The Modern Tontine: Potential and Practice

K. R. Vetzal* G. Westmacott† P. A. Forsyth‡

June 23, 2023

1 Introduction

What is a modern tontine, and what is its role in retirement planning? How should the risks and benefits of a tontine be quantified? What asset allocation should be used to optimize performance? These are some of the questions we sought to address in a recent technical report (Forsyth et al., 2022). This article summarizes the main ideas and results from that report. To avoid repetition, we assume that readers will refer to the technical report and references therein where further clarification is needed.

As with most financial decisions, our analysis involves trading off risk and reward. More details will be given below, but here we briefly note that in our case reward involves total income over a 30-year retirement horizon starting at age 65, while risk depends on the terminal portfolio value at the end of the horizon in the worst 5% of cases. Annual withdrawals are restricted to lie between a lower and an upper bound, and the retiree has to decide how much to invest in equities and fixed income. Due to required minimum yearly withdrawals, it is possible for retirees to exhaust their savings during the retirement period. We consider three strategies:

- *Strategy 1: Constant asset weights and fixed 4% withdrawals.* This is based on the ubiquitous 4% rule (Bengen, 1994), which holds that retirees who rebalance their portfolios annually to appropriate fixed weights (e.g. 60% equities and 40% bonds) can safely spend 4% of their initial retirement assets, adjusted annually for inflation.
- *Strategy 2: Optimal asset weights and variable withdrawals.* In this case, the retiree can choose to spend more than the level afforded by the 4% rule and also has the flexibility to adjust the asset allocation over time in response to past investment returns and withdrawals so as to optimize the risk/reward tradeoff.
- *Strategy 3: Optimal asset weights, variable withdrawals, and a tontine overlay.* This augments Strategy 2 with a tontine overlay, in which a surviving retiree earns mortality credits by participating in a tontine pool.

As will be seen below, allowing for flexible asset weights and withdrawals offers strong improvement over the standard 4% rule, and adding the tontine overlay gives a dramatic further enhancement (i.e. Strategy 3 significantly outperforms Strategy 2, which in turn is considerably

---

*School of Accounting and Finance, University of Waterloo, Waterloo ON Canada N2L 3G1, kvetzal@uwaterloo.ca.

†Richardson Wealth, 540 Bingemans Centre Drive, Suite 100, Kitchener ON Canada N2B 3X9, Graham.Westmacott@RichardsonWealth.com.

‡Cheriton School of Computer Science, University of Waterloo, Waterloo ON Canada N2L 3G1, paforsyt@uwaterloo.ca.

better than Strategy 1). It appears that average annual withdrawals in real terms of close to 7% are achievable with Strategy 3. This is much higher than the 4% rule suggests, while at the same time the tontine overlay offers much lower chance of retirement portfolio depletion after 30 years. This indicates that a modern tontine coupled with suitable strategies for asset allocation and withdrawals may be very attractive, at least for some retirees.

2 Some Background Context

Pension plans can generally be classified as either defined benefit (DB) or defined contribution (DC) schemes. The basic distinction is that DB plans specify the amount of income that a plan member is entitled to receive in retirement while DC plans only specify the amount that the plan member and the employer are required to contribute to retirement savings prior to retirement. DC plans thus place more risk on individual plan members, who must take responsibility for managing their own retirement savings investments. This risk substantially increases in the post-retirement phase when they also need to manage their withdrawals to fund living expenses, while faced with uncertain longevity.[1] A combination of poor investment returns and relatively high withdrawals runs the risk of "ruin", i.e. depleting the investment account prior to death. On the other hand, strong investment performance coupled with parsimonious withdrawals can lead to less enjoyable retirements and unintentionally large bequests.

While there is some variation across countries, the global trend over the past several years has been a shift from DB to DC plans. One indicator of this is the total value of assets held in these different types of plans. A recent study (Thinking Ahead Institute, 2023) provides statistics for the top seven countries by this measure as of 2022.[2] In aggregate, DC plan assets have grown over the past decade by 6.5% per annum, while DB plan assets have risen by 2.1% per annum over the same period. Over the past two decades, the corresponding figures are 7.2% per year for DC plans and 4.4% per year for DB plans. As of 2022 in Australia, DC plans accounted for 87% of total pension plan assets. At the opposite extreme, DC plans amounted for just 5% of total assets in Japan and the Netherlands. Canada occupies the middle ground, with 43% of total pension plan assets in DC plans.[3] That said, there is a marked difference in Canada between the public and private sectors. Overall, membership in Canadian pension plans as of 2021 was about 6.6 million, with about 67% being in DB plans, 18% in DC plans, and the remainder in hybrid plans that combine various features of DB and DC plans (e.g. plans which offer a relatively low defined benefit, augmented by amounts from a separate defined contribution account). However, about 90% of public sector employees who are in registered pension plans are in DB plans. The analogous figure for private sector members of registered plans is around 40% (Statistics Canada, 2022). Many existing DB plans have been closed to new entrants, as companies no longer want to take the risk of managing employees' retirement assets long after they have left the firm.

As noted above, a key difference between DB and DC plans is that DC plan members are exposed to much more risk in terms of managing investments and cash withdrawals during a retirement period of uncertain length. By contrast, DB plan members need not be concerned with investments and withdrawals (assuming plan solvency), and they implicitly benefit from longevity risk pooling: the contributions of plan members who pass away early in their retirements are

---

[1] DC plans also offer some advantages relative to DB plans, such as enhanced portability across different employers and, for individuals who desire to do so, self-management of investments. However, our focus here is on the extra risks placed on plan members.

[2] The countries include Australia, Canada, Japan, the Netherlands, Switzerland, United Kingdom, and the United States.

[3] By way of comparison, 65% of pension plan assets in the United States are in DC plans.

effectively used to augment the retirement benefits for those who live longer. The benefits of longevity risk pooling in DB plans for those who survive to advanced ages can be reduced by features such as guaranteed minimum payouts (e.g. 10 years of benefits) for members who die early, along with survivorship benefits for their spouses, but there is still some wealth transfer from those who die early to provide retirement income over a longer period for those who live a long time. This type of risk pooling is a key characteristic of DB plans, and its absence from traditional DC plans contributes to higher risk for DC plan members.

Annuitization of accumulated savings is an obvious strategy which could in principle be used to manage the risks of investments and portfolio withdrawals during retirement. However, in practice the vast majority of retirees seem highly reluctant to annuitize. There are several reasons for this, including insufficient liquidity, lack of inflation protection, credit risk, etc.[4]

Given the evident distaste for annuitization, it is natural to seek alternatives which are potentially more palatable. Over the past couple of decades, there has been considerable interest in various ideas to augment DC plans with some form of longevity risk pooling other than traditional annuities. Most of these ideas first appeared in the actuarial science literature. Some representative examples include group self-annuitization (Piggott et al., 2005), pooled annuity funds (Stamos, 2008), annuity overlay funds (Donnelly et al., 2014), and tontines (Milevsky and Salisbury, 2015). More recently, several papers on longevity risk pooling have appeared that are more targeted to industry practitioners. Fullmer (2019) and Milevsky (2022) provide excellent discussions of tontines. A useful comparison with traditional annuities is given in Milevsky et al. (2018). Some benefits of longevity risk pooling, plan design, and regulatory issues are discussed by MacDonald et al. (2021) and Bégin and Sanders (2023).

The basic concept of longevity risk pooling is obviously not new. For example, tontines date back hundreds of years (see, e.g. Milevsky and Salisbury, 2015; Milevsky, 2015). However, as discussed by MacDonald et al. (2021), for a variety of reasons including regulatory roadblocks, schemes that use pooling to reduce longevity risk have been quite uncommon in recent decades. One example is the College Retirement Equities Fund developed by TIAA in 1952. A second one in the Canadian context is the Variable Life Pension Annuity feature of the University of British Columbia's DC plan for its faculty, which originated in 1967. However, interest appears to have increased more recently. Australia's QSuper DC plan introduced a "Lifetime Pension" option in 2021. The Purpose Longevity Pension Fund launched in Canada in 2021 is a mutual fund which offers a relatively high target income payout, funded in part by mortality pooling. Somewhat similarly, Guardian Capital introduced its GuardPath Modern Tontine mutual fund in 2022.

