Department of Economics and Finance

Master Degree in Economics and Finance

Chair of Asset Pricing

Rethinking Retirement Withdrawal Rates:

Evaluating the Optimality of Tontines for Decumulation

Supervisor

Nicola Borri

Co-supervisor

Matteo Paradisi

Candidate

Alessandro Ghirardi

770741

Academic Year 2023/2024

Abstract

This thesis examines decumulation strategies for early retirees, proposing an alternative approach. The study first evaluates the safety of spending strategies based on constant withdrawals in real terms, such as the '4% rule'. Secondly, it explores the potential of structuring a Natural Tontine Fund as an effective solution for wealth decumulation. The findings suggest that the '4% rule' overestimates the safe withdrawal rate by approximately 75 basis points. Additionally, extending the retirement period beyond 30 years requires reducing success rates for all withdrawal rates, especially with a high bond allocation. Moreover, a Natural Tontine Fund emerges as an efficient decumulation tool, offering retirees a lifetime income stream with lower risk of failure, higher median payments, and more effective wealth decumulation compared to a constant-amount spending plan.

Contents

List of Figures

Introduction

"Retirement Income: The nastiest, hardest problem in finance"

— William Sharpe

Many countries are encountering a significant demographic shift. As mortality rates decline and the elderly population expands, birth rates decrease, leading to a reduced younger population. In many European countries, public pension systems are structured as pay-as-you-go plans, where current pensions are financed through taxes on the working population. This increasing ratio of retirees to workers increases the instability of these pension systems, with the potential risk that governments may be unable to guarantee stable income for retirees in the future. Considering these trends, there is a growing necessity for individuals to assume greater responsibility for their future retirement income.

Even though planning and saving for retirement may prove to be a daunting task, putting money aside for the future is a challenge that pales in comparison to the difficulties that arise when it comes time to decumulate those savings. Nobel Laureate William Sharpe described the decumulation of wealth during retirement as one of the toughest challenges in finance and declared that there is no 'retirement number' that guarantees financial security[1].

However, a simple approach to retirement spending is often suggested. In a review of the 50 most popular personal finance books, Choi (2022) found that, of the 12 books providing explicit retirement spending advice, seven recommend a 4% spending rule, while four recommend an even higher rate. The '4% rule', which suggests that retirees can safely spend 4% of their investment portfolio in the first year and then

[1] Source: William Sharpe: There Is No Retirement 'Number'

adjust withdrawals according to the inflation rate for the rest of their lives, was initially proposed by William Bengen. In a pivotal study of the field of retirement planning (see Bengen 1994), the author attempted to determine how much retirees can safely spend in retirement and concluded that a 4% initial withdrawal rate, followed by inflation-adjusted withdrawals, could ensure that a retiree's portfolio would last for a 30-year retirement period. This spending rule gained significant popularity due to its simplicity and ease of application; however, distilling a complex problem into a simple rule of thumb has serious flaws, especially when dealing with early retirement.

A 2020 survey published by Vanguard[2] highlighted that 22% of millennials aim to retire before age 60, and the growing interest in *early retirement* is also reflected in the countless books, blogs, and social media accounts dedicated to promoting the values of the F.I.R.E. movement. The concept of Financial Independence, Retire Early (F.I.R.E.) was introduced by the book *'Your Money or Your Life'*, published in 1992 by Vicki Robin and Joe Dominguez. Adherents of the F.I.R.E. movement aim to live a frugal life in order to save and invest a large portion of their salary, with the goal of building an investment portfolio large enough to allow them to retire much earlier than the conventional retirement age.

Early retirees need to establish a spending strategy for a retirement period that will last much longer than the 30 years considered in Bengen's study, and the 4% rule may prove inadequate. Extrapolating the withdrawal rate estimated to be safe over 30 years and simply applying it to a 50- or 60-year retirement period can be flawed and could potentially increase the probability of outliving one's wealth. Furthermore, the 4% rule is based on assumptions that may overestimate success rates, such as neglecting taxes and transaction costs. This study aims to consider factors that are relevant for early retirees to assess their impact on safe withdrawal rates.

Retirees face the risk of living much longer than initially expected and, hence, not having enough resources to cover their most basic needs. Life expectancies exhibit considerable variability, and relying on the simple average is not a wise strategy. Therefore, retirees must either plan for the worst-case scenario—e.g., living up to 110 years—or pool their mortality risk with other retirees. By considering a sufficiently large group of people, some individuals will live longer than others, but it can be expected that

[2]Source: Fuel for the F.I.R.E.: Updating the 4% rule for early retirees

the average will converge to the mean life expectancy. Pooling mortality risk allows for planning based on the average life expectancy of the group rather than for the worst-case scenario.

Many insurance products, such as annuities, are based on the concept of mortality pooling. Research shows that full annuitization of a retiree's wealth can address the risk of outliving their assets (see Davidoff, Brown, and Diamond 2003); however, annuities have lost popularity, and few retirees purchase these products today (see Poterba, Venti, and Wise 2011). The idea of this thesis is to use a financial instrument from the 17th century, the *tontine*, to solve the modern problem of decumulation during retirement. In a tontine, a group of investors decides to pool their funds, and the sponsor of the tontine guarantees a periodic payment that must be divided among the remaining survivors in the fund. Unlike an annuity, where the coupon received by each individual is constant but the total payment made by the sponsor each year depends on the number of survivors, in a tontine, the sponsor makes a fixed payment, and the coupon received by each individual depends on the remaining number of participants in the fund. Therefore, the tontine guarantees a lifetime stream of income, but as the number of participants decreases over time due to deaths, the coupon received by each survivor increases and can vary widely. The *Natural Tontine* structure proposed by Milevsky and Salisbury (2015) aims to overcome the issue of increasing coupons by paying a non-constant Tontine Dividend Rate (the percentage of the fund value paid by the sponsor in each period) to allow survivors to obtain a constant stream of income. The Natural Tontine is a promising investment product, and this thesis investigates possible ways to structure a fund based on this product that could be a viable investment option for early retirees.

The '4% rule' exemplifies the divergence between finance theory and practical advice. While evidence shows that retirees should annuitize their wealth to eliminate the risk of outliving their assets, the most common advice is to follow a simple constant withdrawal strategy. This thesis aims to contribute to the discussion in the early retirement literature, highlighting the flaws of distilling a complex problem into a simple rule and showcasing how early retirees could benefit from advancements in the asset management and insurance industries.

The remainder of the thesis is organized as follows. Chapter 1 is devoted to the analysis of safe withdrawal rates. The chapter begins with an introduction to withdrawal rate strategies and their associated risks. Subsequently, it examines the impact of the assumptions underlying key studies in the field on success rates and identifies sources of inefficiency in these strategies. Chapter 2 focuses on tontines. It starts with an overview of the concept and history of tontines. The chapter then examines the Natural Tontine structure, exploring how a fund based on this product can be constructed. In Chapter 3, the feasibility of structuring a Natural Tontine Fund as a desirable investment tool for early retirees is assessed.

Chapter 1

Safe Withdrawal Rates, An Empirical Analysis

1.1 What are Safe Withdrawal Rates?

The analysis of safe withdrawal rates aims to estimate how much retirees can withdraw from their investment portfolio each year of retirement to cover their needs without worrying about prematurely running out of money. The primary concern for retirees is outliving their own wealth, making it essential to determine a spending plan that can sustain their lifestyle while managing two main sources of risk: longevity risk (the risk that the retiree's lifespan exceeds initial or actuarial expectations) and return risk (the risk of receiving poor real investment returns during retirement).

The approach employed by financial economists to determine the optimal withdrawal rate for a retiree, such as the ones proposed by Hakansson (1970), Merton (1969), or Samuelson (1969), involves specifying a utility function and finding the rate of withdrawals that maximizes it. Although this approach is theoretically sound, it is not the standard practice among retirement planning professionals due to two main drawbacks. First, this strategy requires accurately specifying an explicit utility function for each retiree. Secondly, to simplify the solution, these models often assume that each withdrawal is independent of previous ones, leading to extreme volatility in the withdrawals. Outside the academic environment, practitioners estimate withdrawal rates for their clients under the assumption that they wish to maintain constant withdrawals in real terms throughout the entire retirement period to ensure a stable standard of living.

Retirees generally aim to set an initial amount of money to withdraw in the first year of retirement and then continue withdrawing the same amount in real terms from their investment portfolio each subsequent year. For example, a fictional retiree may determine that they need to spend \$40,000 in the first year. If the retiree wishes to maintain this purchasing power in real terms, they would adjust the withdrawal amount annually by the observed rate of inflation. For instance, if the rate of inflation during the first year is 2.5%, the retiree would withdraw \$41,000 in the second year. The central question for retirees is: How much wealth must be accumulated in the investment portfolio to sustain this spending plan without depleting resources prematurely? Alternatively, given a certain amount of wealth in the investment portfolio, how much can be withdrawn initially while maintaining a constant real spending amount throughout retirement?

The safe withdrawal rate is the percentage of the initial wealth that a retiree can withdraw in the first year and that allows to adjust this amount by the rate of inflation in all subsequent years without depleting the entire investment portfolio prematurely. The safe withdrawal rate is estimated by testing different withdrawal rates on an initial investment portfolio and on different return and inflation paths and then counting how often the portfolio would have run out of money over the entire retirement period. For instance, analysis might conclude that a 4% initial withdrawal rate can be considered safe, as it historically failed only 5% of the time. In this scenario, if an initial expenditure of \$40,000 is to be maintained in real terms, the required wealth to be accumulated in the portfolio is calculated as follows:

$$\$40,000 \times \frac{1}{4\%} = \$1,000,000$$

An initial approach to the topic of retirement withdrawals may rely on average returns and inflation rates. For example, during the 20th century, the average return of the U.S. stock market was approximately 10% annually, with an average inflation rate of 3%, resulting in an average real return of 7%. Initially withdrawing the real return (7%) and then simply adjusting this amount by the rate of inflation may seem like a reasonable plan.

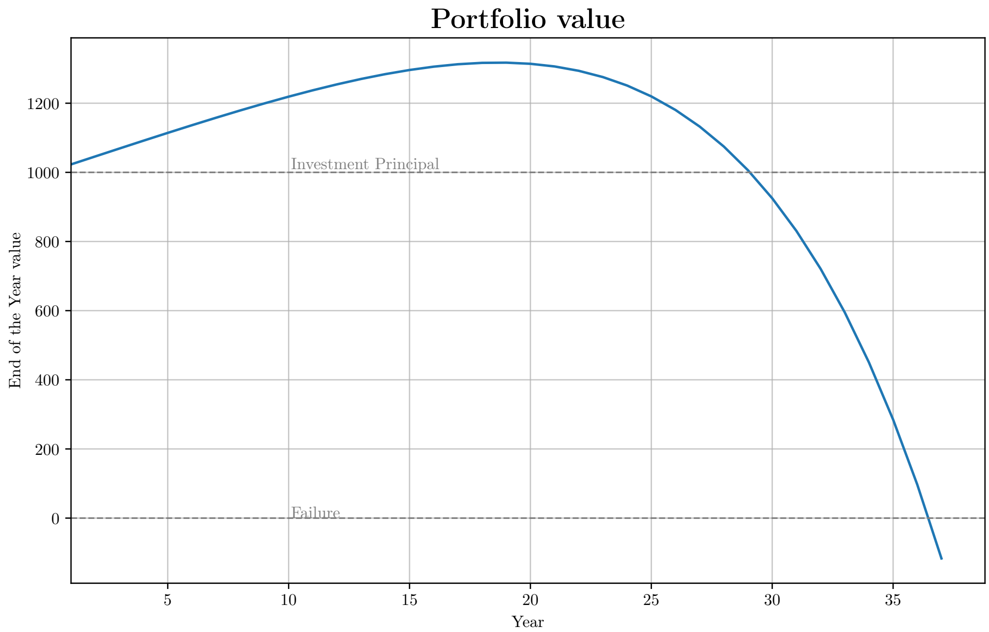

Figure 1.1: Evolution of the portfolio value when initially withdrawing the real return and adjusting for inflation in all the other years. The data are generated assuming a constant 10% nominal return and 3% inflation every year.

Figure 1.1 shows the evolution of the portfolio value when using this withdrawal strategy. This withdrawal strategy functions effectively in the initial years, with the portfolio value remaining relatively stable. In the first year, the portfolio returns 10% and despite withdrawing 7% of the initial value, the portfolio grows by 2.3%. However, as withdrawals continue to grow rapidly due to the compounding effect of inflation, they begin to constitute a significant portion of the portfolio, the growth of the wealth starts to slow down and the portfolio value after reaching a maximum in year 19, starts declining. In year 23, the retiree is withdrawing 10.37% of the portfolio while the investment returns are 'just' 10% each year. At this point, the portfolio value begins to decline rapidly, and by year 37, the portfolio is entirely depleted.

These numbers were generated under the assumption that the realized return and inflation rate are constant each year, at 10% and 3%, respectively. In practice, returns and inflation rates fluctuate, and retirees are financing a constant real cash outflow with a volatile asset. This volatility introduces the so-called *sequence of return risk*, which is the risk that a series of low or negative returns early in the retirement period can

Year	Initial Wealth	Withdrawal	Final Wealth	Withdrawal as % of Initial Wealth	Portfolio % Growth
1	1000.00	70.00	1023.00	7.00%	2.30%
2	1023.00	72.10	1045.99	7.05%	2.25%
3	1045.99	74.26	1068.90	7.10%	2.19%
4	1068.90	76.49	1091.65	7.16%	2.13%
			...		
18	1312.69	115.70	1316.68	8.81%	0.30%
19	1316.68	119.17	1317.27	9.05%	0.04%
20	1317.27	122.75	1313.97	9.32%	-0.25%
			...		
36	285.01	196.97	96.84	69.11%	-66.02%
37	96.84	202.88	-116.65	209.50%	-220.45%

significantly impact the long-term sustainability of the withdrawal plan.

Buy-and-hold investors are indifferent to the order in which they receive their returns and hence are not impacted by the sequence of returns. The commutative property of multiplication states that:

$$(1 + x)(1 + y) = (1 + y)(1 + x)$$

This property governs how returns are computed, and over multiple periods, the return of an investment is simply the geometric average. Thus:

$$(1 + 0.15)(1 - 0.08) = (1 - 0.08)(1 + 0.15)$$

However, this relationship does not hold when there are intermediary cash flows between each period, as in the case of a retiree who withdraws from their portfolio to finance their needs.

