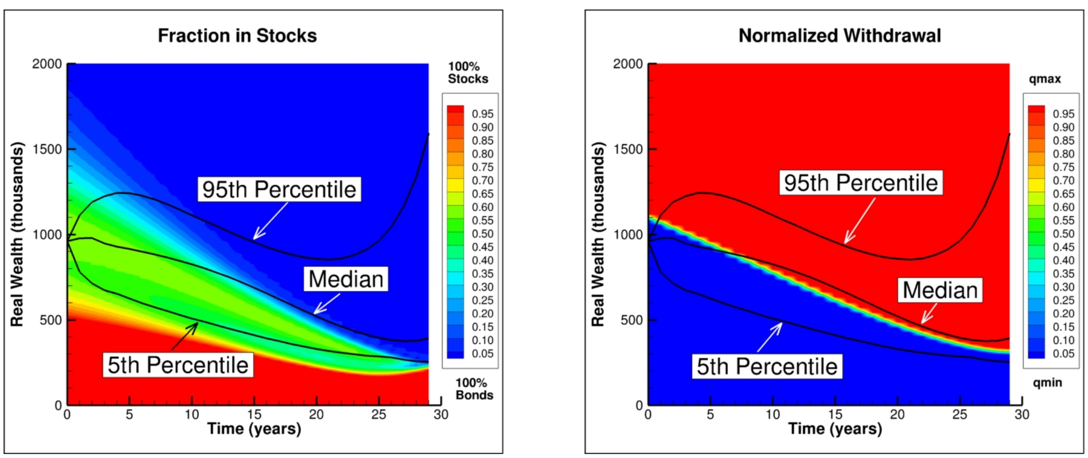
[23]Some providers advertise annuities with inflation protection, however this is simply an escalating nominal payout, based on a fixed escalation rate.

[24]Examination of historical periods of high inflation suggests that a portfolio of short term T-bills and an equal weight stock index generates significant positive real returns, see Forsyth (2022a).



(a) *Percentiles fraction in stocks* (b) *Percentiles wealth* (c) *Percentiles withdrawals*

FIGURE 15.1: *Scenario in Table 11.1. EW-ES control computed from problem EW-ES Problem (7.2). Parameters based on the real CRSP index, and real 30-day T-bills (see Table 9.1). Control computed and stored from the Problem (7.2) in the synthetic market. Control used in the historical market,* 10^6 *bootstrap samples.* $q_{min} = 40, q_{\max} = 80$ *(per year),* $\kappa = 0.18$*.* $W^* = 385$*. Units: thousands of dollars.*



(a) *Fraction in stocks* (b) *Withdrawals*

FIGURE 15.2: *Optimal EW-ES. Heat map of controls: fraction in stocks and withdrawals, computed from Problem EW-ES (7.1). Real capitalization weighted CRSP index, and real 30-day T-bills. Scenario given in Table 11.1. Control computed and stored from the Problem 7.2 in the synthetic market.* $q_{min} = 40, q_{\max} = 80$ *(per year).* $\kappa = 0.18$*.* $W^* = 385$*.* $\epsilon = -10^{-4}$*. Normalized withdrawal* $(q - q_{\min})/(q_{\max} - q_{\min})$*. Units: thousands of dollars.*

initial capital per annum, and a positive ES (expected shortfall) at the 5% level after 30 years.

Consequently, if we consider a retiree with no bequest motive, then joining a tontine pool and following an optimal EW-ES strategy is potentially an excellent alternative to a life annuity. Hence, it could be argued that going forward, the EW-ES optimal tontine pool strategy has less risk than a conventional annuity. However, this implicitly assumes that the idiosyncratic portion of mortality risk is much more significant than the systematic portion, as the tontine pool provides protection against the first component while a traditional annuity in principle protects against both components.

Another point to consider is that the reason that the tontine approach has a higher mean (and median) payout is that it is not guaranteed. There is some flexibility in the withdrawal amounts, and the portfolio contains risky assets. However, the ultimate risk, as measured by the expected shortfall at year 30, is very small. We can also see that the median payout rises rapidly to the maximum withdrawal rate (8% real of the initial investment) within 3-4 years of retirement, and stays at the maximum for the remainder of the 30-year horizon.

As well, the investor forfeits the entire portfolio in the event of death. Although this is often considered a drawback, we remind the reader that annuities and defined benefit (DB) plans have this same property (restricting attention to a single retiree with no guarantee period).[25] Of course, it is possible to overlay various guarantees on to the tontine pool, e.g. a guarantee period, a money back guarantee, or joint and survivor benefits. The cost of these guarantees would, of course, reduce the expected annual withdrawals.