Of course, traditional annuities can also be seen as providing a form of longevity risk pooling. It is worth noting, however, that annuities in principle provide more complete protection than modern tontines or other pooled mortality alternatives. In a modern tontine, surviving investors in the pool receive mortality credits from investors who pass away. This provides protection against *idiosyncratic* longevity risk, i.e. the chance of living longer and requiring more funds than other investors in the pool, on the assumption that the mortality experience of the pool in aggregate is similar to prior expectations. However, to the extent that members of the pool collectively live longer than anticipated, the mortality credits earned will be reduced or delayed. This form of *systematic* longevity risk is not diversified away in the pool and is borne by the participants. However, in the case of a traditional annuity, the provider of the annuity bears both forms of longevity risk. This makes annuities relatively expensive, as providers such as insurance companies must charge higher fees and hold reserves against both types of longevity risk. The overwhelming reluctance of most Canadians to annuitize their retirement savings suggests that this higher level

---

[4] See MacDonald et al. (2013) for a comprehensive discussion of reasons why annuitization is very unpopular.

of insurance may be seen as too costly, further motivating the study of alternatives such as tontines which offer somewhat less protection but at a cheaper price.

3 Investment Scenario

The investment scenario is as follows. To focus solely on the retirement period, consider a 65-year old Canadian male who retires with an initial portfolio of $1 million. This retiree plans to have his portfolio last until age 95, an investment horizon of 30 years. Due to the corrosive effects of inflation on purchasing power over such a long-term horizon, all results presented are in real (i.e. inflation-adjusted terms). We assume that the investor allocates his portfolio between a diversified market capitalization-weighted total return equity index and a short-term government bond index. In particular, the equity index is the Center for Research in Security Prices (CRSP) U.S. market index, while the bond index is the CRSP 30 day U.S. T-bill index.[5] Both of these indexes are adjusted for inflation, and both run on a monthly basis from January 1926 through December 2020. Although we specify a Canadian retiree, we use U.S. financial market data for two reasons:

- it permits easy comparisons with the benchmark case of the 4% withdrawal rule (Bengen, 1994) which was based on U.S. data; and
- we have a long return series, which includes some catastrophic and higly volatile market returns during the Great Depression of the 1930s, giving us a strong stress test of the model.[6]

The retiree allocates his funds to these two assets so as to balance the competing objectives of being able to withdraw and spend a reasonable amount each year while at the same time not depleting the portfolio prior to the end of the investment horizon. We also assume that the retiree owns mortgage-free real estate worth something like $400,000, but that this real estate is not seen by the retiree as an investment asset, but rather as a potential bequest, to be used by the retiree for financial purposes only as a last resort hedge.[7]

3.1 The Modern Tontine

With a tontine, the investor commits to investing in a pooled tontine fund for a period of time. If the investor dies while the tontine is active, the investor's portfolio is divided amongst the remaining members of the fund. If the investor survives until the end of the tontine, she benefits from redistributions from the fund from those who have passed away. These redistributions are variously called mortality credits, survivorship credits, or tontine gains. Reflecting mortality experience, tontine payouts are low for younger participants and increase sharply for older participants. Retirees often prefer higher payouts earlier in retirement and this is consistent with observed spending patterns. A modern tontine allows some flexibility in the payout profile. Some points to note:

- The pooled fund is assumed to be invested so it will earn an investment return in addition to mortality credits. We focus on an individual tontine account, in which each participant can control the asset allocation of their investment and withdrawals.[8]

---

[5] The results presented here were calculated based on Historical Indexes, ©2020 Center for Research in Security Prices, The University of Chicago Booth School of Business. Wharton Research Data Services 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.

[6] By contrast, comparable available Canadian market data starts around 1950.

[7] This hedge provides a safety buffer, allowing for borrowing to fund necessary expenses if the retiree's investment assets are entirely depleted.

[8] Alternatively, a tontine custodian could act on behalf of the retiree to manage the investment and withdrawals.

- The fund return is not guaranteed over any period as it depends on investment market returns and how closely actual mortality tracks expectations.
- The role of the fund manager is to keep track of active investors and to allocate the assets accordingly. Unlike an annuity, there are no guaranteed payouts to be financed so fees should be lower.

In principle, tontines can be structured in many different ways. We consider a tontine that is assumed to have a relatively large number of investors (e.g. on the order of 10,000; as discussed below this helps to ensure that actual deaths in a given period are not likely to be very different from the expected number of deaths, i.e. that idiosyncratic longevity risk is diversified) and which is open-ended, so that new investors join the pool over time. We also assume that a risk pool ownership constraint holds (Fullmer, 2019): no investor in the tontine has such a large amount invested that she would not be fairly compensated even by receiving all of the tontine gains in the pool.

We focus on tontines that are *fair* (Fullmer, 2019; Fullmer and Sabin, 2019). This means that anyone can join a tontine pool but since mortality risk is linked to age, so is the reward: younger participants should expect a lower payout than older participants to be fair.[9] In particular, consider an investor in the tontine pool who has a portfolio worth v_t at the end of period t. This amount would be forfeited to the other members of the pool if this investor dies during period t, which is assumed to occur with probability q_t. The expected amount forfeited at the end of this period for this investor is $q_t v_t$. If this investor does not die during the period, he will receive a mortality credit of c_t at the end of the period. Since the probability that the investor doesn't die during period t is $1 - q_t$, the expected mortality credit or tontine gain is $(1 - q_t)c_t$. For participation in the tontine to be fair for this investor, the expected tontine gain must be equal to the expected amount foreited, i.e.

$$(1 - q_t)c_t = q_t v_t.$$

Rearranging the above expression shows that the expected tontine gain is

$$c_t = \left(\frac{q_t}{1 - q_t}\right) v_t = r v_t$$

where $r = q_t/(1 - q_t)$. A younger pool member will have a lower chance of dying during a period (i.e. a lower value of q_t), which leads to a lower value of r. In other words, as a fraction of the investor's account balance, the expected tontine gain will increase with age.

Aggregating across all members of the pool, it can be shown that expected total mortality credits are balanced by expected total forfeitures. However, in any period the number of actual deaths may not be equal to the expected number of deaths. Following Fullmer (2019), let the *group gain* G_t for period t be defined as the ratio of actual forfeitures to expected mortality credits. The expected tontine gain for all investors is then multiplied by G_t to determine the actual mortality credit distributed to survivors in the pool. The expected value of G_t is 1, but it will randomly deviate from this level as actual forfeitures differ from expected forfeitures. If actual forfeitures are higher than expected, G_t will exceed one, and distributed tontine gains will be raised above their expected values accordingly. Conversely, distributed gains will be reduced if actual forfeitures are lower than anticipated.

---

[9] Providers of tontine-like funds often restrict offerings to relatively narrow age cohorts, which enables them to treat everyone within the fund equally. For example, investors in the GuardPath Modern Tontine mutual fund are currently required to have birth dates from 1957-1961.

The group gain is a simple adjustment mechanism that ensures that the total credits paid out at the end of every period match the total forfeitures actually received. An obvious question is how volatile G is over time. A critical factor is the size of the pool. If it is sufficiently large, then G will have a low standard deviation and it will fluctuate in a narrow range around its expected value of 1. Fullmer and Sabin (2019) conduct a simulation experiment showing that with a pool size of at least 10,000 members (who vary randomly by age, gender, investment policy, and initial portfolio size) the mean of G is very close to 1, and its standard deviation is about 0.1. Our analysis (Forsyth et al., 2022) assumes that $G = 1$, i.e. we have a sufficiently large and open-ended pool (i.e. new members join over time) that errors from assuming $G = 1$ are insignificant. However, a robustness test with G varying randomly in accordance with the simulation of Fullmer and Sabin (2019) showed very little difference compared to the case with G assumed to be constant at 1 in every period.[10] Our analysis calculates tontine gains based on the CPM 2014 mortality table from the Canadian Institute of Actuaries. In addition, we assume that investors in the tontine pool pay fees of 50 basis points per year, as a fraction of their portfolio values at the start of each year, before annual withdrawals but after tontine gains.