Let's consider three scenarios that have the same geometric average:

1. A steady 10% return at the end of year 1 and year 2;
2. A -15% return at the end of year 1 and a 42.35% return at the end of year 2;
3. A 42.35% return at the end of year 1 and a -15% return at the end of year 2.

	Buy and Hold			**Withdraw $10 each year**		
Year	Steady 10%	-15% / 42%	42% / -15%	Steady 10%	-15% / 42%	42% / -15%
0	100.00	100.00	100.00	100.00	100.00	100.00
1	110.00	85.00	142.35	99.00	76.50	128.12
2	121.00	121.00	121.00	97.90	94.66	100.40

For a buy-and-hold investor who starts with $100 in their portfolio, all three scenarios result in the same final value of $121 at the end of year 2. However, for a potential retiree who starts with $100 and withdraws $10 at the end of each period, the final results differ. The second scenario, with a negative return early on, leads to a much lower final value compared to the third scenario, where the negative return occurs at the end of year 2. The rationale is that, in scenario 2, the retiree withdraws from a portfolio significantly diminished by the initial downturn, and the subsequent 42.35% return is therefore calculated on a reduced base. In contrast, experiencing a high return early in the period allows the retiree to outperform the steady +10%/+10% scenario.

This exemplifies why, when examining withdrawal rates, reliance on simple averages should be avoided and historical data or simulated return paths should be used, as market volatility can significantly impact the probability of success for a portfolio.

The use of average values to estimate a retiree's annual spending budget has long been employed by financial planners. However, Bengen (1994) was the first employ a historical approach to determine how much retirees can safely spend in retirement. Bengen used U.S. stock and bond returns from the 1920s up to the publication date to estimate how long a retirement portfolio would have lasted, assuming an initial withdrawal rate and successive inflation-adjusted withdrawals. Assuming a 50% stock and 50% bond allocation, Bengen concluded that an absolutely safe withdrawal rate—one that would have allowed a portfolio to last at least 50 years for all starting dates—is 3%. Retirees with shorter horizons, such as 30 years, could even afford a 4% withdrawal rate. Higher withdrawal rates, for instance 6%, are too risky as they would have failed to last longer than 30 years. Additionally, retirees with bequest motives should consider lowering their withdrawal rate. In the various scenarios analyzed, heavy stock allocations, between 50% and 75%, consistently yielded better results, as they allowed for faster recovery after market downturns and generally enabled the accumulation of larger wealth for the estate.

Bengen also provided guidelines on how retirees should behave during the decumulation phase. Retirees should never be tempted to increase their withdrawals at a rate greater than inflation, as this would deplete the value of the portfolio rather quickly, even during periods of very positive returns. Moreover, this approach could inflate the retiree's lifestyle, making it difficult to return to a more frugal lifestyle if necessary.

Cooley, Hubbard, and Walz (1998) published one of the most influential papers in the field of retirement planning. Their study, also known as the *'Trinity Study'*, employed a historical approach to determine a safe withdrawal rate. Their methodology involved testing different combinations of withdrawal rates, stock allocations, and retirement durations for all available starting dates, then counting the number of occurrences in which the final portfolio value was greater than zero. The authors observed that the safe withdrawal rate varied significantly across different periods. For example, over the timeframe from 1926 to 1995, without adjusting for inflation (i.e., the retiree withdraws the same nominal amount every year), they concluded that a 4% initial withdrawal rate was very reasonable for a 30-years retirement. However, they noted that the financial market conditions in the late 1990s were markedly different from those during the Wall Street Crash of 1929 and before World War II. Therefore, they suggested focusing on the subperiod starting from 1946. In this shorter timeframe, and without adjusting for inflation, they found that a more aggressive 7% withdrawal rate would have been reasonable with high stock allocations.

The authors explained their decision not to adjust for inflation by arguing that the CPI inflation rate overstates the actual inflation of consumption by 1-1.5%, thus the loss in purchasing power was not as significant. However, they also estimated that if retirees wanted to adjust their withdrawals by the rate of inflation, a 4% to 5% withdrawal rate would have been safe over the 1926-1995 period.

The main insight from the study is that the probability of failure drastically increases with longer retirement periods and higher withdrawal rates. Additionally, the authors suggested that including bonds in the portfolio increases success rates for mid to low withdrawal rates, but for higher withdrawal rates, a large allocation to stocks is crucial for increasing the chances of success due to their upside potential. They also observed that, even though there were cases where the portfolio was exhausted, the terminal values were often very large, with mean terminal values greater than median values, indicating a positive skew in the distribution.

Success rates estimated from a small number of scenarios are prone to estimation error. This is particularly true for estimates that rely on overlapping historical scenarios, as in the case of the two studies mentioned. This issue has led some researchers to develop market models—stochastic models of asset returns and inflation rate processes.

Pye (2000) proposed a simulation approach to estimate withdrawal rates, assuming returns are lognormally and independently distributed with a mean of 8% and a standard deviation of 18%, to match the historical moments of U.S. stock returns. The methodology involved setting an initial withdrawal rate and continuing to withdraw the same amount in real terms until that amount was no longer sustainable, at which point the annual withdrawal would be reduced.

Pye found that, despite an expected return of 8%, initially withdrawing 8% leads to serious sustainability risks, even in the first years, with a 35% probability of having to reduce the withdrawal rate by 35% or more during the first 10 years. Reducing the initial withdrawal rate by 25% increased the probability of sustaining it over the entire period without further cuts from 15% to 45%. Further reducing the initial withdrawal rate by 50% raised the probability of sustaining it over the entire period to 80%.

The studies discussed so far have attempted to develop a spending plan for retirees by using the past century of U.S. stock market returns and macroeconomic data to derive a safe withdrawal rate. The implicit assumption in these studies is that if a withdrawal rate guaranteed high success rates in the past, it will likely work in the future as well. However, it is crucial to assess the validity of this assumption and the appropriateness of using the historical experience of the U.S. to model withdrawal rates for future retirees.

Mehra and Prescott (1985) identified that U.S. stocks outperformed bonds by approximately 6% per year over the previous century. The authors could not find a satisfactory explanation for this overperformance, as individuals would need to exhibit extraordinarily high levels of risk aversion to justify this discrepancy—a phenomenon that became known as the 'Equity Risk Premium Puzzle'. Academics have since attempted to explain this puzzle by exploring alternative preference structures or by identifying limitations in the available data. A more recent stream of studies has suggested an alternative explanation: ex-post returns may be biased by the survival of the series, meaning that risk aversion cannot be accurately inferred from empirical analyses of historical data that are conditional on survival.

Jorion and Goetzmann (1999) proposed expanding the data sample to include information from markets that experienced temporary or permanent interruptions. The survival explanation could at least partially account for the Equity Risk Premium Puzzle, as the U.S. experienced a real appreciation of 4.3% per year during the considered period, while the median of all other countries lagged behind at just 0.8%. The conclusion is that relying solely on U.S. historical data to estimate long-term market growth is unjustified, as the U.S. experience does not accurately reflect that of investors in many other developed countries.

Pfau (2017), estimated the success rates of the 4% rule across a wide sample of countries. The author collected data from 1900 to 2015 for 20 developed countries and discovered that only the U.S. and Canada would have sustained the 4% withdrawal rule for a 30-year retirement period. Other countries exhibited failure rates ranging from 8% to 62%, and the global stock market would have historically supported a 3.5% withdrawal rate. However, this study suffers from a significant flaw: the countries included in the dataset were those considered developed at the time of publication. This methodology introduces survivorship bias, as it only considers successful countries that survived and thrived over time, likely inflating the success rates.

Anarkulova et al. 2022 addresses these limitations by using a new dataset specifically designed to overcome the issues of small sample size, survivor bias, and easy data bias. This approach allows investors to form better ex-ante expectations for asset returns using a broader sample of developed markets rather than relying on the small and biased U.S. sample. The dataset spans from 1890 to 2019, and the inclusion of countries in the sample is based on an approach that mitigates survivor bias by relying on ex-ante measures of economic development to determine the initial sample inclusion date for each country. Using this dataset, the authors generated various simulation paths through a block bootstrap methodology, preserving the cross-sectional and time-series properties of long-horizon asset returns. The authors estimated that a 4% withdrawal rate had a 17.4% failure probability for a retired couple and the safe withdrawal rate (with a success rate greater than 95%) was 2.26%, 200 basis points lower than that estimated by considering only the U.S. sample.

1.2 Breaking down the assumptions of the studies

The studies discussed in 1.1 are academic attempts to estimate safe withdrawal rates. However, these studies overlook a range of factors that are crucial for retirees and could significantly impact their chances of financial success. Moreover, these studies were primarily designed for traditional retirement planning and not specifically tailored to the needs of early retirees.

In this section, the study dissects the assumptions underlying these studies and evaluates how they influence success rates, specifically considering a real-world scenario: the case of a 45-year-old Italian who is about to retire.

First, the validity of extrapolating success rates over a 30-year period to plan for much longer retirement horizons is examined, as is necessary for early retirees. The potential impact of an extended retirement duration on success rates will be explored.

The effects of taxes and transaction costs will be considered next. Most studies do not account for these factors; however, understanding the impact of taxes on wealth is crucial for retirees planning their financial future, as will be discussed in the following sections.

Traditional studies define success as any scenario where the final portfolio value remains above zero, implicitly assuming that retirees are not concerned with leaving a bequest and are content with depleting their capital. Therefore, the implications of capital preservation versus capital depletion on success rates will be assessed.

Additionally, the assumption that retirees rely solely on withdrawals from their investment portfolio, without other sources of income, will be explored. In practice, many countries, such as Italy, provide retirees with social security benefits upon reaching a certain age. The impact of including social security income on the safe withdrawal rate will be examined.

This section will analyze the influence of these assumptions on success rates, based on historical market and macroeconomic conditions, under the premise that a withdrawal rate that has been successful in the past will continue to be effective in the future. For our analysis, the dataset provided by Robert Shiller[1] will be used. This dataset spans a significant historical period, with monthly financial data from January 1871 to July 2024. Stock returns are represented by the S&P 500 index, and the dataset also includes dividends paid by the constituent firms, allowing to compute the total return of the index, including both capital gains and dividends. Bond returns are captured through the GS10, which reflects the yield on U.S. Treasury securities with

[1] Data available at Shiller Data.

10-year constant maturities. In addition to stock and bond returns, the dataset includes the Consumer Price Index (CPI).

To estimate the safe withdrawal rate, an overlapping windows algorithm is employed to test various combinations of withdrawal rates and stock-bond allocations across all possible starting dates. The process begins by setting the initial withdrawal rate for the first month and deducting this amount from the initial portfolio value. Each month, the portfolio value is adjusted by that month's nominal rate of return, establishing the initial value for the following month. The subsequent month's withdrawal is adjusted for inflation based on the previous month's observed rate and deducted from the portfolio's value. This procedure is repeated monthly until the end of the specified retirement horizon. For instance, using market data starting from January 1871 for a 50-year period, the first iteration spans January 1871 to January 1921, the second iteration covers February 1871 to February 1921, and so on. The success rate is determined by calculating the percentage of iterations in which the final portfolio value remains above zero.

1.2.1 The implications of a longer retirement duration

As mentioned earlier, many studies on withdrawal rates, including the Trinity Study, focus primarily on how much a retiree should save and subsequently spend during retirement. However, these studies were not designed for early retirees who plan for much longer retirement horizons than the typical 30-year period these studies often consider.

Section 1.1 highlighted that Longevity Risk, the possibility that an individual lives longer than initially anticipated requiring more resources to finance their needs during retirement, is one of the main sources of risk for a retiree alongside Market Risk. Today, extensive data on life expectancy are available. For instance, using the 2020 Italian Mortality Table [2], it is possible to estimate the life expectancy of a hypothetical 45-year-old retiree. The median life expectancy for a 45-year-old Italian in 2020 was 83 years for males and 87 years for females. However, when planning for retirement, relying on the mean or median life expectancy can be risky. Retirees tend to be highly risk-

[2] Data available at Human Mortality Database.

averse and prefer to plan for a longer life expectancy to avoid the risk of outliving their savings. Optimizing for the (almost) worst-case scenario can reduce this risk, but it may also result in over-saving and under-utilizing financial resources.

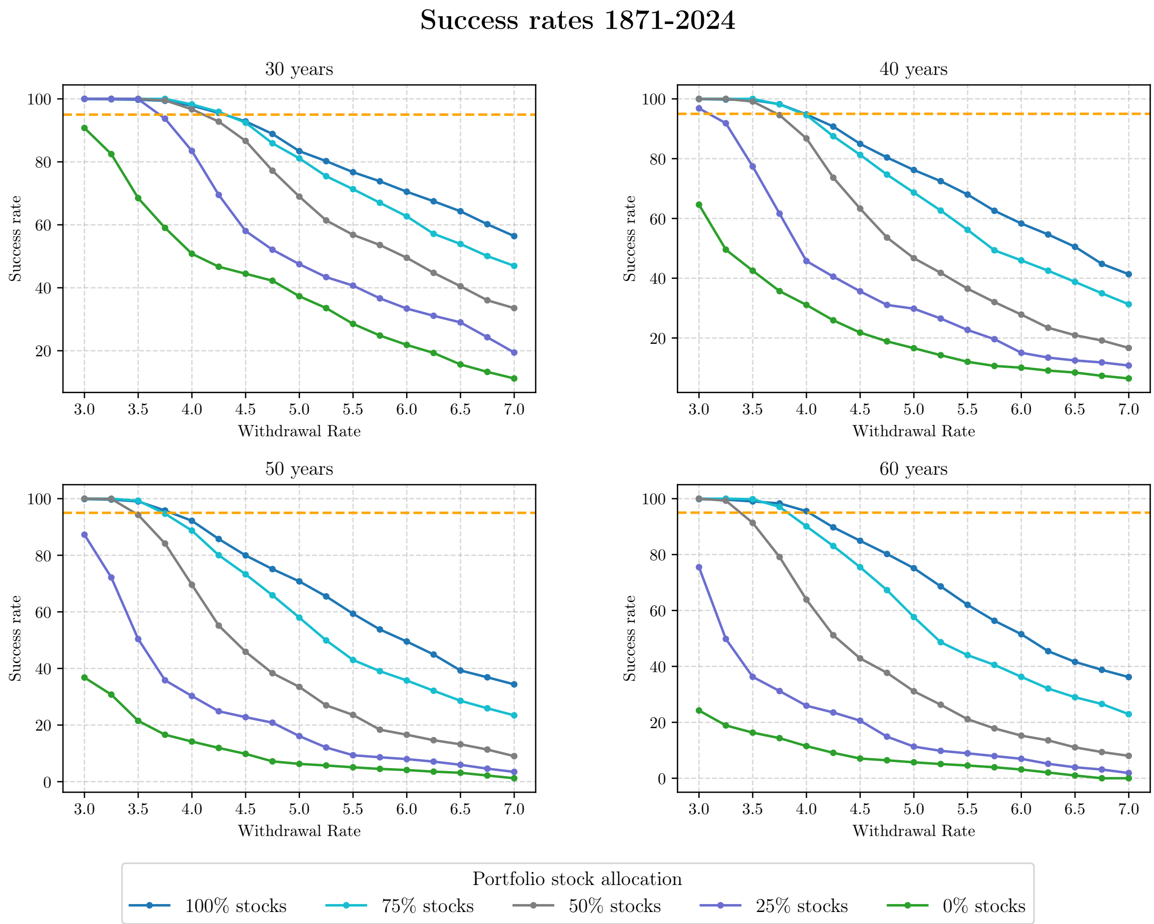

Figure 1.2: Success rates based on US stock and bond returns over the period 1871-2024.