These results are robust to fees in the range of 50-100 bps per year. The long term results are also insensitive to random group gains.[26]

However, the tontine gains (after fees) are comparatively small for retirees in the 65-70 age range. This suggests that it may be optimal to delay joining a tontine until the investor has attained an age of 70 or more.

Although we have explicitly excluded a bequest motive from our considerations, note that the median withdrawal strategy rapidly ramps up to the maximum withdrawal within a few years of retirement, and remains there for the entire remaining retirement period. Although it is commonly postulated that retirement consumption follows a *U-shaped* pattern, recent studies indicate that real retirement consumption falls with age (at least in countries which do not have large end of life expenses)(Brancati et al., 2015). In this case, the withdrawals which occur towards the end of the retirement period may exceed consumption. This allows the retiree to use these excess cash flows as a living bequest to relatives or charities.

17 Conclusions

DC plan decumulation strategies are typically based on some variant of the *four per cent rule* (Bengen, 1994). However, bootstrap tests of these rules using historical data show a significant risk of running out of savings at the end of a 30-year retirement planning horizon.

This risk can be significantly reduced by using optimal stochastic control methods, where the controls are the asset allocation strategy and the withdrawal amounts (subject to maximum and minimum constraints)(Forsyth, 2022b; Forsyth et al., 2020).

However, if we assume the retiree couples an optimal allocation/withdrawal strategy with participation in a tontine fund, then the risk of portfolio depletion after 30 years is virtually eliminated. At the same time, the cumulative total withdrawals are considerably increased compared with the previous two strategies. Of course, this comes at a price: the retiree forfeits her portfolio upon death. Hence the tontine overlay is most appealing to investors who have no bequest motivation, or who manage bequests using other funds.

We should also note that individual tontine accounts allow for complete flexibility in asset allocation strategies and do not require purchase of expensive investment products. These accounts are essentially peer-to-peer longevity risk management tools. Consequently, the custodian of these accounts bears no risk, and incurs only bookkeeping costs. Hence the fees charged by the custodian of these accounts can be very low. If desired, the retiree can pay for additional investment advice in a completely transparent manner.

[25] Moshe Milevsky, an advocate of modern tontines, is quoted in the Toronto Star (April 13, 2021) as noting that *"If you give up some of your money when you die, you can get more when you are alive."*

[26] The randomness of the group gain is due to fact that real tontine pools will be finite and heterogeneous.

However, a potentially significant caveat to our main conclusions is that we have ignored systematic mortality risk (e.g. unexpected improvement in life expectancies). Modelling this would require taking into account an additional risk factor, which we leave as a topic of future work.

18 Acknowledgements

Forsyth's work was supported by the Natural Sciences and Engineering Research Council of Canada (NSERC) grant RGPIN-2017-03760. Vetzal's work was supported by a Canadian Securities Institute Research Foundation grant 2021-2022.

19 Declaration

The authors have no conflicts of interest to report.

Appendix

A Induced Time Consistent Strategy

Denote the investor's initial wealth at t_0 by W_0^-. Then we have the following result:

Proposition A.1 (Pre-commitment strategy equivalence to a time consistent policy for an alternative objective function)**.** *The pre-commitment EW-ES strategy $\mathcal{P}^*$ determined by solving $J(0, W_0, t_0^-)$ (with $\mathcal{W}^*(0, W_0^-)$ from equation (7.5)) is the time consistent strategy for the equivalent problem TCEQ (with fixed $\mathcal{W}^*(0,W_0^-)$), with value function $\tilde{J}(s,b,t)$ defined by*

$$(TCEQ_{t_n}(\kappa/\alpha)): \quad \tilde{J}\left(s,b,t_n^-\right) = \sup_{\mathcal{P}_n \in \mathcal{A}} \left\{ E_{\mathcal{P}_n}^{X_n^+,t_n^+} \left[\sum_{i=n}^{M} \mathfrak{q}_i + \frac{\kappa}{\alpha} \min(W_T - \mathcal{W}^*(0, W_0^-),0) \right.\right.$$
$$\left.\left. \bigg| X(t_n^-) = (s,b) \right] \right\} . \tag{A.1}$$

Proof. This follows similar steps as in Forsyth (2020a), proof of Proposition 6.2, with the exception that the reward in Forsyth (2020a) is expected terminal wealth, while here the reward is total withdrawals. □