3.2 Annual Withdrawals

We specify a lower and an upper bound to annual withdrawals which moderate the tendency for tontine payouts to be back-end loaded. These bounds are inflation-indexed. The minimum withdrawal can be linked to the essential expenditure required by the retiree. We set this lower bound to \$40,000 (4% of the initial portfolio value, which facilitates a convenient comparison with the familiar 4% rule). Of course, requiring that at least \$40,000 be taken out at the start of every year also introduces the risk of running out of money.[11]

We also need to reflect risk-aversion, the well-known human trait of fearing losses more than appreciating gains. From a retirement perspectve this translates to a willingess to impose a minimum withdrawal in exchange for raising the possibility of a shortfall. Utility functions are an alternative way to modelling loss aversion but we argue that it is easier for investors to understand dollar withdrawal limits. Absent other income, the minimum withdrawal is the essential pre-tax income required by the investor. The upper bound reflects the retiree's maximum spending needs. This maximum permitted withdrawal is a matter of investor preference: increasing it reduces the estate value but could leave room for living bequests. These bounds can be incorporated directly into a comprehensive financial plan that deals with all income sources and taxes.[12]

3.3 Measures of Risk and Reward

The risk being addressed is the risk of a shortfall in old age. This is often expressed as the probability of ruin, which is overly dramatic and a bit misleading. In Canada, government benefits (CPP/OAS) provide an income floor and retirees may have other sources of income. The probability of ruin also

---

[10] As pointed out by Fullmer and Sabin (2019), this just involves idiosyncratic longevity risk, not systematic longevity risk. Tontines as a rule do not provide investors with protection against the systematic component of longevity risk. If everyone in the pool dies at a rate that is 10% slower than anticipated, the mortality credits earned by most members will be correspondingly reduced.

[11] While \$40,000 may seem paltry, but it does not include government benefits. A typical Canadian 65-year old could receive about \$20,000 per year of such benefits, which are indexed to inflation. This would result in a minimum total cash flow of about \$60,000 per year, in real terms.

[12] Another reason to specify a maximum allowed withdrawal is to ensure that investors in a tontine are committed to the pool. If investors could simply withdraw the entire portfolio, there would potentially be a situation of moral hazard in which retirees who knew that they were likely to pass away soon could shift their portfolios from the pool to their estates.

fails to consider the dollar value of the shortfall. A 10% probability of failure implies that in one out of 10 cases the portfolio value is reduced to zero before the retiree dies. Whether this happens at age 75 or 85, for example, is not determined: we only know that the portfolio was depleted.

We use expected shortfall (ES) as the risk measure. ES is defined as the mean of the worst 5% of outcomes at the end of retirement. Some points to note about ES:

- The measure is evaluated at the end of the retirement period and is a dollar value. A larger value of ES is preferred: a scenario with an ES of −\$200,000 is worse than a scenario with an ES of −\$100,000. A positive ES implies that the average of the worst 5% of outcomes has a surplus at the end of the retirement period.
- The negative of ES is known as Conditional Value at Risk (CVaR). This is a measure of tail risk often used in the financial industry and by regulators. We use the worst 5% of outcomes, but other choices (e.g. 1% or 10%) are also possible.

Our measure of reward is total expected withdrawals (EW) over the retirement period: the more income the better. Points to note about EW:

- Retirees are not only concerned about income but also the variability of income. A complete retirement plan would account for the timing of other sources of income and pay attention to the sequence of withdrawals and the overall tax efficiency throughout the retirement period.
- Retirees may prefer a particular spending profile such as spending more earlier in retirement.

Considerations such as these are outside our discussion; we focus exclusively on the tradeoff between maximizing EW and maximizing ES. Since higher withdrawals lead to a higher risk of shortfall, our solution has to balance these two requirements. More specifically, we seek to maximize

$$\mathrm{EW} + \kappa \mathrm{ES}$$

where κ can be interpreted as a risk-aversion parameter, subject to the constraints imposed. Another way to think about κ is that it specifies how much weight to place on the ES objective relative to the EW objective. Finding the optimal solution for different values of κ traces out the efficient frontier, i.e. the best possible risk/reward tradeoff.

Requiring minimum annual withdrawals means that it is possible for the retiree's portfolio to be depleted entirely during the retirement period. If this happens, we assume that the retiree withdraws the minimum amount each year, with debt accumulating at an interest rate equal to the prevailing T-bill index rate plus a spread of 2%. This debt could perhaps ultimately be paid off via other means such as a reverse mortgage. This borrowing resulting from insolvency is the only form of debt that we allow: if the investor is solvent, leverage through either borrowing or short positions is not permitted.

Our analysis follows studies such as Bengen (1994) in specifying a 30-year horizon. This is a somewhat conservative assumption. In effect, we are asssessing financial situation of the investor conditional on surviving for 30 years in retirement. We are *not* evaluating the risk of financial ruin, i.e. running out of money. While that could be done by considering the probability of dying in each period, that is not our focus. Instead, we are simply evaluating the investor's position if they survive for 30 years under the three specified strategies.

3.4 Model Solution and Testing Procedures

Forsyth et al. (2022) fit a parametric model to monthly inflation-adjusted U.S. stock index and Treasury bill index data over the period from 1926 to 2020. The exact details of this parametric model are not required for present purposes, but it is worth noting that the model allows for large sudden jumps in asset values (i.e. market crashes) for both indexes. Based on estimated parameters, the optimal strategy in terms of asset allocation and withdrawals is found through sophisticated computational methods, the details of which are again outside our scope here. The main result of this effort, however, is just a pair of extensive tables that for each year over the retiree's 30-year investment horizon and for a large range of values of the retiree's portfolio value in each year specify how much should be withdrawn and how the remaining funds should be split between the two indexes. It is worth emphasizing that the method calculates a solution for the entire retirement period based on information available at the start of retirement. We would expect the model to be updated regularly with actual portfolio data to evaluate the retiree's financial position, as measured by anticipated withdrawals and expected shortfall.

Although Forsyth et al. (2022) provide extensive simulation tests of the asset allocation and withdrawal strategies based on the parametric model, it is more instructive here to concentrate only on a more stringent set of test results. These come from using *stationary block bootstrap resampling* of the historical data (see, e.g. Dichtl et al., 2016). This involves specifying an expected value for the size of a block of data, drawing a block length from a geometric distribution, randomly selecting a starting month, and collecting index returns for that block length. This is repeated with successive blocks of data being pasted together until 30 years of monthly data have been accumulated. This could mean, for example, drawing 24 months of data from the 1960s, followed by 31 months of data from the 1930s, and so on until a 30-year sample path has been constructed. This procedure is then repeated many times to produce a large number of simulated paths. The sampling is done simultaneously from both indexes and with replacement, so it is possible for a single simulated path to contain a block of data from say, the early 1980s with very high interest rates, and subsequent blocks of data that cover the same period. It is also possible to have drastic and sudden changes in the economic environment, as would happen if data sampled from a period of high interest rates such as the early 1980s was immediately followed by data from the 2010s with very low interest rates.

Bootstrap resampling can be viewed as a generalization of traditional back-testing, which often considers how an investment strategy would have performed over a specified horizon, starting at each possible date in a historical sample. For example, Bengen (1994)'s 4% rule was derived by considering 30-year periods of investment returns starting every year from 1926. This tells us how a strategy would have performed over 30 years if it had been initiated in 1926, in 1927, in 1928, etc. Obviously, there is a lot of overlap between nearby starting periods, so the results for these cases will be highly correlated. In addition, the number of different historical scenarios is quite limited: with a 30-year horizon, having even a century of return data will only lead to 71 different samples, many of which contain significant overlap. By breaking up the horizon into smaller periods and pasting together non-overlapping blocks of data, the bootstrap resampling procedure provides a much richer set of possible paths, encompassing many more possible outcomes that are not reflected in the actual historical record. We consider many outcomes that might have happened, rather than just those that did happen.