Figure 1.2 allows to compare the success rates for different retirement horizons: 30 years (for comparison with other studies), 40 years, 50 years, and 60 years. Historically, over a 30-year retirement period, a 4% withdrawal rate has provided high success rates when at least 50% of the portfolio is allocated to stocks. However, for longer retirement horizons, failure rates increase. For a 40-year retirement period, a 4% withdrawal rate yields a 95% success rate only when at least 75% of the portfolio is in stocks; reducing the stock allocation to 50% drops the success rate below the safety threshold of 95% success rate. For a retiree planning to rely on their investments for 50 years, the 4% withdrawal rate provided at best a 92.19% success rate with a 100% stock allocation. For a 60-year retirement period, the success rate just meets the 96% threshold with the same asset allocation. Interestingly, longer retirement periods generally result in lower success rates for a fixed withdrawal rate and asset allocation. However, with a 100% stock allocation and a 4% withdrawal rate, the 60-year retirement horizon surprisingly shows a higher success rate than the 40- and 50-year horizons. This might be due to the longer investment period, which offers more time for the portfolio to recover from market downturns.

From this analysis, it can be concluded that extrapolating the success chances of withdrawal rates from a 30-year horizon and applying them to longer retirement periods is not advisable. Historically, a withdrawal rate between 3.5% and 3.75% can be considered safe for all retirement horizons. These values are the ones that will be used as a baseline in the next section to evaluate the impact of the different assumptions. Furthermore, it is possible to observe that bonds are riskier than stocks for retirement spending: allocating less than 50% of the portfolio to stocks has historically corresponded to success rates below our safety threshold. Allocating between 75% and 100% to stocks has historically yielded the best outcomes for retirees.

The assumption that a portfolio that survives for 30 years without depleting the initial capital will likely survive for another 30 years is not entirely accurate. While it is true that the risk of running out of money after 30 years is low for the safe withdrawal rate, and the median final value is often much higher than the initial value (even after adjusting for inflation), retirement planning should focus on the tails of the distribution rather than the median. Indeed, after 30 years, a significant portion of retirees may not be classified as failures, but they could have a portfolio value below the initial amount adjusted for inflation. Therefore, these retirees that have a wealth below their initial portfolio value after 30 years, may not have sufficient capital to support their withdrawals for another 30 years.

1.2.2 The impact of capital preservation

Most studies in this field consider a retirement plan successful if it avoids running out of money in the final month or year of the analyzed time window. Consequently, these studies implicitly assume that retirees are content with completely depleting their entire wealth by the end of the retirement period. However, some retirees may be interested in preserving their capital, perhaps because they wish to leave a portion of their accumulated wealth as a bequest to their estate. In this context, capital preservation means that retirees aim to maintain a minimum asset level—expressed as a percentage of their initial wealth—at the end of their retirement horizon. Specifically, under full capital preservation, retirees strive to keep the real, inflation-adjusted principal intact.

To assess the impact of capital preservation on safe withdrawal rates, success rates are calculated by targeting a final portfolio value that is 50% and 100% of the initial portfolio value in real terms. In this scenario, success is defined as ending with at least 50% or 100% of the initial value in real terms.

To simplify the interpretation, Figure 1.3 presents the results for an asset allocation of 75% stocks and 25% bonds.

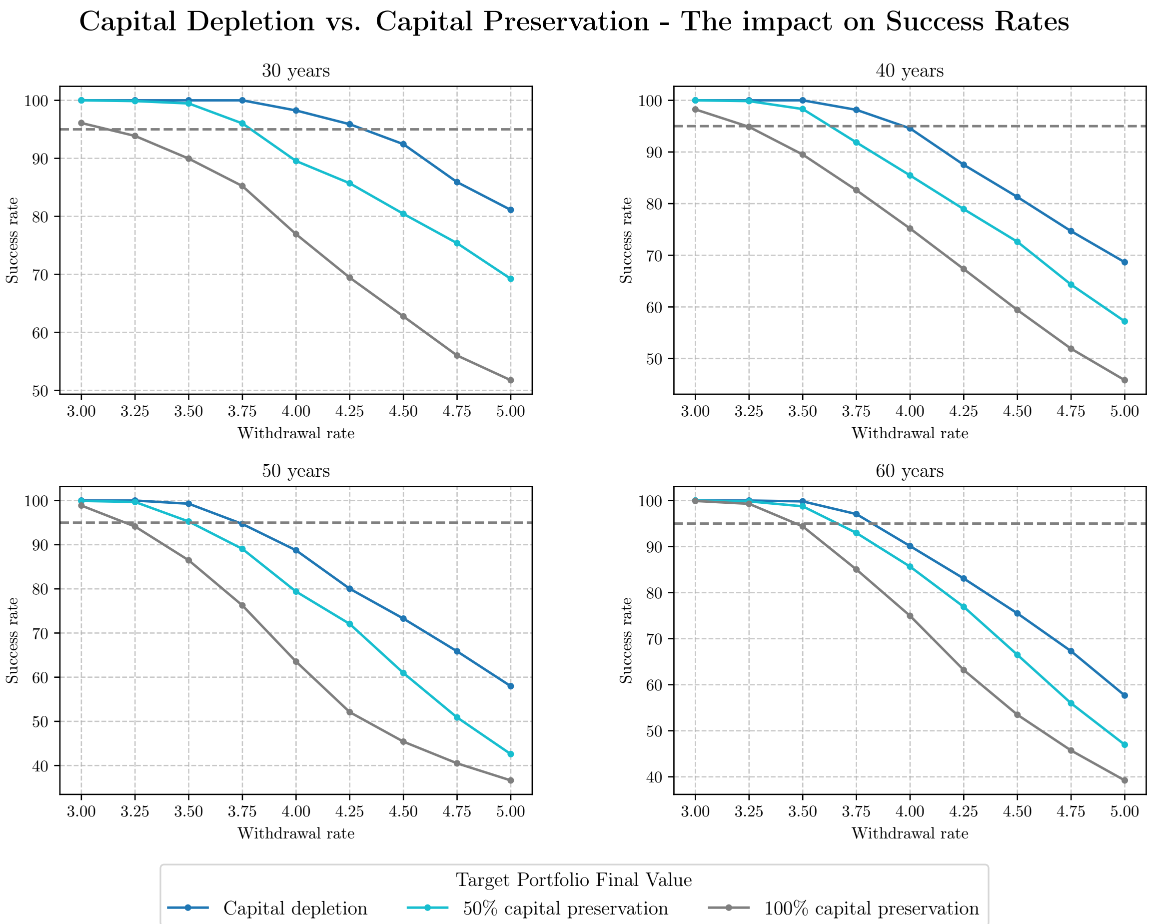

Figure 1.3: The impact of capital preservation on success rates. Results based on US stock and bond returns over the period 1871-2024.

As expected, targeting capital preservation negatively impacts success rates. In the previous section it was demonstrated that, under capital depletion, a 3.75% withdrawal rate can be considered safe even for a 60-year retirement period. However, when aiming for 50% and 100% capital preservation over 30 years, the success rate drops significantly—from 100% to 96.02% and 85.22%, respectively.

For longer retirement durations, the decrease in success rates remains relatively consistent. For example, over a 60-year period, with a 3.75% withdrawal rate, success rates decline from 97.06% to 92.96% and 85.03% when targeting 50% and 100% capital preservation, respectively. This is encouraging news for retirees: if they aim to preserve some of their initial capital, longer retirement horizons do not drastically reduce their chances of success. The reason is that, over longer periods, the high growth potential of stocks can allow the portfolio to expand and eventually recover from market downturns.

The key takeaway for retirees is that if they want to achieve capital preservation, they will need to lower their withdrawal rates. Specifically, when targeting 50% capital preservation in real terms, a safe withdrawal rate appears to be 3.50%, and when targeting full capital preservation, a rate of 3.25% is advisable.

1.2.3 The impact of taxes and transaction costs

Most studies in the field of withdrawal rates, don't adjust their results for taxes or transaction costs. The reason for omitting these factors is that they can vary widely depending on a variety of characteristics specific to each retiree. For example, access to tax-deferred investment accounts varies among individuals, and taxation regimes differ significantly between countries. However, this chapter will try to estimate the impact of taxes and transaction costs on success rates, focusing on a practical case study of an Italian soon-to-be retiree.

The analysis assumes that the retiree has optimized their investment strategy to minimize costs. Recently, there has been a surge in low-cost investment platforms that allow investors to buy and sell securities with minimal fees. Therefore, is assumed that the retiree in consideration will be able to invest through one of these platforms, which charge such trivial fees that they can be neglected in our calculations.

The past decade has seen growing interest from investors in index funds and low-cost ETFs. Exchange-Traded Funds (ETFs) are pooled investment securities that, unlike mutual funds, can be bought and sold on an exchange, similar to stocks. Specifically, passive ETFs are designed to track underlying assets such as an index. In this case, the ETF manager replicates the index's performance by trading its components, eliminating the need for active security selection as in a mutual fund. This approach significantly reduces the intervention of the fund manager, allowing ETFs to charge much lower costs.

Until now, it was assumed that the fictional retiree was investing in a combination of stocks and bonds. Therefore, it is possible to imagine that the retiree is investing in a multi-asset ETF, such as the Vanguard LifeStrategy ETFs [3], which provide exposure to both bonds and stocks with a single investment product and a Total Expense Ratio (TER) of 0.25% per year. Thus, in the calculations, to include the cost of the investment product the TER is subtracted from the performance of stocks and bonds. Specifically, the return of the portfolio will be computed as:

$$R_{\text{Portfolio}} = R_{\text{Stock}} \times W_{\text{Stock}} + R_{\text{Bond}} \times (1 - W_{\text{Stock}}) - \frac{\text{TER}}{12}$$

where:

- $R_{\text{Portfolio}}$ is the net return of the portfolio,
- R_{Stock} is the return of the stock, in our case, the S&P 500 Total Return,
- R_{Bond} is the return of the bond, in our case, GS10, which is the yield on U.S. Treasury securities with 10-year constant maturities,
- W_{Stock} is the weight of stocks in the portfolio,
- TER is the Total Expense Ratio, in our case, 0.25% per year.

Note that the TER is divided by 12 as monthly data are considered and hence it is assumed that this cost is charged monthly.

In the Italian tax system, one of the primary taxes on investment portfolios is the 'Imposta di Bollo', a fixed annual charge of 0.2% of the portfolio's value. This tax is typically levied once per year, but for simplicity in the calculations, it is assumed that this tax is paid monthly, similar to the Total Expense Ratio (TER). Thus, the net return of our portfolio can be computed as:

[3] Product details available at Vanguard LifeStrategy ETF

$$R_{\text{Portfolio}} = R_{\text{Stock}} \times W_{\text{Stock}} + R_{\text{Bond}} \times (1 - W_{\text{Stock}}) - \frac{\text{TER}}{12} - \frac{0.20\%}{12}$$

In the Italian fiscal system, taxation on investment products is governed by Decreto Legge 66/2014 [4]. Under the current tax regime, capital gains (or losses) from selling an ETFs are classified as 'redditi da capitale' and are subject to a 26% tax. When an ETF is sold, a 26% tax is applied to the capital gain, which is the difference between the selling price and the weighted purchase cost (WPC)—the weighted average of all the prices at which the ETF was previously purchased. However, if the ETF is sold at a loss, meaning the selling price is lower than the weighted purchase price, no tax is applied. ETFs that include government bonds are an exception, with capital gains taxed at a lower rate of 12.5%. Capital gains from multiasset ETFs have a taxation of 26% on the stock part and of the 12.5% on the bonds. For instance the capital gain tax on 80% stocks and 20% bonds Multiasset ETF is given by

$$\text{CG Tax} = 26\% \times W_{\text{Stock}} + 12.5\% \times (1 - W_{\text{Stock}}) = 26\% \times 80\% + 12.5\% \times 20\% = 23.3\%$$

Using an ETF instead of individual stocks and bonds simplifies the tax calculation by eliminating the need to account for the so-called 'zainetto fiscale', which allows investors to offset future tax liabilities when selling stocks or bonds at a loss.

In the calculations, taxes require tracking the current price of the security and the weighted purchase price. The algorithm computes taxes as follows:

- If $\text{Price}_t > \text{WPC} \times \text{Price}_0$, where Price_t is the price at time t of the hypothetical ETF, WPC is the weighted purchase price, and Price_0 is the ETF's price at the retirement date. Retirees build their investment portfolio over a long time, by frequently purchasing new shares. Hence, it is possible to say that the WPC in most cases would differ from Price_0. For instance, WPC may be 80% of Price_0 and in this case, the capital gain on which the tax is paid will be higher than in the case in which $\text{WPC} = \text{Price}_0$. The tax payable on each share sold at time t

[4] Details available at Decreto Legge 66/2014

is:

$$\text{Tax}_t = (0.26 \times W_{\text{Stock}} + 0.125 \times (1 - W_{\text{Stock}})) \times (\text{Price}_t - \text{WPC} \times \text{Price}_0)$$

- If $\text{Price}_t < \text{WPC} \times \text{Price}_0$, no tax is payable at time t for each share sold.