Remark A.1 (An Implementable Strategy)**.** *Given an initial level of wealth W_0^- at t_0, then the optimal control*[27] *for the pre-commitment problem (7.2) is the same optimal control for the time consistent problem*[28] *$(TCEQ_{t_n}(\kappa/\alpha))$ (A.1), $\forall t > 0$. Hence we can regard problem $(TCEQ_{t_n}(\kappa/\alpha))$ as the* EW-ES induced time consistent strategy*. Thus, the induced strategy is implementable, in the sense that the investor has no incentive to deviate from the strategy computed at time zero, at later times (Forsyth, 2020a).*

Remark A.2 (EW-ES Induced Time Consistent Strategy)**.** *In the following, we will consider the actual strategy followed by the investor for any $t > 0$ as given by the induced time consistent strategy $(TCEQ_{t_n}(\kappa/\alpha))$ in equation (A.1), with a fixed value of $\mathcal{W}^*(0, W_0^-)$, which is identical to the EW-ES strategy at time zero. Hence, we will refer to this strategy in the following as the EW-ES strategy, with the understanding that this refers to strategy $(TCEQ_{t_n}(\kappa/\alpha))$ for any $t > 0$.*

B Numerical Techniques

We solve problems (7.2) using the techniques described in detail in Forsyth and Labahn (2019); Forsyth (2020a; 2022b). We give only a brief overview here.

We localize the infinite domain to $(s,b) \in [s_{\min}, s_{\max}] \times [b_{\min}, b_{\max}]$, and discretize $[b_{\min},b_{\max}]$ using an equally spaced $\log b$ grid, with n_b nodes. Similarly, we discretize $[s_{\min}, s_{\max}]$ on an equally spaced $\log s$ grid, with n_s nodes. Localization errors are minimized using the domain extension method in Forsyth and Labahn (2019).

At rebalancing dates, we solve the local optimization problem (8.7) by discretizing $(\mathfrak{q}(\cdot), \mathfrak{p}(\cdot))$ and using exhaustive search. Between rebalancing dates, we solve the two dimensional partial integro-differential equation (PIDE) (8.10) using Fourier methods (Forsyth and Labahn, 2019; Forsyth,

[27]To be perfectly precise here, in the event that the control is non-unique, we impose a tie-breaking strategy to generate a unique control.

[28]Assuming that the same tie breaking strategy is used as for the pre-commitment problem.

2022b). Finally, the optimization problem (8.4) is solved using a one-dimensional optimization technique.

We used the value $\epsilon = -10^{-4}$ in equation (8.2), which forces the investment strategy to be bond heavy if the remaining wealth in the investor's account is large, and $t \to T$. Using this small value of gave the same results as $\epsilon = 0$ for the summary statistics, to four digits. This is simply because the states with very large wealth have low probability. However, this stabilization procedure produced smoother heat maps for large wealth values, without altering the summary statistics appreciably.

B.1 Convergence Test: Synthetic Market

We compute and store the optimal controls from solving Problem 7.2 using the parametric model of the stock and bond processes. We then use the stored controls in Monte Carlo simulations to generate statistical results. As a robustness check, we also use the stored controls and simulate results using bootstrap resampling of historical data.

Table B.1 shows a detailed convergence test for the base case problem given in Table 11.1, for the EW-ES problem. The results are given for a sequence of grid sizes, for the dynamic programming algorithm in Section 8 and Appendix B. The dynamic programming algorithm appears to converge at roughly a second order rate. The optimal control computed using dynamic programming is stored, and then used in Monte Carlo computations. The Monte Carlo results are in good agreement with the dynamic programming solution. For all the numerical examples, we will use the 2048×2048 grid, since this seems to be accurate enough for our purposes.

	Algorithm in Section 8 and Appendix B			Monte Carlo	
Grid	ES (5%)	$E[\sum_i \mathfrak{q}_i]/M$	Value Function	ES (5%)	$E[\sum_i \mathfrak{q}_i]/M$
512×512	108.13	67.99	2059.60	123.26	68.04
1024×1024	158.88	67.79	2063.19	164.45	67.81
2048×2048	201.88	67.56	2064.27	203.87	67.56
4096×4096	206.56	67.54	2064.54	207.70	67.54

TABLE B.1: *Convergence test, real stock index: deflated real capitalization weighted CRSP, real bond index: deflated 30 day T-bills. Scenario in Table 11.1. Parameters in Table 9.1. The Monte Carlo method used* 2.56×10^6 *simulations. The MC method used the control from the algorithm in Section 8.* $\kappa = 0.185, \alpha = .05$. *Grid refers to the grid used in the Algorithm in Section B:* $n_x \times n_b$, *where* n_x *is the number of nodes in the* $\log s$ *direction, and* n_b *is the number of nodes in the* $\log b$ *direction. Units: thousands of dollars (real).* M *is the total number of withdrawals (rebalancing dates).*

C Continuous Withdrawal/Rebalancing Limit

In order to develop some intuition about the nature of the optimal controls, we will examine the limit as the rebalancing interval becomes vanishingly small.