Retiree	Canadian male, age 65 at retirement
Horizon	30 years
Portfolio value at retirement date	\$1,000,000
Minimum annual withdrawal	\$40,000 (real, at start of each year)
Maximum annual withdrawal	\$80,000 (real, at start of each year)
Investable assets	CRSP U.S. T-bill index, CRSP U.S. total return stock market index (capitalization-weighted)
Portfolio rebalancing frequency	Annual
Borrowing spread	2% (only if portfolio is insolvent, otherwise no shorting or borrowing is permitted)
Available real estate hedge	\$400,000
Strategy #1:	
Annual withdrawals	\$40,000 at $t = 0,1,2,\ldots,29$ years
Investment policy	Fixed portfolio weights
Strategy #2:	
Annual withdrawals	\$40,000 – \$80,000 at $t = 0,1,2\ldots,29$ years
Investment policy	Optimal portfolio weights
Strategy #3:	
Annual withdrawals	\$40,000 – \$80,000 at $t = 0,1,2\ldots,29$ years
Investment policy	Optimal portfolio weights
Tontine gains	Based on CPM 2014 mortality table (from Canadian Institute of Actuaries)
Tontine fee	0.5% of the portfolio value at the start of each year

TABLE 1: *Summary of investment scenario and alternative strategies considered. Strategy #1 does not involve a direct optimization of an objective function. For Strategies #2 and #3, the retiree chooses the annual withdrawals and the portfolio weights to maximize* $EW + \kappa ES$ *where EW is expected total withdrawals, ES is expected shortfall after 30 years, and* κ *is a specified risk-aversion parameter.*

4 Results

This section presents the main results from the stationary block bootstrap tests of Forsyth et al. (2022). Unless otherwise stated, the results shown here are based on 1 million resampled paths with an expected blocksize of 24 months.[13] Withdrawals and rebalancing of investment portfolio weights occur at the start of each year, over an investment horizon of 30 years. The investment portfolio weights are rebalanced annually. Before providing the specific results, Table 1 summarizes the investment scenario and the three strategies considered. Both Strategy #2 and Strategy #3 involve maximizing $\text{EW} + \kappa\text{ES}$, where EW is expected total withdrawals, ES is expected shortfall after 30 years, and κ is a risk-aversion parameter to be specified. Various illustrative values of κ will be considered. Finally, we remind the reader that all quantities presented below are in real terms so, for example, the minimum and maximum withdrawal amounts of \$40,000 and \$80,000 respectively are adjusted each year for inflation.

---

[13] Some comparative results with other expected blocksizes can be found in Forsyth et al. (2022).

4.1 Strategy #1: Constant Asset Weights and Fixed 4% Withdrawals

The first set of results are summarized in Table 2, for a variety of constant asset weights. The first column of the table indicates the fraction of the retiree's portfolio that is invested in the equity market index, with the remaining fraction being invested in the T-bill index. The initial retirement portfolio is worth $1 million and withdrawals are fixed at 4% annually. Therefore, withdrawals are constant at 40 (in thousands of dollars) per year regardless of the equity weight, as shown in the second column of the table. The next column shows expected shortfall. Recall that this is the average of the terminal portfolio value for the worst 5% of simulated paths. All of the values of this column are negative, indicating a significant probability of portfolio depletion. Since the reward of total annual withdrawals) is constant, the best risk/reward tradeoff occurs where the risk is lowest (i.e. the maximum expected shortfall). In this instance, the "efficient frontier" is a single point, indicated by the shaded row of the table. This point occurs for an equity weight of 40% which results in an expected shortfall of about $-$305,000$. Higher or lower equity index allocations lead to reduced expected shortfall. The last column of the table shows the median terminal value of the investor's portfolio. This increases consistently with the allocation to the equity index, reaching a maximum of about $2.2 million (more than double the initial portfolio value) at the last equity weight shown of 80%. If the goal is to use retirement funds for retirement, it is inefficient to leave large sums as an unintended inheritance for fear of running out of money.

The scenario here is quite similar to that of Bengen (1994), but our results are considerably less optimistic. To recap, Bengen (1994) found that 4% was a safe withdrawal rate given an equally-balanced allocation to stocks and bonds (the retiree's portfolio was never depleted over a 30-year horizon), and this has been a common piece of financial advice over the past 30 years since his article appeared. In contrast, Table 2 shows that there is some chance of depletion for all constant weight portfolios considered. The 5% expected shortfall (recall that is the average outcome in the worst 5% of cases) is always negative, with a magnitude of one-third or more of the initial investment capital. What accounts for this? An obvious factor is that our data covers a much longer sample period with almost three additional decades, much of which contained periods of relatively low real interest rates.[14] However, our use of the stationary block bootstrap resampling approach is probably the main reason, as it provides a much more richer set of investment return paths.[15]

4.2 Strategy #2: Optimal Asset Weights and Variable Withdrawals

In this setting the retiree can adjust the portfolio composition over time on an annual basis and also has the flexibility to increase withdrawals. As indicated above, the retiree trades off reward (total withdrawals) against risk (expected shortfall) in order to determine how much to withdraw and what portfolio weights to use.

Table 3 presents the results for a range of values of the risk-aversion parameter κ. Expected annual withdrawals decrease with κ. This is because the investor puts more emphasis on expected shortfall as κ increases. As the last row of the table indicates, when κ is very high the investor

---

[14] Another difference is that we use a short-term Treasury bill index, whereas Bengen (1994) used intermediate-term Treasuries. To the extent that the yield curve was upward-sloping on average, we should expect somewhat worse performance than reported by Bengen.

[15] Along somewhat different lines, Anarkulova et al. (2022) use block bootstrap resampling on historical developed international market returns. They conclude that the withdrawal rate has to be reduced to just 2.26% provided that the retiree is willing to incur a 5% chance of financial ruin, assuming a 60-40 equity/bond asset allocation. While Anarkulova et al.'s use of non-U.S. data is part of the reason for the implication that the 4% withdrawal rate is much riskier than commonly thought, the resampling approach also contributes to this conclusion.

Equity Fraction	Annual Withdrawal ($ thousands)	Expected Shortfall (5%) ($ thousands)	Median Terminal Portfolio Value ($ thousands)
0%	40	−508.44	−155.04
10%	40	−418.02	−10.98
20%	40	−350.00	164.75
30%	40	−312.24	382.16
40%	40	−305.52	649.04
50%	40	−326.40	966.61
60%	40	−370.18	1,336.31
70%	40	−432.55	1,759.66
80%	40	−509.00	2,232.29

TABLE 2: *Results for Strategy #1. Based on 1 million resampled investment return paths with expected blocksize of 24 months and fixed withdrawals of $40,000 at the start of each year over the investment horizon of 30 years. Risk is measured as expected shortfall, which is the average terminal portfolio value for the worst 5% of sample paths. Reward is total annual withdrawals, which is constant at $1.2 million (i.e. withdrawals of $40,000 per year for 30 years), expressed on an annual basis here for simplicity. The shaded row indicates the best risk/reward tradeoff.*

simply withdraws the minimum amount of $40,000 per year, in an effort to maximize the expected shortfall. In this case, the expected shortfall is about −$219,000. However, this is inefficient: it is possible to achieve higher withdrawals for the same (or higher) expected shortfall.[16] For example, compare the last row of the table with that for $\kappa = 2.50$: this lower value of κ leads to about the same expected shortfall, but average annual withdrawals of about $56,000 rather than the minimum of $40,000. Alternatively, the value of $\kappa = 5.00$ leads to higher withdrawals (approximate annual average of $52,000) *and* higher expected shortfall (−$209,000) compared to the case where $\kappa \to \infty$. In fact, the last five rows of Table 3 all represent inefficient points, and this is highlighted by shading these rows.

Comparing Tables 2 and 3, it can be seen that allowing for higher (but flexible) withdrawals and departing from fixed asset allocation weights leads to significant improvements. The best case from Table 2 has withdrawals pinned to the minimum value of $40,000 per year and an expected shortfall of about −$306,000. The same level of withdrawals leads to an expected shortfall of −$219,000 with the varying asset allocation. As observed above, this is actually inefficient since higher expected shortfall and higher average annual withdrawals can be achieved. Alternatively, from Table 3 with $\kappa = 1$ we have an expected shortfall of −$290,000, a slight improvement over the minimum expected shortfall with constant weights of −$306,000. However, the average annual withdrawal when $\kappa = 1$ in Table 3 is almost $62,000, more than 50% higher than the best case from Table 2. The median terminal portfolio values are generally much bigger for the constant weight strategy, especially if the allocation to equity is high. Part of the reason is the higher average annual withdrawals for Strategy #2. In addition, Strategy #2 actively tries to protect against low

---

[16] Our strategy is determined by using the parametric market model. Under this model, all strategies for different values of κ are efficient. However, recall that the results in Table 3 correspond to "out-of-sample" bootstrap resampling tests. In this case, some of the efficient strategies become inefficient out-of-sample. This implies that a retiree should stick with strategies which are efficient both in-sample and out-of-sample.