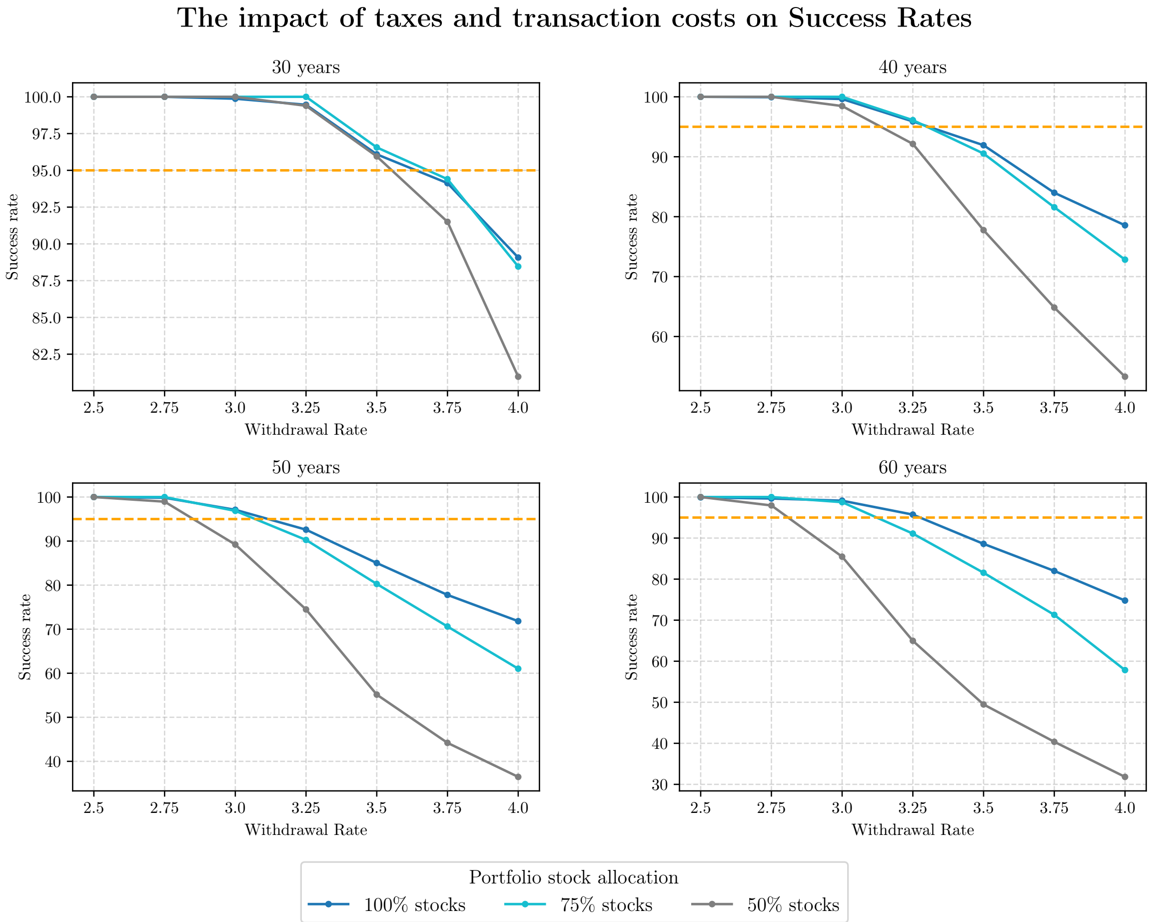

Figure 1.4: The impact of taxes and transaction costs on success rates. Results based on US stock and bond returns over the period 1871-2024.

Figure 1.4 illustrates the success rates relative to the withdrawal rate when accounting for taxes and transaction costs. The drop in success rates compared to the base case is significant, particularly for longer horizons. This greater impact on success rates over extended periods can be attributed to the substantial capital gains accumulated over 60 years. As retirees withdraw larger amounts to cover taxes on these capital gains, their wealth depletes more rapidly.

While a 3.75% withdrawal rate seemed safe in our baseline scenario, accounting for taxes and transaction costs reveals that the probability of failure is now far from acceptable. For high stock allocations (75% or 100%), the safe withdrawal rate should be reduced by 50 basis points, to 3%, to achieve a reasonable success rate. For more conservative allocations, such as 50% stocks and 50% bonds, the success rates fall below the 95% threshold even with a 3% withdrawal rate. Therefore, in these cases, the safe withdrawal rate must be further reduced to 2.75%, particularly for longer retirement horizons.

1.2.4 The impact of additional cashflows

Studies in the field of safe withdrawal rates often assume that retirees will not have any additional sources of income during their retirement period. However, it is plausible to consider that an early retiree might receive positive cash flows, such as salaries from casual jobs, inheritances from relatives, or income from pensions. While modelling the impact of salaries or inheritances is not feasible due to their dependence on individual choices and circumstances, it is possible to estimate the impact of pensions, as they are determined by predefined rules.

To account for the impact of social security pensions, the case study of a fictitious 45-year-old Italian soon-to-be retiree is still considered. In the Italian pension system, an individual has the right to receive the ‘pensione di anzianità’ upon reaching the age of 67, provided they have worked and contributed regularly for at least 20 years. If this condition is met, once an individual turns 67, they are entitled to receive a monthly allowance of at least 1.5 times the ‘Assegno Sociale’, which in 2024 amounts to €6,947.33 per year [5]. Therefore, the monthly allowance of the ‘pensione di anzianità’ is €868.26 per month once they reach 67.

The calculation assumes that the requirements to receive social security pensions will not change in the future. Furthermore, it is assumed that the cost of living and the income from pension will grow at the rate of inflation, and therefore, the pension will cover a fixed proportion of monthly spending. Hence, in the calculation it is assumed that once the retiree reaches the age of eligibility for the social security pension, the withdrawal amount will be reduced by the proportion of monthly expenses covered by the pension. For instance, it is possible to consider a retiree who withdraws $40,000

[5] For further details: ‘L’assegno Sociale INPS 2024’

annually in real terms until age 67. If the pension covers, say, 40% of their financial needs, the retiree will then withdraw only $\$40,000 \times (1 - 0.40) = \$24,000$ from their investment portfolio moving forward. Assuming that the pension is indexed to the cost of living, it can be further assumed that the pension will continue to cover a constant proportion of the retiree's living expenses.

According to the most recent data from ISTAT, the median cost of living for a single person in Italy is €1,773.12 per month. Consequently, social security will help cover approximately 48.96% of monthly needs. Therefore, the calculation assumes that when a retiree turns 67, only the remaining portion not covered by social security will be withdrawn from the portfolio, amounting to 51.04% of the monthly needs.

The impact of pension on success rates

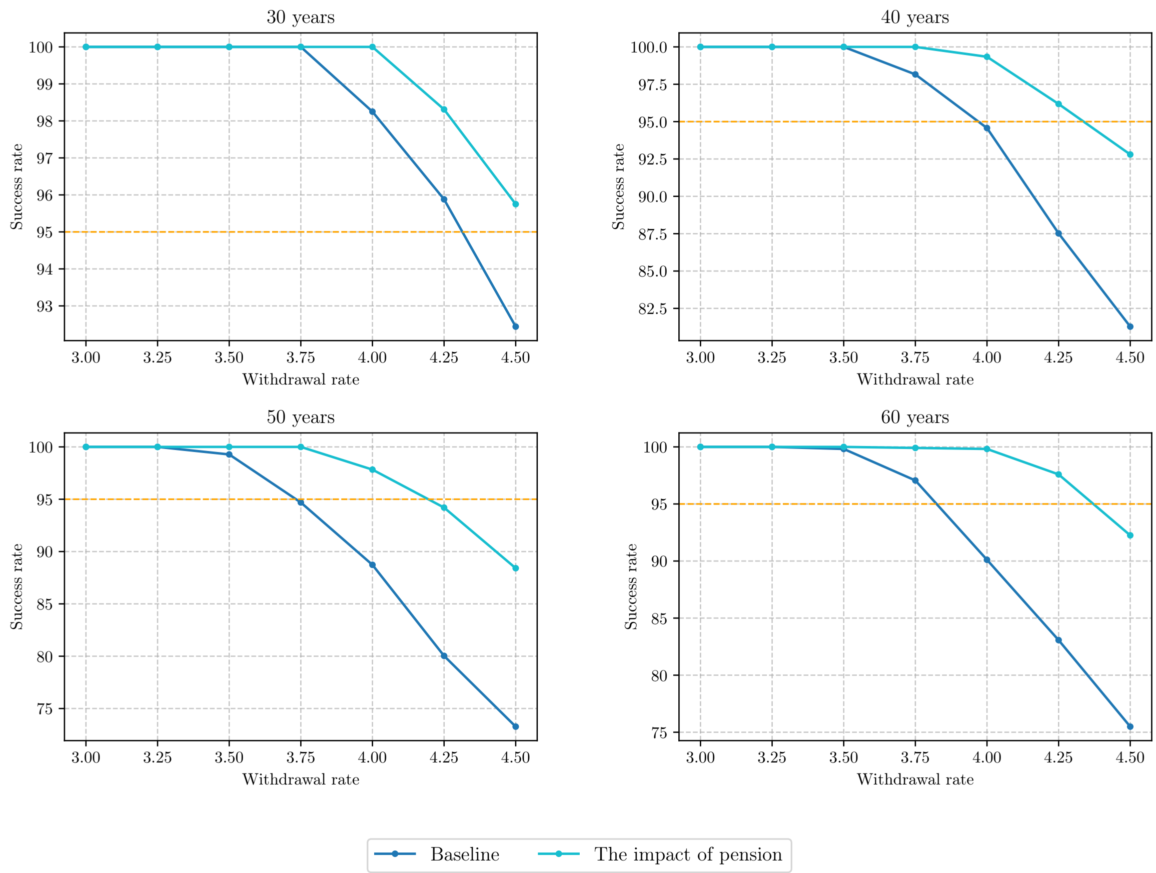

Figure 1.5: The impact of income from pension on success rates. Results based on US stock and bond returns over the period 1871-2024.

Figure 1.5 illustrates the impact of accounting for social security pensions on success rates. It is possible to observe that longer retirement periods experience the greatest improvement in success rates. Specifically, assuming that the fictional retiree is 45 years old and will only begin receiving a pension at age 67, a 30-year retirement would provide pension benefits for just 8 years, whereas a 60-year retirement would provide income from the pension for 38 years. Consequently, in the case of a longer retirement duration, the retiree benefits from a reduction in the monthly withdrawal amount for a more extended period, thereby significantly improving the success rate.

Indeed, when factoring in the income from the pension, the 4% rule is safe for all retirement durations. For a 30-year retirement, it would even be possible to increase the withdrawal rate above 4.5%, while for a 40-year retirement, a 4.25% withdrawal rate is still feasible.

1.3 Is the '4% rule' really desirable?

In the previous section, the impact of each of these assumptions on success rates was analyzed. The next logical step is to put everything together and estimate what could be a safe withdrawal rate.

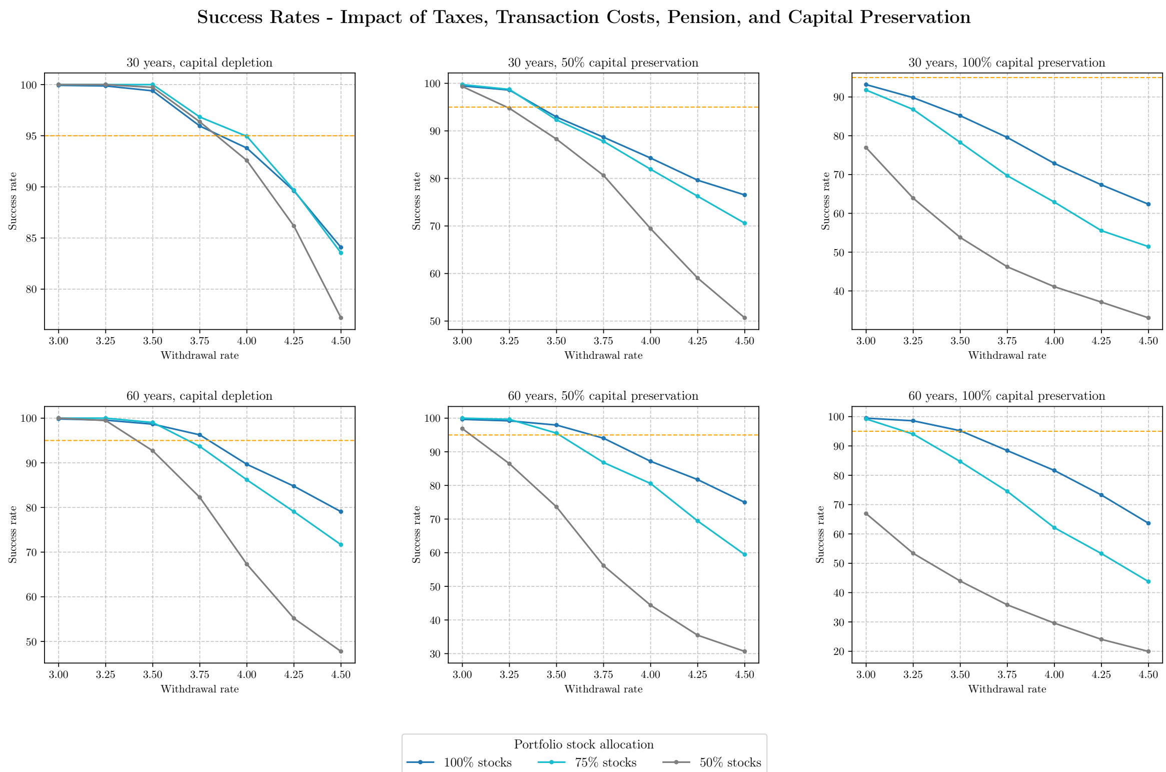

Figure 1.6: Distribution of Terminal Values of the portfolio as multiplier of the inflation adjusted principal. Results based on US stock and bond returns over the period 1871-2024.

Figure 1.6 shows the success rates against different withdrawal rates for 30 and 60 years, targeting different final values, and considering the impact of taxes, transaction costs, and income from pensions based on the criteria described in the previous sections. As discussed in the dedicated section, capital preservation impacts shorter retirement durations more heavily than longer ones. Indeed, over 30 years, 50% capital preservation requires retirees to sacrifice 75 basis points on their initial withdrawal rate, from 4% to 3.25%. If they aim for full capital preservation, they will need to accept a withdrawal rate below 3%. However, over 60 years of retirement, targeting 50% capital preservation requires retirees to give up just 25 basis points and 50 if they aim for full capital preservation. Furthermore, the chart highlights the importance of a high stock allocation: investing less than 75% of the portfolio in stocks causes sharp drops in success rates, especially for longer retirement periods and higher capital preservation targets.