Proposition C.1 (Bang-bang withdrawal control in the continuous withdrawal limit)**.** *Assume that*

- *the stock and bond processes follow (4.3) and (4.4),*
- *the portfolio is continuously rebalanced, and withdrawals occur at a continuous (finite) rate* $\hat{\mathfrak{q}} \in [\hat{\mathfrak{q}}_{\min}, \hat{\mathfrak{q}}_{\max}]$,

	Algorithm in Section 8 and Appendix B			Monte Carlo	
Grid	ES (5%)	$E[\sum_i \mathfrak{q}_i]/T$	Value Function	ES (5%)	$E[\sum_i \mathfrak{q}_i]/M$
512×512	-203.31	54.08	860.033	-191.99	53.96
1024×1024	-191.40	53.58	889.613	-188.07	53.53
2048×2048	-188.91	53.57	898.712	-188.14	53.55
4096×4096	-188.04	53.54	901.106	-187.95	53.53

TABLE B.2: *No tontine case. Convergence test, real stock index: deflated real capitalization weighted CRSP, real bond index: deflated 30 day T-bills. Scenario in Table 11.1, but no tontine. Parameters in Table 9.1. The Monte Carlo method used* 2.56×10^6 *simulations. The MC method used the control from the algorithm in Section 8.* $\kappa = 3.75, \alpha = .05$. *Grid refers to the grid used in the Algorithm in Section B:* $n_x \times n_b$, *where* n_x *is the number of nodes in the* $\log s$ *direction, and* n_b *is the number of nodes in the* $\log b$ *direction. Units: thousands of dollars (real).* M *is the total number of withdrawals (rebalancing dates).* $W^* = -106.476$ *on the finest grid.*

- *the HJB equation for the EW-ES problem in the continuous rebalancing limit has bounded derivatives w.r.t. total wealth,*
- *in the event of ties for the control* $\hat{\mathfrak{q}}$, *the smallest withdrawal is selected,*

then the optimal withdrawal control $\hat{\mathfrak{q}}^*(\cdot)$ *for the EW-ES problem* $(PCES_{t_0}(\kappa))$ *and for the EW-LS problem* $(EWLS_{t_0}(\hat{\kappa}))$ *is bang-bang,* $\hat{\mathfrak{q}}^* \in \{\hat{\mathfrak{q}}_{\min}, \hat{\mathfrak{q}}_{\max}\}$.

Proof. This follows the same steps as in Forsyth (2022b). □

Remark C.1 (Bang-bang control for discrete rebalancing/withdrawals). *Proposition C.1 suggests that, for sufficiently small rebalancing intervals, we can expect the optimal* $\mathfrak{q}$ *control (finite withdrawal amount) to be bang-bang, i.e. it is only optimal to withdraw either the maximum amount* $\mathfrak{q}_{\max}$ *or the minimum amount* $\mathfrak{q}_{\min}$. *However, it is not clear that this will continue to be true for the case of yearly rebalancing (which we specify in our numerical examples), and finite amount controls* $\mathfrak{q}$. *In fact, we do observe that the finite amount control* $\mathfrak{q}$ *is very close to bang-bang in our numerical experiments, even for yearly rebalancing.*

D Detailed Efficient Frontiers: Synthetic Market

κ	$E[\sum_i \mathfrak{q}_i]/T$	ES(5%)	$Median[W_T]$	W^*
0.15	70.06	-309.569	189.48	0.490
0.17	70.04	-270.13	185.19	0.489
0.18	68.51	46.77	599.42	385.28
0.185	67.56	203.87	820.65	585.97
0.20	66.41	384.76	1058.40	802.40
0.25	63.85	732.34	1517.04	1220.33
0.30	62.22	912.29	1754.40	1439.83
0.50	58.48	1209.40	2120.59	1802.19
1.0	54.81	1372.46	2327.42	2021.22
10.0	48.96	1457.52	2484.58	2151.79
∞	40.00	1460.76	2885.85	2173.04

TABLE D.1: *EW-ES synthetic market results for optimal strategies, assuming the scenario given in Table 11.1. Tontine gains assumed. Stock index: real capitalization weighted CRSP stocks; bond index: real 30-day T-bills. Parameters from Table 9.1. Units: thousands of dollars. Statistics based on* 2.56×10^6 *Monte Carlo simulation runs. Control is computed using the Algorithm in Section 8 and Appendix B, stored, and then used in the Monte Carlo simulations.* $q_{\min} = 0.40$, $q_{\max} = 80$ *(annually).* $T = 30$ *years,* $\epsilon = -10^{-4}$.