κ	Average Annual Withdrawal ($ thousands)	Expected Shortfall (5%) ($ thousands)	Median Terminal Portfolio Value ($ thousands)
0.18	69.91	−805.65	−31.84
1.00	61.77	−290.03	−40.87
1.50	59.21	−248.15	−77.26
1.75	58.16	−235.46	−78.50
2.50	56.02	−219.00	−81.84
3.75	53.78	−209.90	−80.68
5.00	52.43	−207.15	−77.25
6.25	51.74	−209.02	−78.11
7.50	51.26	−210.38	−78.48
10.00	50.58	−212.41	−77.95
100.00	47.72	−217.82	−67.91
∞	40.00	−219.16	17.34

TABLE 3: *Results for Strategy #2. Based on 1 million resampled investment return paths with expected blocksize of 24 months and withdrawals of at least $40,000 and at most $80,000 at the start of each year over the investment horizon of 30 years. Risk is measured as expected shortfall, which is the average terminal portfolio value for the worst 5% of sample paths. Reward is total annual withdrawals, which are converted here to an average annual basis for simplicity. The shaded rows of the table represent inefficient points as it is possible to achieve higher average annual withdrawals for the same (or higher) expected shortfall.*

expected shortfall by adjusting the asset allocation and withdrawals. Finally, we emphasize that the median terminal portfolio value is simply a consequence of following the specified allocation and withdrawal strategies: it is not part of the objective which trades off the reward of total withdrawals against risk as measured by expected shortfall.

4.3 Strategy #3: Optimal Asset Weights, Variable Withdrawals, and a Tontine Overlay

We now expand the setting of Strategy #2 with the tontine overlay, which offers the additional potential benefit of earning tontine gains. As noted previously, we assume that investors in the tontine pool pay annual fees of 50 basis points as a percentage of the values of their portfolios.

Table 4 provides the results. Similar to Table 3, there are some inefficient points for high values of κ, as indicated by the shaded rows of the table. For example, consider the case with $\kappa = 10$, which results in average annual withdrawals of $49,420 and an expected shortfall of about $1.276 million. This is inefficient since, for example, average withdrawals of $54,500 and an expected shortfall of $1.285 million can be achieved by setting $\kappa = 1$. However, the main point is the dramatically improved results when the tontine overlay is included (Table 4) compared to when it is not (Table 3). Although the values of κ are somewhat different between the tables, the range of average annual withdrawals is quite comparable, from a high of about $70,000 (representing a real average annual withdrawal rate of around 7%, almost double the 4% rule) to an (inefficient) low of $40,000. However, the expected shortfall numbers are much higher with the tontine overlay, as is the median expected terminal portfolio value. Keep in mind that with the tontine overlay,

κ	Average Annual Withdrawal ($ thousands)	Expected Shortfall (5%) ($ thousands)	Median Terminal Portfolio Value ($ thousands)
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	1,038.07
0.25	63.84	863.20	1,500.51
0.30	62.08	1,011.55	1,739.21
0.50	58.13	1,211.18	2,115.22
1.00	54.50	1,285.93	2,330.33
10.00	49.42	1,275.98	2,485.58
∞	40.00	1,280.97	2,982.41

TABLE 4: *Results for Strategy #3. Based on 1 million resampled investment return paths with expected blocksize of 24 months and withdrawals of at least $40,000 and at most $80,000 at the start of each year over the investment horizon of 30 years. Risk is measured as expected shortfall, which is the average terminal portfolio value for the worst 5% of sample paths. Reward is total annual withdrawals, which are converted here to an average annual basis for simplicity. The shaded rows of the table represent inefficient points as it is possible to achieve higher average annual withdrawals for the same (or higher) expected shortfall. The investor participates in a tontine pool, with an annual fee of 50 basis points.*

the portfolio value upon the retiree's death goes to other investors in the tontine pool, not to the retiree's estate. To take a specific point of comparison, without the tontine overlay average yearly withdrawals of about $54,000 correspond to an expected shortfall of about −$210,000 and a median terminal portfolio value of −$81,000; see the row of Table 3 with $\kappa = 3.75$. If we include the tontine overlay, average annual withdrawals of around $54,000 are associated with an expected shortfall of almost $1.29 million and a median terminal portfolio value of $2.33 million; see the row of Table 4 with $\kappa = 1$.

An overall comparison of the three strategies is provided in Figure 1, which plots expected shortfall on the horizontal axis and average annual withdrawals on the vertical axis. Since we are seeking to maximize both average annual withdrawals and expected shortfall, we want to move as far as possible toward the upper right of the plot. Since Strategy #1 specifies constant annual withdrawals of $40,000, there is just a single efficient point for this strategy (all other possible points would lie to the left of the one shown, with the same withdrawals and lower expected shortfall). By allowing for variable withdrawals and flexible asset allocation, Strategy #2 achieves higher average withdrawals and some possible minor improvement of expected shortfall. In turn, adding the tontine overlay (Strategy #3) shifts the frontier significantly to the right, giving much higher expected shortfall for comparable withdrawals. Overall, participation in the tontine pool appears to provide solid insurance against portfolio depletion while simultaneously offering an attractive average annual withdrawal rate.

A potential concern is that all results presented thus far assume an expected blocksize of 24 months in the resampling procedure. Are our main conclusions robust to different expected block-

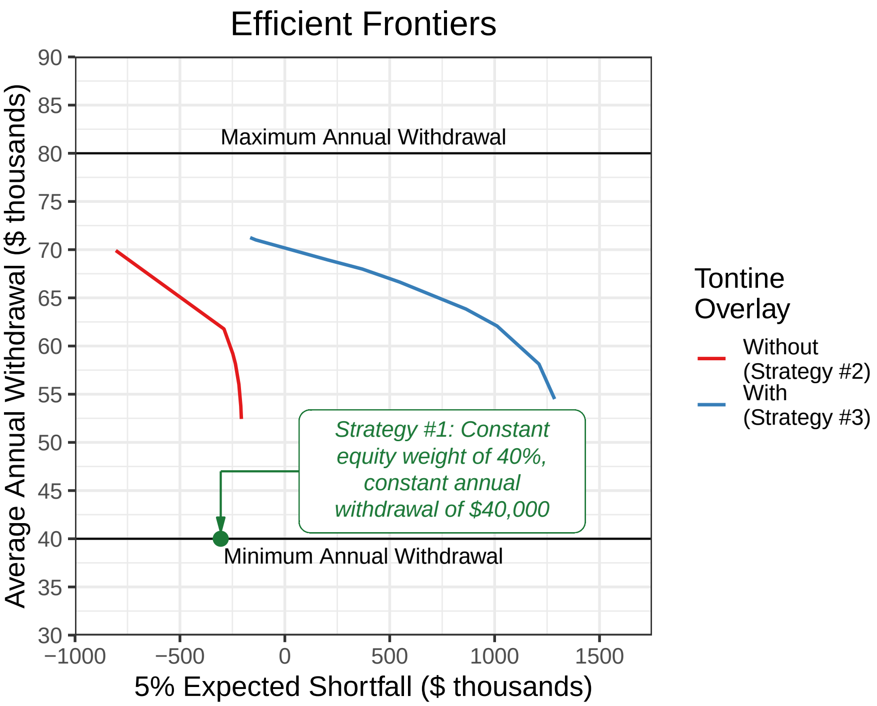

FIGURE 1: *Efficient frontiers for the three different strategies. Based on 1 million resampled investment return paths with expected blocksize of 24 months. There is just a single efficient point for Strategy #1, as indicated by the labelled point.*

sizes? Figure 2 plots the efficient frontiers for Strategy #3 with expected blocksizes of 6, 12, 24, and 60 months. We see very little difference between the frontiers with expected blocksizes of 6 and 12 months. The 24 month case and the 60 month case lie further to the upper right, indicating a better risk/reward tradeoff. On the whole, however, the plot is quite reassuring: while there is clearly some sensitivity to the assumed expected blocksize, the differences are fairly minor, and certainly insufficient to raise doubts about the main conclusion that the tontine overlay in conjunction with flexible asset allocation and withdrawals provides quite effective longevity insurance and attractively high average withdrawals.