The conclusion of this analysis is that academic studies have overlooked various factors crucial for early retirees, and the 4% rule seems far from safe. Early retirees should consider withdrawal rates of 3.50% over 60 years if they plan to deplete their capital or closer to 3.25% if they aim for capital preservation.

It is also insightful to analyze withdrawal rates not only through the lens of success rates but by inspecting the distribution of terminal wealth values.

Figure 1.7 depicts the distribution of the final portfolio value as a multiple of the real principal, i.e., the principal adjusted for observed inflation each year, assuming an asset allocation of 75% stocks and 25% bonds. Specifically, the range between the 5th and 95th percentiles and the median value is represented. The observation that the 90% confidence interval (represented by the sky-blue area) falls below 0 indicates that, in more than 5% of the cases, the specific withdrawal rate failed. However, for all analyzed retirement durations and withdrawal rates, the median terminal value has consistently remained above the initial inflation-adjusted wealth. This insight underscores the importance of focusing on the left tail of the distribution rather than on the median values.

Retirees understand that even though there is a chance of failure in their withdrawal strategy, there is substantial upside potential for their wealth, which they could eventually pass on to their estate. For example, in the case of a 60-year retirement period

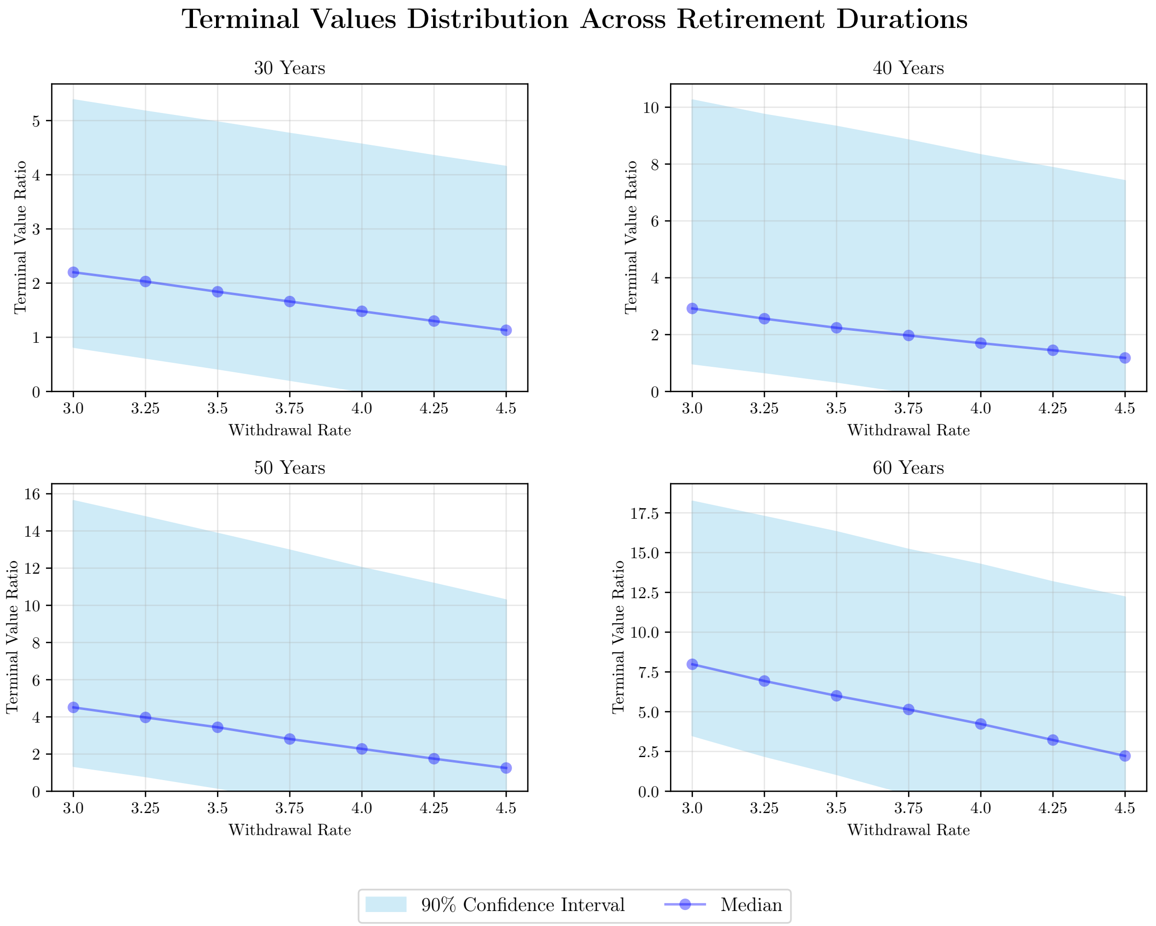

Figure 1.7: Terminal values when accounting for taxes, transaction costs, income from pension and different capital preservation targets. Results based on US stock and bond returns over the period 1871-2024.

with a withdrawal rate of 3.5%, deemed safe, the median terminal wealth is exactly 6 times the initial inflation-adjusted wealth, and in the 95th percentile, it could be 16.30 times.

On the one hand, there is always the rare but extremely impactful risk of running out of resources, while on the other hand, there are very good chances to end up with a very large investment portfolio. This very wide range of possible outcomes does not seem to be desirable for retirees.

Scott, Sharpe, and Watson (2008) argued that supporting a constant spending plan using a volatile investment policy is fundamentally flawed. A retiree using a 4% rule faces spending shortfalls when risky investments underperform and may accumulate wasted surpluses when they outperform, and in any case, could likely purchase exactly the same spending distributions more cheaply. In essence, retirees implementing a decumulation plan based on withdrawal rates are overpaying for the potential investment gains that they do not need to meet their retirement income goals, resulting in an inferior spending plan.

Figure 1.7 illustrates this issue: retirees choose a withdrawal rate that minimizes the failure rate in the worst-case scenario, leading to the accumulation of large wealth that will likely remain untouched. This inefficiency is one of the most significant drawbacks of strategies based on an initial withdrawal rate and inflation-adjusted withdrawals in all subsequent years, raising the need to structure a more efficient decumulation strategy.

Chapter 2

Natural Tontine

2.1 Why tontines?

Tontines, introduced in the 17th century, were unique investment structures where a group of investors pooled their funds, paid a lump sum, and received annual payments until their death. Unlike annuities, when a member died, their share was redistributed among the surviving members. For example, consider 1,000 investors each contributing $1,000 to purchase a $1 million bond with a 3% annual coupon. The bond would pay $30,000 yearly, to be shared among the participants. If all members were alive at the first payment, each would receive $30. However, after 10 years, with supposedly only 750 survivors, each would receive $40. After 20 years, with only 100 survivors, each would receive $300. This process continued until at least one person remained in the fund, making investors speculate on their life expectancy for potentially large payments if they outlived other members.

The concept of the tontine was first proposed by Lorenzo De Tonti, a Neapolitan banker, that in the 1650s was serving as a financial consultant to the French crown (McKeever 2010). At that time, France was engaged in the Thirty Years' War, and King Louis XIV sought new financing methods to support the military efforts. In 1653, De Tonti suggested the tontine structure to raise funds, highlighting its potential benefits for both investors and the kingdom. However, the idea was initially rejected due to concerns about the lack of reliable life expectancy estimates.

The first successful tontine was established in Kampen, in the Netherlands, in 1670, and other Dutch cities soon followed. France issued its first tontine shortly after, and in 1693, the English government organized its first state tontine under King William III (Milevsky and Salisbury 2015). This tontine aimed to raise £1 million to finance the war against France, allowing British citizens to subscribe with a £100 premium. The tontine guaranteed a 10% dividend for the first seven years and 7% thereafter, as long as the nominee (who could differ from the investor) was alive.

Over time, tontines became a popular tool for project financing. By the 1850s, U.S. insurance companies also began offering these products, which soon became common in Europe. However, a series of scandals, such as manipulation of tontine member registers, excessive fees, and conflicts of interest regarding investments, led the New York State Insurance Commission to ban tontines in 1910. Many other countries soon followed suit.

Today, tontines in Europe are regulated by Directive 2002/83/EC, which allows new pension products to be structured around the tontine principle. A growing number of insurance companies, pension funds like QSuper in Australia [1] and Le Conservateur in France [2], and fintech firms such as Tontine Trust [3], have begun offering products based on tontines. Recent academic literature has proposed new tontine designs, arguing that they could be effective tools for financing retirement.

Modern tontines are often categorized as explicit or implicit. Explicit tontines use predefined rules to redistribute assets upon a member's death, providing an explicit 'longevity credit' each year. Notable works in this field include the Fair Tontine Annuity (Sabin 2010), the Annuity Overlay Fund (Donnelly, Guillén, and Nielsen 2014), and the Pooled Annuity Fund (Stamos 2008). In implicit tontines, individuals receive income based on the number of survivors in the pool. Key publications in this area include The Group Self-Annuitization (Piggott, Valdez, and Detzel 2005) and the Optimal Retirement Tontine (Milevsky and Salisbury 2015).

[1] Details available at QSuper.
[2] Details available at La Tontine Par le Conservateur.
[3] Details available at Tontine Trust.

2.2 The design of a Natural Tontine

In a historical tontine, the dividend rate offered to investors remains constant, leading to increasing payments for the survivors as the number of individuals in the fund decreases exponentially over time. The initial example illustrates this: 1000 individuals invest \$1000 each and the fund pays a 3% coupon in perpetuity and distributes this amount among the survivors. At the first payment date, if all the individuals are still alive, each one of them receives $\frac{1000\times\$1000}{1000} \times 3\% = \$30$. If at future date t there are only 800 survivors in the fund, the tontine dividend is $\frac{1000\times\$1000}{800} \times 3\% = \$37.5$, where \$30 are the coupon payment and \$7.50 are the mortality credits. The coupon payment made to each survivor is constant over time, but as the number of remaining investors decreases, each survivor's payment increases due to the mortality credits. This payment profile may not be attractive to pensioners who prefer stable cash flows over time.

To address this, Milevsky and Salisbury (2015) designed a new tontine structure called the 'Natural Tontine'. This structure aims to determine a payment rate that guarantees a constant stream of income over time, thereby maximizing the discounted expected utility of consumption for retirees.

The authors of the Natural Tontine proposed a product that is expected to cost 1\$ and that continuously pays out the amount $d(t)$ rather than at a monthly or annual frequency. Furthermore, the authors also assume a constant risk-free interest rate r and an objective survival function ${}_tp_x$, that identifies the probability of individual x to survive t years. The obvious comparator for this sort of product is an annuity in which the annuitant pays 1\$ at $t = 0$ and then receives a lifetime income stream of $c(t)$, the payout rate per survivor. For this annuity to be fairly priced, with a sufficiently large client base, all the payments made in perpetuity must be funded by the initial premium invested at the risk-free rate, implying the following constrain on the annuity payout function $c(t)$:

$$\int_0^\infty e^{-rt}\, {}_tp_x\, c(t)\, dt = 1 \tag{2.1}$$

Letting $u(c)$ denote the instantaneous utility of consumption, an annuitant with lifetime ζ and without intention to leave any bequest to its heirs will choose a life annuity payout function for which $c(t)$ maximizes the discounted lifetime utility:

$$\max_{c(t)} \mathbb{E}\left[\int_0^{\zeta} e^{-rt} u(c(t))\, dt\right] = \max_{c(t)} \left[\int_0^{\infty} e^{-rt}\ {}_tp_x\, u(c(t))\, dt\right]$$

By the Euler-Lagrange Theorem, it is possible to prove the existence of a constant λ such that:

$$e^{-rt}\ {}_tp_x u'(c(t)) = \lambda e^{-rt}\ {}_tp_x \quad \text{for every } t$$

Simplifying the equation on both sides shows that $u'(c(t)) = \lambda$ which is constant and if the utility function $u(c)$ is strictly convex, the optimal annuity function $c(t)$ is also constant. This stable level of income can be determined by solving the fair pricing constraint of the annuity in equation 2.1, showing that the optimized life annuity payout function is:

$$c_t \equiv c_0 = \left[\int_0^{\infty} e^{-rt}\ {}_tp_x\, dt\right]^{-1} \tag{2.2}$$

Moving to the tontine structure proposed by Milevsky and Salisbury, the main purpose is to maximize the expected utility by determining the optimal level of payout $d(t)$, i.e. the percentage of the fund value paid to the annuitants. Indeed, differently from the fair annuity or the the historical tontine, there is no point in the dividend payout of the tontine to be fixed at a constant level, in which every payment is always the same fixed percentage of the current value of the fund. At the inception of the tontine, n soon-to-be retirees will each pay 1$ to buy the annuity, and in every point in the future, the number of individuals still alive and hence in the fund is a random variable $N(t)$. Assuming an homogeneous cohort and that an individual is alive, at the future time t, the number of survival is given by a binomial distribution with parameter ${}_tp_x$, hence, stating it more precisely, $N(t) - 1 \sim Bin(n-1, {}_tp_x)$. The individual's discounted lifetime utility is given by:

$$\begin{aligned}
\max_{d(t)} \mathbb{E}\left[\int_0^{\zeta} e^{-rt} u\left(\frac{nd(t)}{N(t)}\right) dt\right] &= \max_{d(t)} \int_0^{\infty} e^{-rt}\, t_p\, \mathbb{E}\left[u\left(\frac{nd(t)}{N(t)}\right) \mid \zeta > t\right] dt \\
&= \max_{d(t)} \int_0^{\infty} e^{-rt}\, t_p \sum_{k=0}^{n-1} \binom{n-1}{k} (tp_x)^k (1 - tp_x)^{n-1-k} \\
&\quad \times u\left(\frac{nd(t)}{k+1}\right) dt
\end{aligned}$$

where $nd(t)$ is the total fund payment at time t, which must be divided by the number of people that are still alive, given by the random variable $N(t)$. As in the previous case, the maximization problem is subject to a constraint. The sponsor of the tontine should be able to sustain the withdrawals in perpetuity with the n premiums initially collected. This constraint is expressed by the following equation:

$$\int_0^\infty e^{-rt} d(t)\, dt = 1 \tag{2.3}$$

At this point it would be possible to set $d(t) = d(0)$ and force it to remain constant at the value r, but as it will be showed the optimal value will be far from constant. By the Euler-Lagrange Theorem it follows that there is a constant λ such that the optimal $d(t)$ satisfies for every t:

$$e^{-rt}\, {}_tp_x \sum_{k=0}^{n-1} \binom{n-1}{k}\, {}_tp_x^k (1 - {}_tp_x)^{n-1-k} \frac{n}{k+1} u'\left(\frac{nd(t)}{k+1}\right) = \lambda e^{-rt}$$

Assuming Constant Relative Risk Aversion (CRRA), the solution is greatly simplified. Letting:

- $u(c) = c^{1-\gamma}/(1-\gamma)$ if $\gamma \neq 1$
- $u(c) = log(c)$ when $\gamma = 1$

and defining

$$\theta_{n,\gamma}(p) = E\left[\left(\frac{n}{N(p)}\right)^{1-\gamma}\right] = \sum_{k=0}^{n-1} \binom{n-1}{k} p^k (1-p)^{n-1-k} \left(\frac{n}{k+1}\right)^{1-\gamma}$$

where $N(p) - 1 \sim Bin(n-1, p)$. By setting $\beta_{n,\gamma} = p\theta_{n,\gamma}(p)$ is obtained that under the initial constraint in equation 2.3 and CRRA utility, the optimal tontine withdrawal rate is $D^{OT}_{n,\gamma}(p) = D^{OT}_{n,\gamma}(1)\beta_{n,\gamma}(p)^{1/\gamma}$, where

$$D^{OT}_{n,\gamma}(1) = \left[\int_0^\infty e^{-rt} \beta_{n,\gamma}({}_tp_x)^{1/\gamma} dt\right]^{-1}$$

For a sufficiently large n, the Natural Tontine is quite optimal for all given risk aversion coefficients γ. Indeed, when an individual has $\gamma \neq 1$ the welfare loss they experience is very small compared to the tontine structure that is optimal only with

$\gamma = 1$, leading the authors to conclude that the Natural Tontine can be the basis for future products of this kind as it is quite optimal for all longevity risk aversion coefficients and precisely optimal when considering a logarithmic utility function.