κ	$E[\sum_i \mathfrak{q}_i]/T$	ES(5%)	$Median[W_T]$	W^*
0.180	69.17	-823.76	-2.51	-691.81
1.0	61.38	-319.66	-39.47	-229.18
1.5	58.98	-260.92	-65.88	-179.60
1.75	57.97	-242.34	-74.74	-161.25
2.5	55.86	-211.03	-81.44	-132.87
3.75	53.55	-188.14	-81.11	-107.00
5.0	52.08	-177.88	-78.39	-90.10
6.25	51.29	-173.59	-79.08	-89.03
7.5	50.72	-171.05	-79.30	-88.25
10.0	49.89	-168.16	-78.77	-87.18
100.0	46.41	-162.86	-68.28	-77.47
∞	40.00	-162.67	+5.72	-76.00

TABLE D.2: *EW-ES synthetic market results for optimal strategies, assuming the scenario given in Table 11.1. No tontine gains assumed. Stock index: real capitalization weighted CRSP stocks; bond index: real 30-day T-bills. Parameters from Table 9.1. Units: thousands of dollars. Statistics based on* 2.56×10^6 *Monte Carlo simulation runs. Control is computed using the Algorithm in Section 8 and Appendix B, stored, and then used in the Monte Carlo simulations.* $q_{\min} = 0.40$, $q_{\max} = 80$ *(annually).* $T = 30$ *years,* $\epsilon = -10^{-4}$.

E Detailed Efficient Frontiers: Historical Market

κ	$E[\sum_i \mathfrak{q}_i]/T$	ES(5%)	$Median[W_T]$
0.15	71.25	-165.23	157.16
0.17	71.01	-138.15	153.13
0.18	68.94	204.20	573.29
0.185	67.99	369.26	769.96
0.20	66.64	546.98	1038.07
0.25	63.84	863.20	1500.51
0.30	62.08	1011.55	1739.21
0.5	58.13	1211.18	2115.22
1.0	54.50	1285.93	2330.33
10.0	49.42	1275.98	2485.58
∞	40.00	1280.97	2892.41

Table E.1: *EW-ES historical market results for optimal strategies, assuming the scenario given in Table 11.1. Tontine gains assumed. Stock index: real capitalization weighted CRSP stocks; bond index: real 30-day T-bills. Parameters from Table 9.1. Units: thousands of dollars. Statistics based on 10^6 bootstrap simulation runs. Expected blocksize = 2 years. Control is computed using the Algorithm in Section 8 and Appendix B, stored, and then used in the bootstrap simulations. $q_{\min} = 40$, $q_{\max} = 80$ (annually). $T = 30$ years, $\epsilon = -10^{-4}$.*

κ	$E[\sum_i \mathfrak{q}_i]/T$	ES(5%)	$Median[W_T]$
0.18	69.91	-805.65	-31.84
1.0	61.77	-290.03	-40.87
1.5	59.21	-248.15	-77.26
1.75	58.16	-235.46	-78.50
2.5	56.02	-219.00	-81.84
3.75	53.78	-209.90	-80.68
5.0	52.43	-207.15	-77.25
6.25	51.74	-209.02	-78.11
7.5	51.26	-210.38	-78.48
10.0	50.58	-212.41	-77.95
100.0	47.72	-217.82	-67.91
∞	40.00	-219.16	+17.34

Table E.2: *EW-ES historical market results for optimal strategies, assuming the scenario given in Table 11.1. No Tontine gains assumed. Stock index: real capitalization weighted CRSP stocks; bond index: real 30-day T-bills. Parameters from Table 9.1. Units: thousands of dollars. Statistics based on 10^6 bootstrap simulation runs. Expected blocksize = 2 years. Control is computed using the Algorithm in Section 8 and Appendix B, stored, and then used in the bootstrap simulations. $q_{\min} = 40$, $q_{\max} = 80$ (annually). $T = 30$ years, $\epsilon = -10^{-4}$.*

References

Ameriks, J., R. Veres, and M. J. Warshawsky (2001). Making retirement income last a lifetime. *Journal of Financial Planning 14:12*, 60–74.