5 The Tontine Overlay: A Closer Look

The results above suggest that the tontine overlay potentially offers some attractive features to retirees. We now turn to a more detailed examination of its characteristics. Throughout this section we assume that the risk-aversion parameter $\kappa = 0.18$. The results presented would obviously vary somewhat for alternative values of κ, but the general characteristics would be similar. We begin with the annual withdrawals. Recall that in our illustrations we allow the investor to withdraw each year any amount between $40,000 and $80,000. Figure 3 plots the optimal withdrawals determined from solving the parametric model as a function of portfolio values as a heatmap.[17] For the time being, ignore the superimposed red lines and observe that the underlying heatmap shows that it is

---

[17] As mentioned above (see Section 3.4), the model solution procedure generates detailed tables showing optimal withdrawals and asset allocation over time across a range of portfolio values. The heatmaps given here are a graphical depiction of these tables.

FIGURE 2: *Efficient frontiers for Strategy #3. Based on 1 million resampled investment return paths with various expected blocksizes.*

generally optimal to withdraw either the minimum amount of $40,000 or the maximum amount of $80,000, although there is a very narrow range that is hard to discern on the plot where it is optimal to withdraw an intermediate amount between the maximum and minimum permitted withdrawals. The minimum amount should be withdrawn when the portfolio value is relatively low, with the transition to the maximum amount occurring at steadily decreasing portfolio values over time.[18] Given the initial portfolio value of $1 million (i.e. 1,000 in the plot since units are thousands of dollars), the first withdrawal at year 0 should be the minimum of $40,000. Subsequent withdrawals will depend on investment returns, which of course vary randomly. The superimposed red lines give an indication of the distribution of the portfolio value, based on 1 million resampled investment return paths with expected blocksize of 24 months. The solid line is the median of the portfolio value distribution, after withdrawals at the start of each year.[19] This median line generally trends downward over time, until the last few years, staying close to the boundary between the region where it is optimal to withdraw the minimum and the region where it is best to withdraw the maximum. During the first few years, it lies in the area corresponding to minimal withdrawals, but thereafter it is in the maximal withdrawal zone. This indicates that the retiree should expect to be conservative, withdrawing the minimum amount at the start of retirement and, assuming decent investment performance, transitioning to withdrawing the maximum amount down the road. The

---

[18] Forsyth et al. (2022) show that with continuous rebalancing, the optimal withdrawal is a "bang-bang" control, i.e. it is always optimal to withdraw either the minimum or maximum permitted amount. Continuous rebalancing is clearly unrealistic, but our results indicate that the same conclusion generally holds in our more realistic setting with annual rebalancing.

[19] The line starts out at 960 because the initial withdrawal of $40,000 reduces the portfolio value from $1 million to $960,000 at the outset.

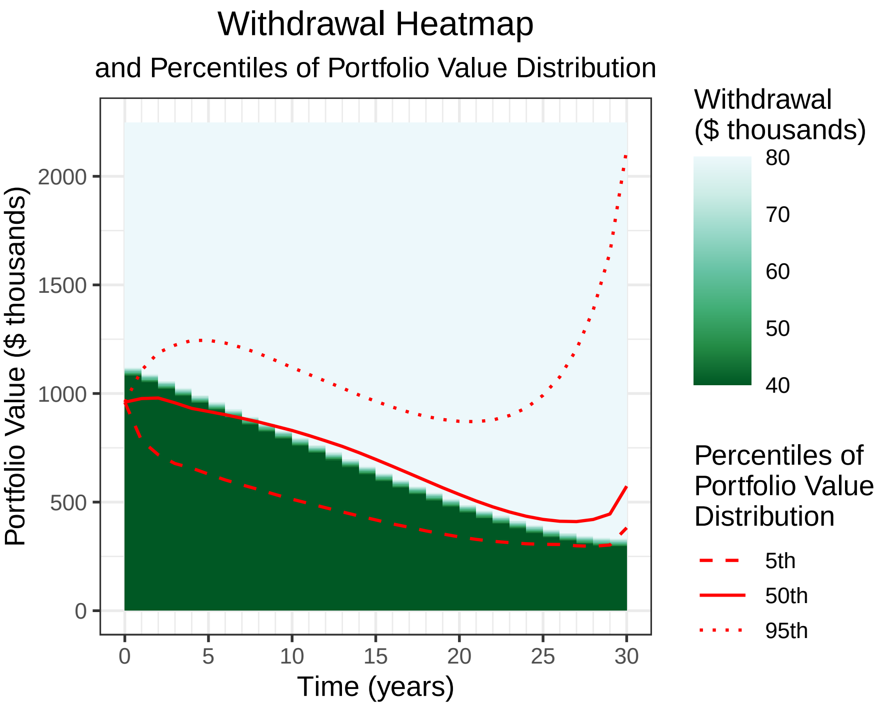

FIGURE 3: *Heatmap of withdrawals as a function of portfolio value over time for* $\kappa = 0.18$. *The percentiles of the portfolio value distribution superimposed on the heatmap are based on 1 million resampled investment return paths with expected blocksize of 24 months.*

upper red dotted line represents the 95th percentile of the portfolio value distribution. If the investment experience is exceptionally good, then the retiree can switch very early to maximum annual withdrawals and sustain them, with the portfolio value still rising significantly in the last few years due to the combination of favourable market returns and tontine gains, which become more significant in the later years. The lower red dashed line depicts the 5th percentile of the portfolio value distribution, and it shows that with really poor investment portfolio returns the retiree will have to make do with the minimum withdrawal of $40,000 throughout virtually the entire retirement horizon, despite the increased tontine gains in the later years.

Figure 4 provides a slightly different perspective by directly plotting the 5th, 50th, and 95th percentiles of the distribution of withdrawals over time. For the median case (i.e. 50th percentile), withdrawals start out at the minimum value of $40,000 for the first couple of years but quickly increase, reaching the maximum of $80,000 by the start of the 5th year and remaining at this level throughout the balance of the retirement period. The 95th percentile follows a similar pattern, reaching the maximum withdrawal even faster (by the start of the second year). However, the 5th percentile remains at the minimum withdrawal of $40,000 until very late in the retirement period. There are no guarantees here: while it is likely that the retiree will at some point be able to increase withdrawals, it can take quite a long time for this to happen if the investment returns are exceptionally poor.

It is also instructive to consider the investment allocation over time. Figure 5 provides a heatmap of the fraction optimally allocated to the equity market index over time as a function of the portfolio value. Ignoring the superimposed red lines for the moment, it is apparent that there are basically three zones:

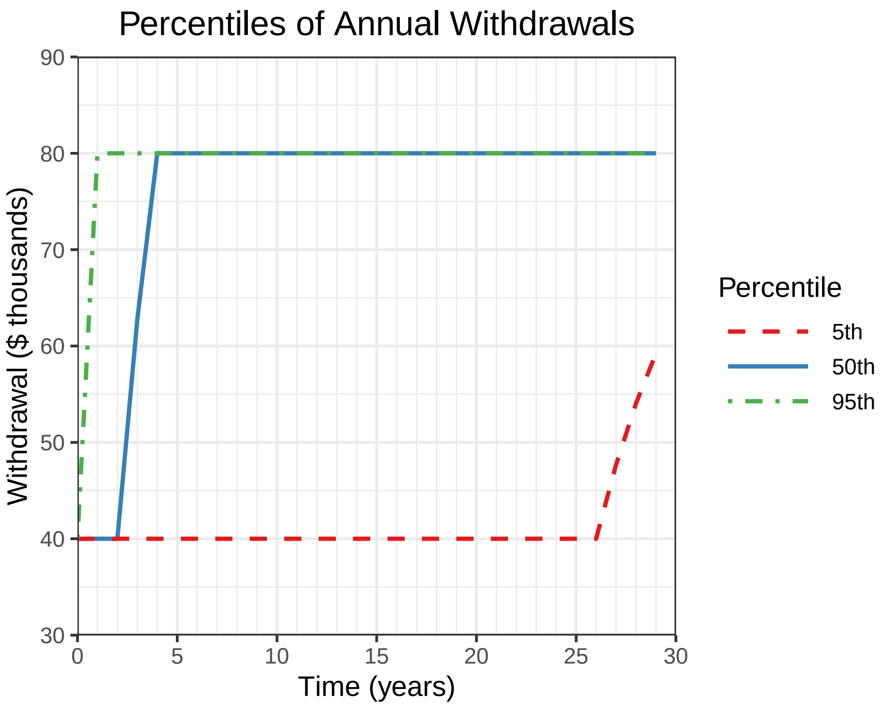

Figure 4: *Percentiles of withdrawals over time for $\kappa = 0.18$. Based on 1 million resampled investment return paths with expected blocksize of 24 months.*