But how does the payment profile of this product compare to a traditional tontine?

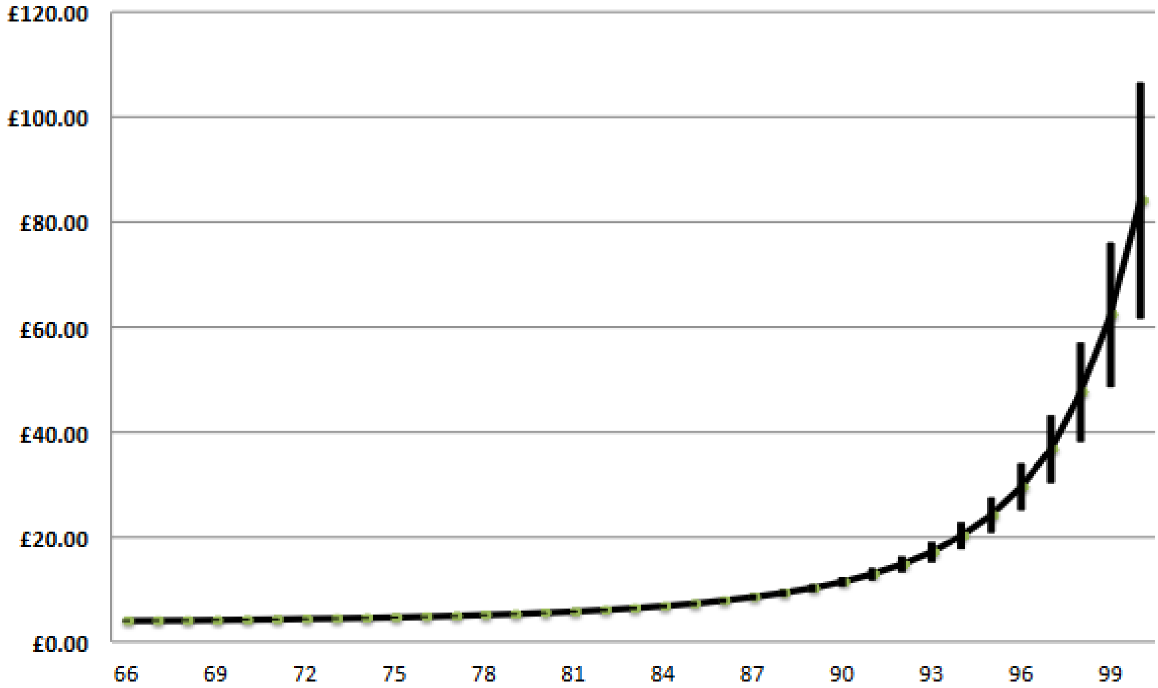

Figure 2.1: Range of Flat 4% Tontine Payout purchased at 65: Gompertz Mortality assuming n=400, m=88.721, b=10. Source: Milevsky and Salisbury 2015

In a traditional tontine scheme such as the one represented in Figure 2.1, individuals can expect low payments early on that will get very large, but also variable, in the last couple of years thanks to the mortality credits. However this payments are only experienced if one is fortunate enough to be still alive at the time. Individuals are more interested in maximising their utility (and standard of living) over the entire lifetime, hence this payment profile is not very appealing.

In contrast, in the Natural Tontine proposed by the authors, the expected tontine payout per survivor is relatively constant over time, and therefore also the discounted expected utility is much higher than in the traditional structure. Figure 2.2 represents the payout profile of a Natural Tontine. The chart is obtained by solving for the value of $D^{OT}_{n,\gamma}(1)$ and then reconstructing $D^{OT}_{n,\gamma}({}_tp_x)$ for $n = 400$, $r = 0.04$ and $\gamma = 1$.

It is important to distinguish once more between the actual interest rate paid to the entire pool (i.e. the percentage of the fund value that is distributed) and the expected

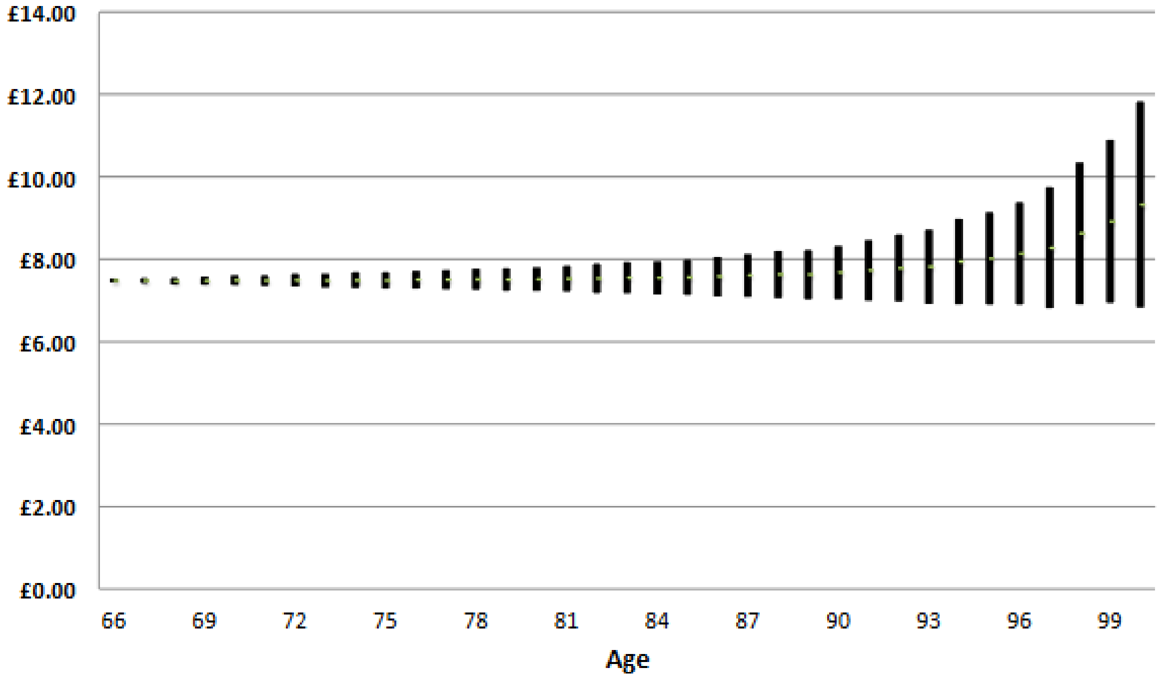

Figure 2.2: Range of Optimal Tontine Payout at 4%: Gompertz Mortality assuming n=400, m=88.721, b=10. Source: Milevsky and Salisbury 2015

dividend of the Natural Tontine which will be relatively stable over time.

The bottom line of the Natural Tontine is that the present value of the interest paid to the entire pool over time is exactly equal to the original contribution made by the pool itself. There is simply rearrangement and parsing of the cash flows over time. The historical tontine in which dividends to the entire pool are a constant percentage (e.g. 4%) of the value of the fund are suboptimal because they create an increasing consumption profile that is undesirable due to its variability. However, a tontine scheme in which interest payments made to the entire pool early on are higher (e.g. 8%) and then decline over time to lower interest rate (e.g. to 1%), is in fact the optimal design, because the dividend amount (in dollar terms) is more constant as the total payment made by the fund is to be divided among a shrinking number of individuals.

2.3 Building a simulation of a Natural Tontine

In the previous section, it was discussed how the Natural Tontine design proposed by Milevsky and Salisbury can be considered optimal, meaning that it maximize the discounted expected utility of consumption and allows for stable cash flows over time.

In this section, a simulation for a Natural Tontine Fund is created. The simulation integrates two main sources of randomness: one for the investment returns and one for the mortality model to project the number of survivors. These projections are then combined deterministically using a specific sharing rule to determine the tontine dividend. Initially, the structure of the fund is introduced using a 'toy example' to demonstrate the properties of this type of product. For this example, simplifying assumptions are made: LogNormally Distributed returns and a Gompertz Mortality Law are assumed.

Later, in subsection 2.3.5, the quality of these assumptions is assessed, and real data is used to evaluate how this fund would have performed.

2.3.1 Modelling mortality: Gompertz Mortality Law

In a tontine, each dividend payment is distributed among the remaining participants. Therefore, the simulation first requires a method to estimate the number of survivors at each future payment date. This necessitates the introduction of a survival probability curve, which represents the likelihood that an investor alive at age x will still be alive at age $x + t$, and thus be eligible for payments at time t.

In the 19th century, the British actuary Benjamin Gompertz proposed a mortality law, now known as Gompertz's law, which posits that the natural logarithm of adult mortality rates increases linearly with age, implying exponential growth in mortality rates over time. Gompertz observed that not only do mortality rates increase with age, but they do so at approximately the same percentage each year. Specifically, if the mortality rate at age y is $q\%$, then at age $y + 1$, the mortality rate becomes $q(1 + z\%)$, and at age $y + 2$, it becomes $q(1 + z\%)^2$, and so forth. Consequently, mortality rates are an exponentially increasing function of age, meaning that when the logarithm of annual mortality rates is computed, it approximates a straight line determined by a slope parameter and an intercept parameter.

The instantaneous hazard rate, representing the death rate at a given age, is denoted by λ_x, where x is the individual's age. According to Gompertz's law, this rate is expressed as:

$$\lambda_x = h_0 e^{gx}$$

Here, h_0 and g are constants, leading to the equivalent logarithmic form:

$$\ln[\lambda_x] = \ln[h_0] + gx$$

In this equation, h_0 represents the initial mortality hazard rate at time zero (the intercept), and g denotes the mortality growth rate (the slope coefficient). To simplify the subsequent mortality modelling, the hazard rate can be parameterized as follows (see Milevsky 2022):

$$\lambda_x = \frac{1}{b} e^{(x-m)/b}$$

Taking the natural logarithm, this becomes:

$$\ln[\lambda_x] = \underbrace{-\ln[b] - \frac{m}{b}}_{\ln[h_0]} + \underbrace{\frac{1}{b}}_{g} x$$

Taking the natural logarithm, this becomes: Furthermore, the survival probability can be formulated as:

$$\Pr[T_x \geq t] = \exp\left\{e^{(x-m)/b}\left(1 - e^{t/b}\right)\right\}$$

where $\Pr[T_x \geq t]$ represents the probability that the remaining lifetime random variable T_x is at least t.

The parameters in this model have specific interpretations: m is the modal value of life expectancy, which is different from the average value, and b is the dispersion parameter, which is not the same as the standard deviation.

```python
def TPXG(x, t, m, b):

    survival_prob = np.exp(np.exp((x - m) / b) * (1 - np.exp(t / b)))

    return survival_prob
```

Listing 2.1: Python function to compute the survival probability according to a

Gompertz Mortality Law

The Python function represented in Listing 2.1 will be utilized in the remainder of this example to compute survival probabilities based on the Gompertz Mortality Law. This function requires four main inputs:

- `x` is the current age of the individual,
- `t` is the time horizon for which the survival probabilities are computed,
- `m` and `b` are the parameters of the Gompertz Mortality Law introduced earlier.

This function calculates the probability that an individual aged x will survive for another t years according to the Gompertz Mortality Law. For example, the probability that a 60-year-old will live another 40 years (and thus reach 100 years old) is 6.94%, assuming $m = 90$ and $b = 10$. However, if the individual is already 80 years old, the probability of reaching 100 increases to 17.94%, because the probability is conditional on the individual surviving until age 80, thus discounting the probability of dying between ages 60 and 80. The older the individual, the greater the probability of reaching any given age.

To understand the impact of the Gompertz parameters on survival probabilities, consider the following examples: if the parameter m is set to 100 instead of 90, the probability that a 60-year-old will live for another 40 years increases to 37.47%. Conversely, if the parameter b is changed from 10 to 5 (while resetting m to 90), the probability of a 60-year-old reaching the age of 100 drops to a mere 0.06%. Indeed, a smaller dispersion parameter b translates to less variation in the age of death, making it increasingly difficult to outlive the modal age.