Anarkulova, A., S. Cederburg, and M. S. O'Doherty (2022). Stocks for the long run? Evidence from a broad sample of developed markets. *Journal of Financial Economics 143:1*, 409–433.

Bär, M. and N. Gatzert (2023). Products and strategies for decumulation of wealth during retirement: insights from the literature. *North American Actuarial Journal 27:2*, 322–340.

Bengen, W. (1994). Determining withdrawal rates using historical data. *Journal of Financial Planning 7*, 171–180.

Bernhardt, T. and C. Donnelly (2018). Pension decumulation strategies: A state of the art report. Technical Report, Risk Insight Lab, Heriot Watt University.

Bernhardt, T. and C. Donnelly (2019). Modern tontine with bequest: Innovation in pooled annuity products. *Insurance: Mathematics and Economics 86*, 168–188.

Brancati, C., B. Beach, and B. Frankin (2015). Understanding retirement journeys: Expectations vs reality. International Longevity Centre UK, `https://www.bl.uk/collection-items/understanding-retirement-journeys-expectations-vs-reality`.

Bräutigam, M., M. Guillén, and J. P. Nielsen (2017). Facing up to longevity with old actuarial methods: a comparison of pooled funds and income tontines. *Geneva papers on risk and insurance 42*, 406–422.

Cavaglia, S., L. Scott, K. Blay, and S. Hixon (2022). Multi-asset class factor premia: A strategic asset allocation perspective. *The Journal of Portfolio Management 48:9*, 14–32.

Chen, A., M. Guillen, and M. Rach (2021). Fees in tontines. *Insurance: Mathematics and Economics 100*, 89–106.

Chen, A., P. Hieber, and J. K. Klein (2019). Tonuity: A novel individual-oriented retirement plan. *ASTIN Bulletin 49*, 5–30.

Chen, A., P. Hieber, and M. Rach (2021). Optimal retirement products under subjective mortality beliefs. *Insurance: Mathematics and Economics 101:A*, 55–69.

Chen, A. and M. Rach (2022). The tontine puzzle. SSRN 4106903.

Chen, M., M. Shirazi, P. A. Forsyth, and Y. Li (2023). Machine learning and Hamilton-Jacobi-Bellman Equation for optimal decumulation: a comparison study. arXiv 2306.10582.

Cogneau, P. and V. Zakalmouline (2013). Block bootstrap methods and the choice of stocks for the long run. *Quantitative Finance 13*, 1443–1457.

Cont, R. and C. Mancini (2011). Nonparametric tests for pathwise properties of semimartingales. *Bernoulli 17*, 781–813.

Cui, X., D. Li, X. Qiao, and M. Strub (2022). Risk and potential: An asset allocation framework with applications to robo-advising. *Journal of the Operations Research Society of China 10*, 529–558.

Daily, J., E. Palmer, and J. Mizell (2023). A primer on retirement savings for pediatricians. *Clinical Pediatrics 62:7*, 678–682.

Dang, D.-M. and P. A. Forsyth (2016). Better than pre-commitment mean-variance portfolio allocation strategies: a semi-self-financing Hamilton-Jacobi-Bellman equation approach. *European Journal of Operational Research 250*, 827–841.

Denuit, M., P. Hieber, and C. Y. Robert (2022). Mortality credits within large survivor funds. *ASTIN Bulletin 52(3)*, 813–834.

Denuit, M. and R. Vernic (2018). Bivariate Bernoulli weighted sums and distribution of single period Tontine benefits. *Methodology and Computing in Applied Probability 20*, 1403–1416.

Dichtl, H., W. Drobetz, and M. Wambach (2016). Testing rebalancing strategies for stock-bond portfolios across different asset allocations. *Applied Economics 48*, 772–788.

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

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

Forsyth, P. and G. Labahn (2019). $\epsilon-$Monotone Fourier methods for optimal stochastic control in finance. *Journal of Computational Finance 22:4*, 25–71.

Forsyth, P. A. (2020a). Multi-period mean CVAR asset allocation: Is it advantageous to be time consistent? *SIAM Journal on Financial Mathematics 11:2*, 358–384.

Forsyth, P. A. (2020b). Optimal dynamic asset allocation for DC plan accumulation/decumulation: Ambition-CVAR. *Insurance: Mathematics and Economics 93*, 230–245.