- If the portfolio value is quite low, it is optimal to invest entirely in the equity index. The reason is that even with minimum withdrawals, the expected shortfall is very low (i.e. it is quite likely that the portfolio value will become negative). The only way to avoid this is to invest in the riskier equity index and hope for a recovery through strong returns.
- If the portfolio value is quite high, it is optimal to invest entirely in the T-bill index. There is very little chance that the portfolio will become insolvent, even if the maximum withdrawal is taken out each year. In essence, the retiree has been fortunate and can easily afford the maximum withdrawals without needing to have a relatively risky portfolio composition.
- For moderate portfolio values, it is optimal to split the portfolio between the equity index and the T-bill index, with a higher allocation to the former if the portfolio value is relatively low.

Overall, the optimal asset allocation weights and the tontine are complementary: varying the allocation weights is most effective early in retirement when mortality credits are smallest, but the converse is true as the retiree ages.

The three red lines in Figure 5 correspond to the 5th (dashed line), 50th (solid line), and 95th (dotted line) percentiles of the portfolio value distribution from 1 million resampled paths with an expected blocksize of 24 months. All three lines start out at year 0 with an allocation to equities that is close to 60%. The median (50th percentile) case involves a gradual decline in portfolio values until the last few years of the retirement period, with an eventual transition to being entirely invested in the T-bill index. The 95th percentile case shows an increase in portfolio value over the first few years, followed by a gradual decline and then a sharp increase over the

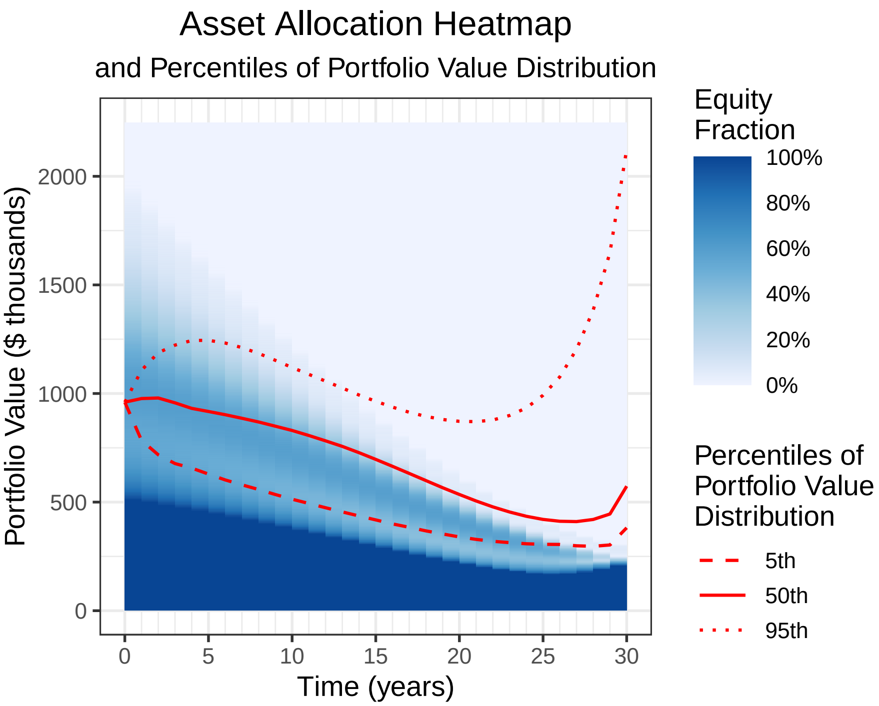

FIGURE 5: *Heatmap of asset allocation (expressed as the fraction invested in the equity index) as a function of portfolio value over time for* $\kappa = 0.18$*. The percentiles of the portfolio value distribution superimposed on the heatmap are based on 1 million resampled investment return paths with expected blocksize of 24 months.*

last few years. The 5th percentile case has a consistently declining trend until the very end of the retirement period, but never crosses into the lowest zone where it is optimal to invest entirely in the equity index.

Finally, Figure 6 shows the percentiles of the distribution of the fraction of portfolio value that is optimally allocated to the equity index over time. The allocation to the equity index begins at slightly under 60%, and generally trends downward over time. The 5th percentile of the distribution declines quickly, reaching 0% (i.e. the portfolio is entirely invested in the T-bill index) before year 15 and remaining there. The median allocation to the equity index trends downward at a slower rate, but reaches 0% in the last few years. Of course, by this point the expected tontine gains are quite significant, so sufficiently high anticipated overall returns are available without taking on much additional equity market risk. For example, at age 90 the tontine mortality credit is about 14% per year, and this increases to around 33% per year at age 95. The 95th percentile of the equity allocation stays close to 60% for about two decades, before dropping rapidly. Overall, the picture from Figures 5 and 6 is reasonably attractive for retirees who may not want a large exposure to equity markets, as the chance of needing to be completely invested in equities at any point is less than 5%.

6 Additional Observations

The illustrative example considered above is just one possible form of tontine. In our case, the tontine has a total payout of portfolio market returns and mortality credits to fund withdrawals.

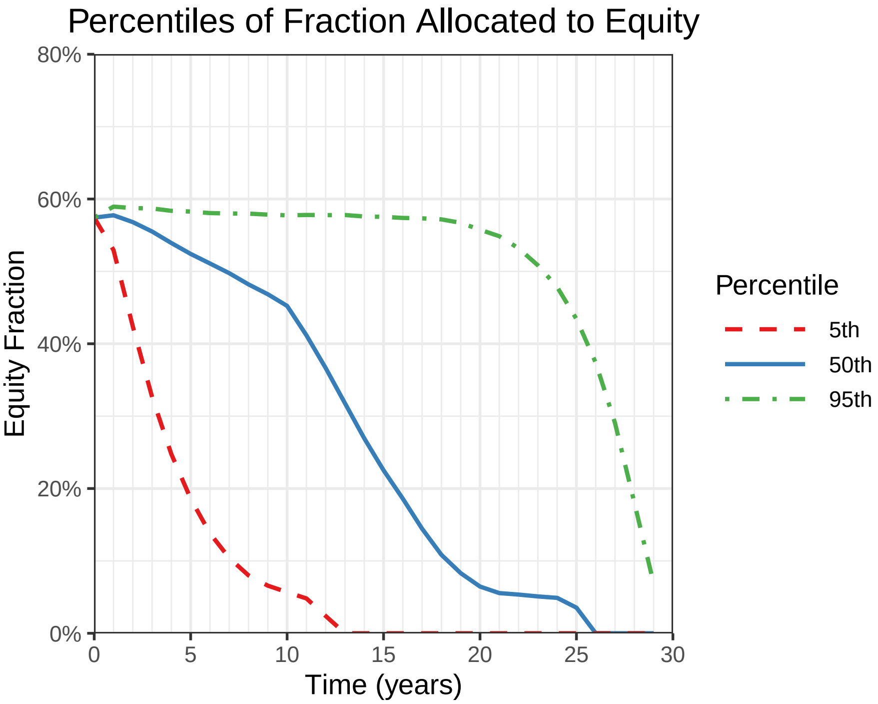

Figure 6: *Percentiles of optimal equity fraction distribution over time for* $\kappa = 0.18$. *Based on 1 million resampled investment return paths with expected blocksize of 24 months.*

Other options could have accumulation of all proceeds and a lump sum payment to survivors after a specified period of time, or paying out all gains annually without limitation. In practice, some tontines offer a money back guarantee such as a return of the initial investment in nominal terms minus any withdrawals until the time of death. Such additional guarantees would have to be hedged, reducing tontine benefits.