In section 2.2, it was argued that, assuming a homogeneous cohort of individuals, the number of survivors at the future time t, represented by the random variable $N(t)$, is given by a binomial distribution with parameter ${}_tp_x$, hence, $N(t) \sim Bin(n, {}_tp_x)$. The Gompertz Mortality Law can be used as the parameter of the binomial distribution to model the number of individuals alive at time t.

```python
def deaths_simulation(N, TH, GLO, x, m, b):

    # Placeholders
```

```
    GLIVE = np.zeros((N, TH), dtype=int) # Matrix for survivors
    GDEAD = np.zeros((N, TH), dtype=int) # Matrix for deaths

    # Simulate the deaths and survivors
    for i in range(N):
        # Simulate deaths in year 1
        prob_death_year_1 = 1 - TPXG(x, 1, m, b)
        GDEAD[i, 0] = binom.rvs(GL0, prob_death_year_1)
        # Subtract deaths from GL0 to get survivors
        GLIVE[i, 0] = GL0 - GDEAD[i, 0]
        # Loop through remaining years
        for j in range(1, TH):
            prob_death_year_j = 1 - TPXG(x + j, 1, m, b)
            GDEAD[i, j] = binom.rvs(GLIVE[i, j - 1], prob_death_year_j)
            # Count survivors
            GLIVE[i, j] = GLIVE[i, j - 1] - GDEAD[i, j]

    # Now GLIVE and GDEAD matrices contain the simulated paths
    return GDEAD, GLIVE
```

Listing 2.2: Python function to create N simulated paths of the number of survivors according to the Gompertz Mortality Law

The Python script reported in Listing 2.2 generates `N=10000` simulation paths for the evolution of the number of participants in the fund. It assumes an initial number of `GL0=1000` participants and a fund time horizon of `TH=30` years, after which the fund will be shut down. Each year, the number of deaths and the number of people still alive are recorded in the arrays `GDEAD` and `GLIVE`, respectively.

In the script, the Gompertz Mortality Law was used to compute the survival probability for an individual between ages x and $x+1$. `(1 - TPXG(x, 1, m, b))` was then used as the parameter for a binomial distribution to generate the actual number of individuals in the pool who died each year in each simulation.

Figure 2.3 presents the 98% confidence interval for the number of deaths and the number of people alive for each year considered. There is a notable difference between the 99th percentile and the 1st percentile, illustrating the substantial variability of

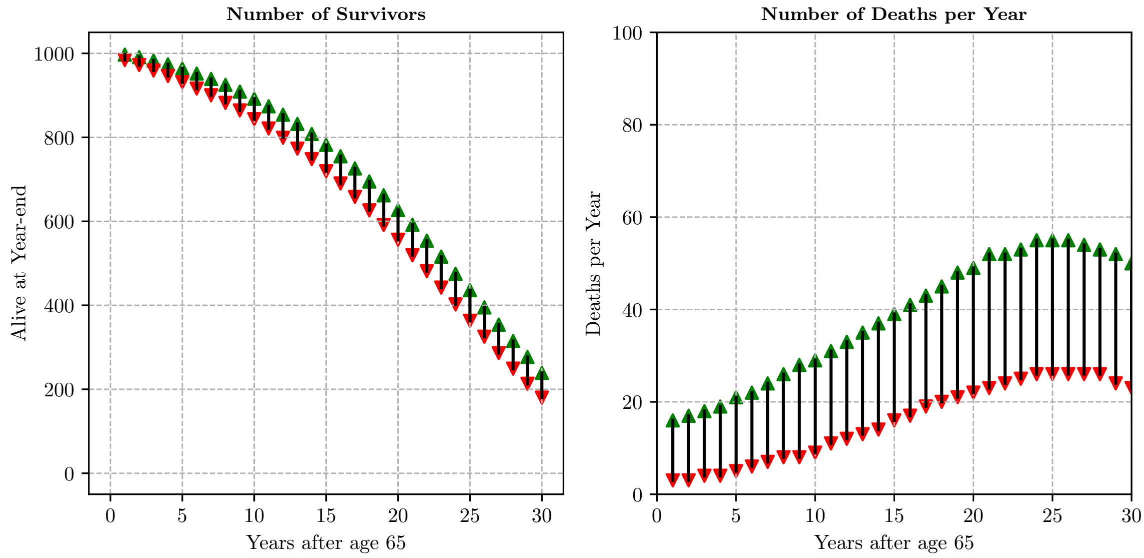

Figure 2.3: 98% confidence intervals for simulated lifespan and deaths. Original Pool Size = 1000. Fixed $m = 90$ and $b = 10$ with Gompertz Mortality Law.

the simulated paths, even with a relatively large initial pool of 1000 individuals. The variability can be explained by the fact that the Gompertz Mortality Law computed the expected number of survivors each year, while the actual number of survivors is given by a binomial distribution with a parameter equal to expected number of deaths.

2.3.2 Modelling Returns: LogNormal distribution

This initial 'toy example' will assume that returns are LogNormally distributed and time-independent. Using the notation $\tilde{R}[i, j]$ to denote the effective periodic return in simulation i and year j, the assumption of LogNormality implies that $\ln[1 + \tilde{R}[i, j]]$ is normally distributed. The quantity $\ln[1 + \tilde{R}[i, j]]$ is also known as the continuously compounded investment return, represented by $\tilde{r}[i, j]$, where $\tilde{R}[i, j] = e^{\tilde{r}[i,j]} - 1$. In the following paragraphs, the expected value of the continuously compounded returns will be denoted by ν and the standard deviation by σ. It is important to note that under the assumption of continuously compounded returns $\tilde{r}$, the theoretical skewness and kurtosis are 0 and 3, respectively.

In simulating portfolio returns, the most important variable is the continuously compounded long-term *assumed* rate of return (ARR) earned by the Natural Tontine, denoted as r. In this initial 'toy example', it was assumed r to be 4%. This figure is crucial for determining the initial tontine dividend payout and for forecasting dividend payouts to surviving investors in the coming years.

This assumed rate of return will differ from the realized investment return (RIR), which is denoted by $r[i, j]$, where i represents the simulation number and j represents the year in which the return is obtained. Although the RIR will vary from the ARR, the average RIR across a large number of simulations will converge to the ARR.

```
# Set base parameters
EXR = 0.04 # Mean
SDR = 0.02 # Standard Deviation
TH = 30 # number of years to simulate
N = 10000 # number of simulation

# Define placeholders
PORET = np.zeros((N, TH)) # matrix for portfolio returns
STPRV = np.zeros((N, TH)) # matrix for sthocastic present values

# Simulate N paths of TH returns
for i in range(N):
    PORET[i, :] = np.exp(np.random.normal(EXR, SDR, TH)) - 1
```

Listing 2.3: Python script to simulate the portfolio return assuming a LogNormal distribution

Script 2.3 generates the matrix `PORET`, which contains the portfolio returns for each year. As previously mentioned, to generate these returns a LogNormal distribution with a mean value (ν) of `EXR` = 0.04 and a standard deviation (σ) of `SDR` = 0.03 was assumed. It is important to note that `EXR` represents the expected continuously compounded investment return, also described as the *geometric mean* of the investment returns, and is equivalent to r.

Figure 2.4 depicts the 98% confidence interval for the investment return in each year. The parameters that generated the returns remained constant across all years, resulting in no significant differences between the returns from year to year, given the large number (10,000) of simulations.

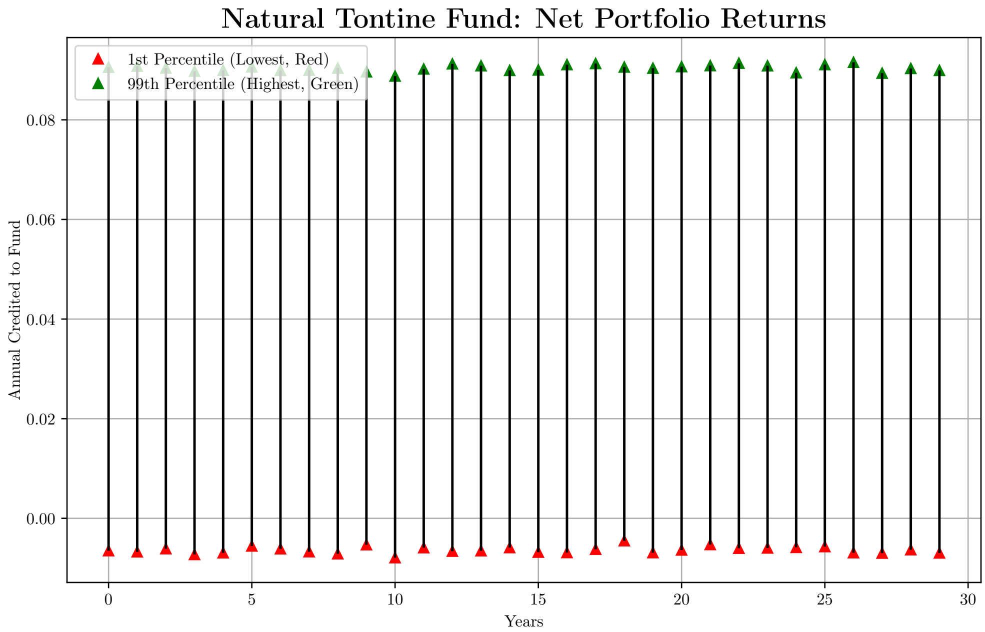

Figure 2.4: 98% confidence intervals for simulated returns. LogNormally distributed returns with $\nu = 0.04$, $\sigma = 0.03$

2.3.3 The Tontine Dividend Rate

The Gompertz Mortality Model deaths and the LogNormally distributed returns are the two sources of randomness in this initial example of the Natural Tontine Fund. However, to derive the tontine dividend, a deterministic rule is needed.

The Tontine Dividend Rate, denoted by κ_i, is the percentage of the fund value distributed to surviving members at the end of period i. For instance, if F_{10} denotes the fund value at the end of year 10, the tontine dividend shared by all survivors is $\kappa_{10} \times F_{10}$, and the tontine dividend per survivor is $\frac{\kappa_{10} \times F_{10}}{GL_{10}}$, where GL_{10} is the number of survivors at the beginning of year 10.

If the sponsor of the tontine fund wishes to maintain a stable dividend for survivors over time, κ_i is determined endogenously rather than being set exogenously by the sponsor.

The value of κ_i is the reciprocal of the price of a *Temporary Life Income Annuity* that pays an annual cash flow to an individual aged x until death or a specified date, whichever comes first.

In the previous chapter, it was discussed how the authors of the Natural Tontine structure proved that for a fairly priced *Life Only Immediate Annuity*, which allows an individual to pay a certain premium upfront and receive a lifetime income, the payout that maximizes the utility for the annuitant is constant.

This property of the Life Only Income Annuity shown in equation 2.2 can rewritten as:

$$c_0 = \frac{1}{\int_0^\infty e^{-rt}\, {}_tp_x \, dt}$$

The denominator of the above expression is the discounted actuarial present value of an annuity, often represented as $a(x, r)$. In general, the discounted actuarial present value of a life annuity is given by (see Promislow 2014):

$$a(x, r) = \int_0^\infty e^{-rt} \Pr[T_x \geq t] dt = \int_0^\infty e^{-rt} ({}_tp_x) dt$$

Assuming Gompertz Mortality, there is an analytical solution for $a(x, r)$ (see Milevsky 2006), expressed as:

$$a(x, r) = \frac{b\Gamma(-rb, e^{(x-m)/b})}{\exp\{(m - x)r - e^{(x-m)/b}\}},$$

where $\Gamma(A, B)$ is the incomplete Gamma function.

The Python function reported in Listings 2.4 shows how, in discrete time, the value of a Temporary Life Income Annuity is the sum from age x to age y of the discounted actuarial present values, which are the product of the Gompertz Survival Rates and constant interest discount rates (see Milevsky 2022).

```
def TLIA(x, y, r, m, b):
    # Actuarial Present Value (APV) function
    def APV(t):
        return np.exp(-r * t) * TPXG(x, t, m, b) # product of the
            Gompertz mortality and the constant interest dicount rates

    # Summation of APV from 1 to (y - x)
    apv_sum = sum(APV(t) for t in range(1, y - x + 1))
    return apv_sum1
```

Listing 2.4: Python function to value a Temporary Life Income Annuity

For example, if a 65-year-old buys a Temporary Life Income Annuity that pays a \$1 dividend until age 95, assuming $r = 4\%$, $m = 90$, and $b = 10$, the annuity value is \$13.04. However, if the annuity pays dividends until age 100, the value increases slightly to \$13.20 due to the longer payment period.

As mentioned at the beginning of this section, the reciprocal of the value of a Temporary Life Income Annuity is κ_i, the Tontine Dividend Rate. For instance, the dividend received by a 65-year-old would be 7.67% if the annuity paid dividends until age 95, or 7.55% if dividends were paid until age 100. Extending the payment horizon further will not significantly reduce the yield, as the present value of payments and survival probabilities diminish.

2.3.4 The Natural Tontine Fund

This section explains how the Natural Tontine Fund values are simulated using the random returns and lifetimes generated in previous sections, along with the deterministic rule for determining the Tontine Dividend Rate.

Although the core of this chapter aims to bridge the theoretical advancements of the Natural Tontine with practical applications, this is not intended as a step-by-step guide for building a Natural Tontine fund business in real life. Therefore, two main assumptions are made to simplify the computations without affecting the interpretation of the fund's mechanism.

First, it is assumed that all fund members are homogeneous and belong to the same cohort, sharing the same age, birthday, and a fund launch date that coincides with their birthday. Additionally, each participant is considered to have the same life expectancy. Secondly, it is assumed that all participants invest the same dollar amount in the fund.

While incorporating a mix of ages, genders, and varying investment amounts does not alter the fundamental mechanics of the fund, it does add complexity to the computations, making the underlying fund mechanism harder to understand.

For the remainder of the chapter, the following values will be used for the inputs:

- The starting age `x` is 65 years.
- The horizon of the fund `TH` is 30 years.
- `GL0`, the initial number of participants in the fund, is 1,000.
- The initial investment `f0` is \$100,000 each.