Forsyth, P. A. (2022a). Asset allocation during high inflation periods: A stress test. `https://cs.uwaterloo.ca/~paforsyt/inflation_stress_test.pdf`.

Forsyth, P. A. (2022b). A stochastic control approach to defined contribution plan decumulation: “The Nastiest, Hardest Problem in Finance”. *North American Actuarial Journal 26:2*, 227–252.

Forsyth, P. A. and K. R. Vetzal (2019). Optimal asset allocation for retirement savings: deterministic vs. time consistent adaptive strategies. *Applied Mathematical Finance 26:1*, 1–37.

Forsyth, P. A., K. R. Vetzal, and G. Westmacott (2020). Optimal asset allocation for a DC pension decumulation with a variable spending rule. *ASTIN Bulletin 50:2*, 419–447.

Forsyth, P. A., K. R. Vetzal, and G. Westmacott (2022). Optimal control of the decumulation of a retirement portfolio with variable spending and dynamic asset allocation. *ASTIN Bulletin 51:3*, 905–938.

Fuentes, O. M., R. K. Fullmer, and M. Garcia-Huitron (2022). A sustainable, variable lifetime retirement income solution for the Chilean pension system. SSRN 4045646 `https://ssrn.com/abstract=4045646`.

Fullmer, R. K. (2019). Tontines: a pracitioner’s guide to mortality-pooled investments. CFA Institute Research Foundation `https://www.cfainstitute.org/en/research/foundation/2019/tontines`.

Fullmer, R. K. and J. B. Forman (2022). State sponsored pensions for private sector workers. In O. S. Mitchell (Ed.), *New models for managing longevity risk*, Chapter 10, pp. 171–206. Oxford: Oxford University Press.

Fullmer, R. K. and M. J. Sabin (2019). Individual tontine accounts. *Journal of Accounting and Finance 19:8*, 31–61.

Gemmo, I., R. Rogalla, and J.-H. Weinart (2020). Optimal portfolio choice with tontines under systematic longevity risk. *Annals of Actuarial Science 14*, 302–315.

Guyton, J. T. and W. J. Klinger (2006). Decision rules and maximum initial withdrawal rates. *Journal of Financial Planning 19*(3), 48–58.

Hatch, J. E. and R. W. White (1985). Canadian stocks, bonds, bills and inflation: 1950-1983. Financial Analysts Research Foundation Mongraph Series, Number 19.

Hieber, P. and N. Lucas (2022). Modern life care tontines. *ASTIN Bulletin 52(2)*, 563–589.

Hill, C. (2016). Older people fear this more than death. `https://www.marketwatch.com/story/older-people-fear-this-more-than-death-2016-07-18`.

Irlam, G. (2014). Portfolio size matters. *Journal of Personal Finance 13*(2), 9–16.

Kou, S. G. (2002). A jump-diffusion model for option pricing. *Management Science 48*, 1086–1101.

Kou, S. G. and H. Wang (2004). Option pricing under a double exponential jump diffusion model. *Management Science 50*, 1178–1192.

Li, Y. and P. A. Forsyth (2019). A data driven neural network approach to optimal asset allocation for target based defined contribution pension plans. *Insurance: Mathematics and Economics 86*, 189–204.

Lin, Y., R. MacMinn, and R. Tian (2015). De-risking defined benefit plans. *Insurance: Mathematics and Economics 63*, 52–65.

MacDonald, B.-J., B. Jones, R. J. Morrison, R. L. Brown, and M. Hardy (2013). Research and reality: A literature review on drawing down retirement financial savings. *North American Actuarial Journal 17*, 181–215.

MacDonald, B. J., B. Sanders, L. Strachan, and M. Frazer (2021). Affordable Lifetime Pension Income for a Better Tomorrow. How we can address the 1.5 trillion decumulation disconnect in the Canadian retirement income system with Dynamic Pension pools. National Institute on Ageing, Ryerson University and Global Risk Institute, `https://globalriskinstitute.org/publication/affordable-lifetime-pension-income-for-a-better-tomorrow/` .

MacMinn, R., P. Brockett, J. Wang, Y. Lin, and R. Tian (2014). The securitization of longevity risk and its implications for retirement security. In O. S. Mitchell, R. Maurer, and P. B. Hammond (Eds.), *Recreating Sustainable Retirement*, pp. 134–160. Oxford: Oxford University Press.