Whatever the particular structure of a tontine, the most obvious objection is that there is no terminal value upon death. However, other retirement products have similar constraints: a lifetime annuity, a DB pension plan, or OAS in Canada have no or limited survivor benefits. The purpose of the tontine is to provide additional income for older investors. It should be used in combination with other investment strategies to meet the investor's goals. As noted retirement finance expert Moshe Milevsky has observed, "if you give up some of your money when you die, you can get more when you're alive".[20] This is in direct contrast to term insurance where premiums are paid while you are alive to benefit others when you are dead.

Government benefits, DB pension plans, and lifetime annuities all provide guaranteed lifetime income. A modern tontine does not provide guaranteed income but benefits from mortality credits, an exposure to market returns and possibly a redemption value. Conventional income funds offer no guarantees and rely solely on the market performance of the underlying securities.

What factors should be considered by investors (or their advisors) when deciding how much, if any, to allocate to a modern tontine? A reasonable starting point is to assess if a retiree has a significant expected shortfall by the end of their retirement using their current investment mix. Although traditional financial planning tools do not compute expected shortfall directly, it is not difficult to estimate. The shortfall could be addressed by home equity as a contingent asset; in many financial planning scenarios the principal residence is assumed to be part of the estate. Illiquidity, transaction costs, bequests, and a preference for aging in place can be barriers to using home equity. An allocation to a modern tontine may be an attractive way to address an expected shortfall.

Tontine products currently available in Canada are generally split into age cohorts. Some focus on enhanced payouts, while others offer a lump sum payment after 20 years, for example. Which is more appropriate depends on the retiree's situation and concerns. Consider the following two examples:

- Michelle would like a higher income stream throughout retirement. Michelle is divorced and wants to stay in her home for as long as possible. Her adult children have no interest in benefitting from an estate. Michelle could consider a decumulation fund appropriate for her age cohort that blends a diversified investment fund with a tontine layer. If instead Michelle was married and likely through age or health to predecease her spouse, then the redemption value of the fund becomes an important consideration. She should contrast the benefit of a tax-free spousal rollover of registered assets with the potential loss of capital with a tontine product.

- Sanjay, age 65, is more concerned about running out of money in his 80s and beyond. He doesn't want to commit to using home equity. Sanjay might consider an accumulation tontine that pays out after a certain period such as 20 years. He could allocate something like $50,000 in his TFSA and benefit from tax-free capital growth. The tontine could be incorporated into Sanjay's financial plan as a lump sum payout with an appropriate estimate of the range of possible payouts. Although he would not directly derive income from the tontine, he could modify the asset allocation for his other investments to boost his expected income in the early

---

[20] Quoted in New type of pension plan funnels more cash to retirees while they're alive - Toronto Star, April 13, 2021.

years of retirement while maintaining an acceptable level of expected shortfall. Sanjay may consider his tontine investment as a form of insurance; if markets perform poorly, he knows that he will get a boost in his mid 80s to carry him through his final years.

For both Michelle and Sanjay, monitoring the risk of portfolio depletion as retirement progresses is crucial, especially in the early years when sequence of returns risk has the largest impact.

7 Summary

It has often been observed that the shift from DB to DC pension plans places much more responsibility on plan members, who need to make investment choices both pre- and post-retirement and have the additional burden of managing portfolio withdrawals during retirement in the face of longevity risk. Traditional solutions such as the benchmark 4% withdrawal rule combined with constant asset allocation weights (Bengen, 1994) appear to be much riskier than commonly perceived, especially if one considers either non-U.S. data (Anarkulova et al., 2022) or block bootstrap resampling of historical market returns (or both).

The first strategy considered in this article is essentially the traditional 4% rule for withdrawals, with various alternative fixed portfolio weights for the investment portfolio. Despite the relatively low level of withdrawals, the expected shortfall (i.e. terminal portfolio value on average in the worst 5% of cases) is significantly negative. It is clear that the initial portfolio value of $1 million does not guarantee that inflation-adjusted annual withdrawals of $40,000 can be sustained over a 30-year retirement horizon.

Strategy #2 allows the investor to optimally adjust the portfolio weights and withdraw up to $80,000 per year, with a minimum withdrawal of $40,000. While there is some notable improvement over Strategy #1 in terms of average annual withdrawals, the expected shortfall is still significantly negative.

Strategy #3 adds the tontine overlay to Strategy #2, and shows strong improvement. Real annual withdrawal rates of close to 7% appear to be possible, with a positive expected shortfall. The tontine overlay plus the flexibility to adapt portfolio weights and withdrawals to evolving circumstances (as in Strategy #2) appears to provide very effective insurance against portfolio depletion in retirement. We have used a pure form of tontine, without any redemption value or lump sum payouts. Such additional features would reduce the efficacy relative to the case considered here.

We emphasize that our analysis concentrates on how effective the insurance provided by the tontine is, assuming that it is needed (i.e. the investor has a long retirement period). Of course, a missing element from the overall picture is the cost of this insurance. While we have included an annual fee of 50 basis points to administer the tontine, the much larger consideration would be the reduction in the estate's portfolio value for tontine pool members who die relatively early. This is what supports the increased withdrawals for long-lived members.

It is also worth recalling a point we have mentioned earlier: the tontine structure provides insurance against idiosyncratic longevity risk, but not against systematic longevity risk. Insurance against both types of risk can be purchased with annuities, but these are quite costly and to date have not been popular with retirees.

There are many alternative ways to structure a tontine pool to provide longevity insurance. We have considered just a single case here as a proof of concept. We have assumed a large open-ended pool and a fair tontine, somewhat along the lines described by Fullmer (2019) and Fullmer and Sabin (2019), with investors retaining the ability to select their own portfolio allocations and having some flexibility about the scope of annual withdrawals. An encouraging result is that it appears that notably higher real withdrawal rates can be sustained without incurring signficant risk of portfolio depletion in retirement. In addition to providing an important benefit to retirees, this also provides a strong boon to individuals currently saving for retirement: higher real withdrawal rates mean that less money needs to be accumulated for retirement, so that more employment income can be used for other purposes.

We have used an idealized model to show the potential of the modern tontine to enhance retirement income and reduce shortfall risk. To effectively incorporate a modern tontine into holistic retirement planning we use total (or average) expected withdrawals and expected shortfall as key measures of reward and risk.

Modern tontines may be valuable for retirees who must increasingly rely on their own resources to plan retirement. There is no universal retirement asset, but the modern tontine complements current tools and has recently made an appearance in the Canadian investment landscape.

Investment products like modern tontines or traditional annuities require targeting and explanation to those who would most benefit. Retirees whose desired expenditure requires sustained withdrawals from investments in excess of expected real returns would be a natural clientele for modern tontines. With the advent of higher inflation, longer retirement, and diminished anticipated investment returns, this market is likely to expand.

Acknowledgements:

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

References

Anarkulova, A., S. Cederburg, M. S. O'Doherty, and R. Sias (2022). The safe withdrawal rate: Evidence from a broad sample of developed markets. Working paper, Eller College of Management, University of Arizona.

Bégin, J.-F. and B. Sanders (2023). Benefit at risk in lifetime pension pools. Society of Actuaries Research Institute and Canadian Institute of Actuaries.

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

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

Forsyth, P. A., K. R. Vetzal, and G. Westmacott (2022). Optimal performance of a tontine overlay subject to withdrawal constraints. Working paper, available at `https://arxiv.org/abs/2211.10509`.

Fullmer, R. K. (2019). Tontines: A practitioner's guide to mortality-pooled investments. CFA Institute Research Foundation Brief.

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

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 Iinstitute.

Milevsky, M. A. (2015). *King William's Tontine: Why the Retirement Annuity of the Future Should Resemble Its Past.* Cambridge University Press.

Milevsky, M. A. (2022). *How to Build A Modern Tontine: Algorithms, Scripts and Tips.* Springer.

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

Milevsky, M. A., T. S. Salisbury, G. Gonzalez, and H. Jankowski (2018). Annuities versus tontines in the 21st century: A Canadian case study. Society of Actuaries.

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

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

Statistics Canada (2022). Pension plans in Canada, as of January 1, 2021. Available online at https://www150.statcan.gc.ca/n1/daily-quotidien/220718/dq220718a-eng.htm.

Thinking Ahead Institute (2023). Global Pension Assets Study. Available online at https://www.thinkingaheadinstitute.org/research-papers/global-pension-assets-study-2023/.