Mancini, C. (2009). Non-parametric threshold estimation models with stochastic diffusion coefficient and jumps. *Scandinavian Journal of Statistics 36*, 270–296.

Milevsky, M. (2022). *How to build a modern tontine*. Toronto: Springer.

Milevsky, M. A. and T. S. Salisbury (2015). Optimal retirement income tontines. *Insurance: Mathematics and Economics 64*, 91–105.

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

Ni, C., Y. Li, P. A. Forsyth, and R. Caroll (2022). Optimal asset allocation for outperforming a stochastic benchmark. *Quantitative Finance 22(9)*, 1595–1626.

OECD (2019). Pension markets in focus. https://www.oecd.org/daf/fin/private-pensions/Pension-Markets-in-Focus-2019.pdf.

Patton, A., D. Politis, and H. White (2009). Correction to: automatic block-length selection for the dependent bootstrap. *Econometric Reviews 28*, 372–375.

Peijnenburg, K., T. Nijman, and B. J. Werker (2016). The annuity puzzle remains a puzzle. *Journal of Economic Dynamics and Control 70*, 18–35.

Pfau, W. D. (2015). Making sense out of variable spending strategies for retirees. *Journal of Financial Planning 28*(10), 42–51.

Pfau, W. D. (2018). An overview of retirement income planning. *Journal of Financial Counseling and Planning 29:1*, 114–120. DOI:10.1891/1052-3073.29.1.114.

Pfeiffer, S., J. R. Salter, and H. E. Evensky (2013). Increasing the sustainable withdrawal rate using the standby reverse mortgage. *Journal of Financial Planning 26:12*, 55–62.

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

Politis, D. and J. Romano (1994). The stationary bootstrap. *Journal of the American Statistical Association 89*, 1303–1313.

Politis, D. and H. White (2004). Automatic block-length selection for the dependent bootstrap. *Econometric Reviews 23*, 53–70.

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

Rockafellar, R. T. and S. Uryasev (2000). Optimization of conditional value-at-risk. *Journal of Risk 2*, 21–42.

Ruthbah, U. (2022). The retirement puzzle. *Australian Journal of Management 47:2*, 342–367.

Sabin, M. J. (2010). Fair tontine annuity. SSRN 1579932.

Sabin, M. J. (2011). A fast bipartite algorithm for fair tontines. SSRN 1848737.

Sabin, M. J. and J. B. Forman (2016). The analytics of a single period tontine. SSRN 2874160.

Scott, J. S., W. F. Sharpe, and J. G. Watson (2009). The 4% rule - at what price? *Journal Of Investment Management Third Quarter*. https://ssrn.com/abstract=1115023.

Shefrin, H. M. and R. H. Thaler (1988). The behavioral life-cycle hypothesis. *Economic Inquiry 26*, 609–643.

Simonian, J. and A. Martirosyan (2022). Sharpe parity redux. *The Journal of Portfolio Management 48:9*, 183–193.

Strub, M., D. Li, X. Cui, and J. Gao (2019). Discrete-time mean-CVaR portfolio selection and time-consistency induced term structure of the CVaR. *Journal of Economic Dynamics and Control 108*. Article 103751 (electronic).

Thaler, R. H. (1990). Anomalies: Savings, fungibility, and mental accounts. *Journal of Economic Perspectives 4*(1), 193–205.

van Benthem, L., L. Frehen, J. Cramwinckel, and R. de Kort (2018). Tonchain: the future of pensions? *VBA Journal 134:Summer*, 18–22.

van Staden, P., P. Forsyth, and Y. Li (2023). Beating a benchmark: dynamic programming may not be the right numerical approach. *SIAM Journal on Financial Mathematics 14:2*, 407–451.

van Staden, P. M., D.-M. Dang, and P. Forsyth (2018). Time-consistent mean-variance portfolio optimization: a numerical impulse control approach. *Insurance: Mathematics and Economics 83*, 9–28.

van Staden, P. M., D.-M. Dang, and P. Forsyth (2021). The surprising robustness of dynamic mean-variance portfolio optimization to model misspecification errors. *European Journal of Operational Research 289:2*, 74–792.

Waring, M. B. and L. B. Siegel (2015). The only spending rule article you will ever need. *Financial Analysts Journal 71*(1), 91–107.

Weinert, J.-H. and H. Gründl (2021). The modern tontine: an innovative instrument for longevity risk management in an aging society. *European Actuarial Journal 11*, 49–86.

Winter, P. and F. Planchet (2022). Modern tontines as a pension solution: a practical overview. *European Actuarial Journal 12*, 3–32.