- 10,000 simulations (N).
- Returns are generated assuming an expected return EXR=4% (equivalent to the discount rate r) and a standard deviation SDR=3%.
- The number of survivors is generated using a Gompertz Mortality Law with parameters m=90 and b=10.

```
def tontine_fund_nominal(N, TH, x, r, m, b, f0, GL0, PORET, GLIVE):

    # Define placeholders
    DETFV_nominal = np.zeros((N, TH))
    TONDV_nominal = np.zeros((N, TH))

    # Calculate vector of kappa values
    kappa = np.array([1 / TLIA(x + i - 1, x + TH, r, m, b) for i in range(1, TH + 1)])

    # Simulate Natural Tontine Fund with stochastic returns and deaths
    for i in range(N):
        # Dividend and fund value at end of year 1
        TONDV_nominal[i, 0] = kappa[0] * f0
        DETFV_nominal[i, 0] = f0 * GL0 * (1 + PORET[i, 0]) - TONDV_nominal[i, 0] * GLIVE[i,0]

        # Dividend and fund value for remaining years
        for j in range(1, TH):
            TONDV_nominal[i, j] = kappa[j] * DETFV_nominal[i, j - 1] / GLIVE[i,j-1]
            DETFV_nominal[i, j] = max(DETFV_nominal[i, j - 1] * (1 + PORET[i, j]) - TONDV_nominal[i, j] * GLIVE[i,j], 0)

    return DETFV_nominal, TONDV_nominal
```

Listing 2.5: Python function to simulate a Natural Tontine Fund

Script 2.5 computes the evolution of fund values, by implementing a recursive methodology involving the following steps:

1. The algorithm starts by setting the initial Natural Tontine Fund value to the initial investment of each individual multiplied by the initial number of members.

2. The fund value increases by the investment return in the first year and is then reduced by the total tontine dividend payout made to survivors at the end of the year.

3. The fund value at the end of the first year becomes the fund value at the beginning of the second year, and the second step of the algorithm is repeated until the final year.

The Tontine Dividend Rate κ_i, represented in the code by the variable `KAPPA`, is central to the Natural Tontine Fund. It is built to maximize participants' utility. As mentioned in section 2.3.3, κ_i is computed as the reciprocal of the value of an annuity bought by an `x`-year-old that pays dividends until age `x` + `TH`. For example, at the fund's inception, all annuitants are 65 years old, and κ_0 is the reciprocal of the value of an annuity bought by a 65 year old investor that pays dividends for the next 30 years, until age 95. In the second year, κ_1 is the reciprocal of the value of an annuity bought by a 66 year old investor that pays dividends for the next 29 years, still until age 95. Therefore, the Tontine Dividend Rate is determined at the fund's inception and increases over time as the annuitants age.

The fund adjusts the yearly payout such that, in times of loss, the sponsor reduces the tontine dividend in dollar terms. When the fund performs better than expected, the dividends increase in dollar terms. This is achieved using the factor κ, which is not constant from year to year as in a classical tontine, but is designed to be equitable for all participants and safe for the sponsors. It allows modulating payouts to keep the fund solvent.

Even though κ is determined at time zero, the dollar value of tontine dividends is not set in advance, as it depends on actual investment returns and the number of survivors. For example, if in year 7 $\kappa_7 = 8\%$ and the fund value is \$50 million with 840 survivors, each participant would receive a tontine dividend of $\frac{50{,}000{,}000 \times 0.08}{840} = \$4,761.90$. If there are more survivors, such as 910, the structure of the Natural Tontine adjusts the dividend to keep the fund solvent. Assuming the fund value is still \$50 million and $\kappa_7 = 8\%$, the tontine dividend is reduced to $\frac{50{,}000{,}000 \times 0.08}{910} = \$4,395.60$.

Similarly, if the fund value is lower, the κ mechanism reduces the tontine dividend to maintain solvency. Conversely, if there are fewer survivors or a higher fund value, κ modulates the payments accordingly, allowing for larger dividend payments.

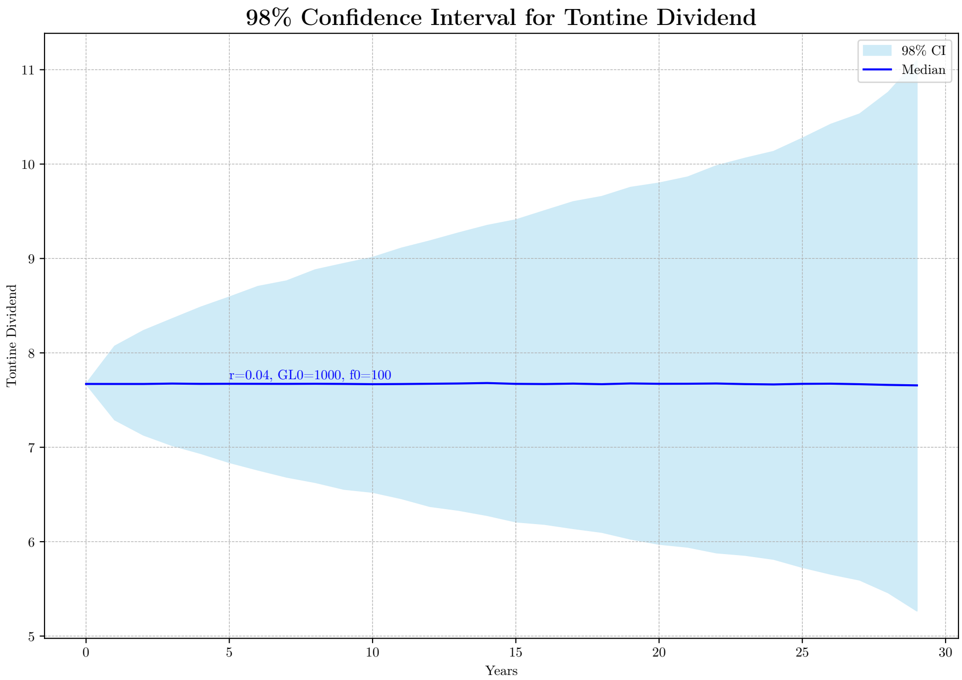

Figure 2.5: 98% confidence intervals for Natural Tontine dividend, assuming sthocastic returns and Gompertz Mortality

Figure 2.5 displays the 98% confidence interval of the Natural Tontine Dividend paid to each survivor each year. The tontine dividend is set at the beginning of each year and paid to survivors at the end of the year. Therefore, in the first year, the dividend is fixed across all scenarios at `KAPPA[0]`$\times$`f0`, which in the example is \$7,670.86. As the years progress and returns and survivor numbers vary across simulations, uncertainty increases, as seen in the widening gap between the 99th and 1st percentiles.

Despite the increased variability in the dollar amount of tontine dividends, the median value remains constant over time. The median dividend over 30 years is \$7,670.86, matching the initial dividend, while the mean is slightly higher at \$7,697.86, showing slight upward asymmetry. This demonstrates that the Natural Tontine structure maintains a flat payment profile over time, resulting in a larger discounted expected utility compared to a classical tontine where payments increase over time. To prove this feature, it is possible to fit a linear regression of the median tontine dividend amount against a constant and a time trend. The regression slope is statistically indistinguishable from zero, confirming the stability of tontine dividends over time as shown in 2.6.

```
mtd = []
t = np.arange(1, TH + 1)
for j in range(TH):
    mtd.append(np.quantile(TONDV[:, j], 0.50))
mtd = np.array(mtd)


t = np.arange(1, TH + 1)
t_with_const = sm.add_constant(t)  # Add a constant term for the
    intercept
model = sm.OLS(mtd, t_with_const).fit()

#model.summary()

# Extract the coefficients
intercept = model.params[0]
slope = model.params[1]

print(f"Intercept: {intercept:.04f}")
> Intercept: 7.6745

print(f"Slope: {slope:.04f}")
> Slope: -0.0007
```

Listing 2.6: OLS regression to prove the stability of the median tontine dividend

Figure 2.6 shows the 98% confidence interval of the Natural Tontine Fund value each year. Here, the band between the upper and lower percentiles remains constant over time and narrows as the fund's horizon approaches. The scarce variability in the fund value across simulation is guaranteed by the role of κ. If the fund experiences very good returns in one year, κ will result in a larger dividend payout to survivors the next year, reducing the fund value and preventing excessive growth. This mechanism ensures perfect decumulation of the fund value over time, with the entire value paid

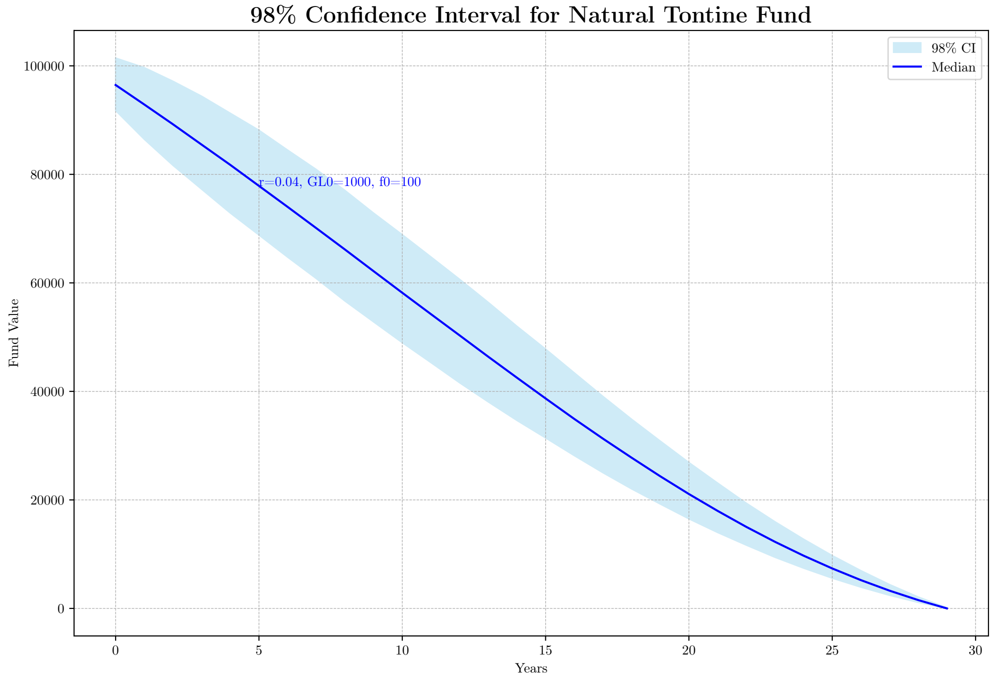

Figure 2.6: 98% confidence intervals for Natural Tontine Fund value, assuming sthocastic returns and Gompertz Mortality

out to the remaining individuals in the final year.

2.3.5 From fictionary data to a real world example

The previous section of this chapter outlined the main features of a fund based on the Natural Tontine structure developed by Milevsky and Salisbury (2015). To keep the example simple and focus on the insights of the tontine structure rather than on the inputs of the model, simplifying assumptions about the two sources of randomness, the number of survivors and the investment returns, were made. The purpose of this section is to transition from the simplified example to a model that incorporates real-world data instead of fictional inputs.

Fitting a real Mortality Table

In recent years, the Gompertz Mortality Model has faced increasing criticism in academic literature (for example see Li et al. 2021), as the law appears to be inaccurate at advanced ages, where the mortality curve tends to plateau. Since the focus of this study is on retirees, this issue is particularly relevant. Most insurance actuaries nowadays work with discrete mortality tables while academic researchers in actuarial finance prefer to operate with continuous mortality rates such as the Gompertz Law. In this subsection will be laid down a methodology to fit a discrete mortality table to the Gompertz Mortality Law parameters m and b to overcome the shortcomings of the Law.

However, before explaining how a discrete mortality table can be approximated by the Gompertz Mortality Law, it is important to highlight the significance of setting the correct parameters. In the 'toy example', it was assumed a Gompertz Mortality Law with parameters $m = 90$ and $b = 10$ to generate the number of survivors and derive the value of the Temporary Life Income Annuities. As a result, individuals died and survived at the exact rate assumed in the annuity valuation. Setting accurate parameters for annuity valuation is essential to maintain a constant dividend. If annuitants live longer than assumed, the tontine mechanism would have to reduce the dividend in dollar terms to remain solvent.

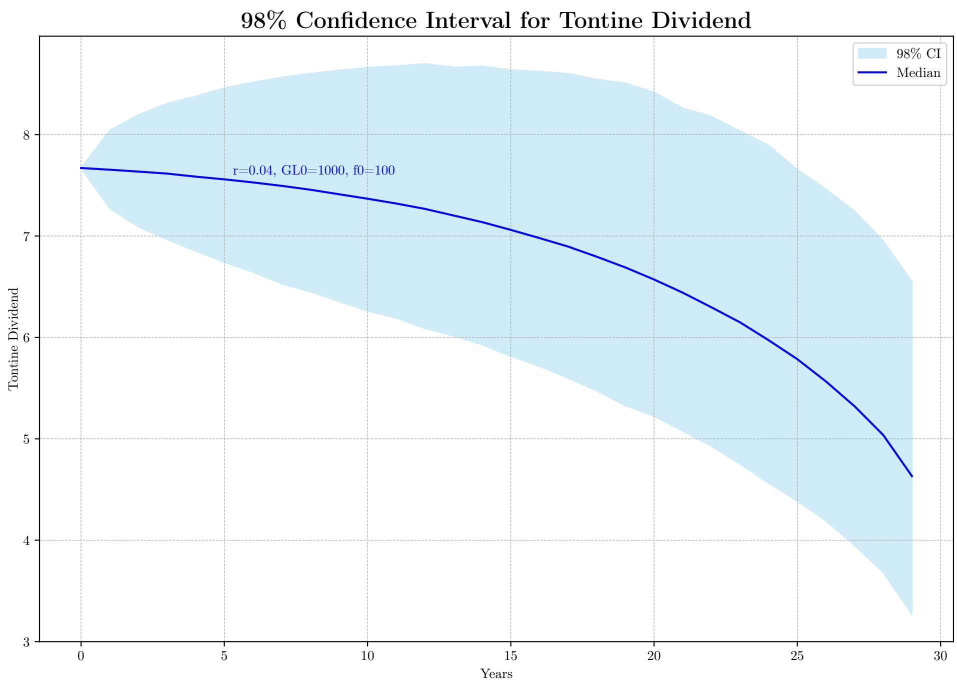

Figure 2.7: 98% confidence intervals for the Natural Tontine dividend, assuming that the number of survivors is generated with $m = 93$ and the annuities are priced with $m = 90$

Figure 2.7 illustrates what would happen if people lived longer than expected by the fund's sponsors, for example, if the survival function used $m = 93$ while the annuity was valued with $m = 90$. It is possible to observe that the tontine dividend is not stable over time and exhibits a downward trend. By running a regression of the median tontine dividend against a constant and a time trend, the slope coefficient results to be -0.0907, indicating a time dependence of the tontine dividend. This violates one of the main insights of the Natural Tontine structure, which is the stability of the dividend. However, the mechanism of κ ensures that the fund remains solvent and never runs out of money as Figure 2.8 shows.

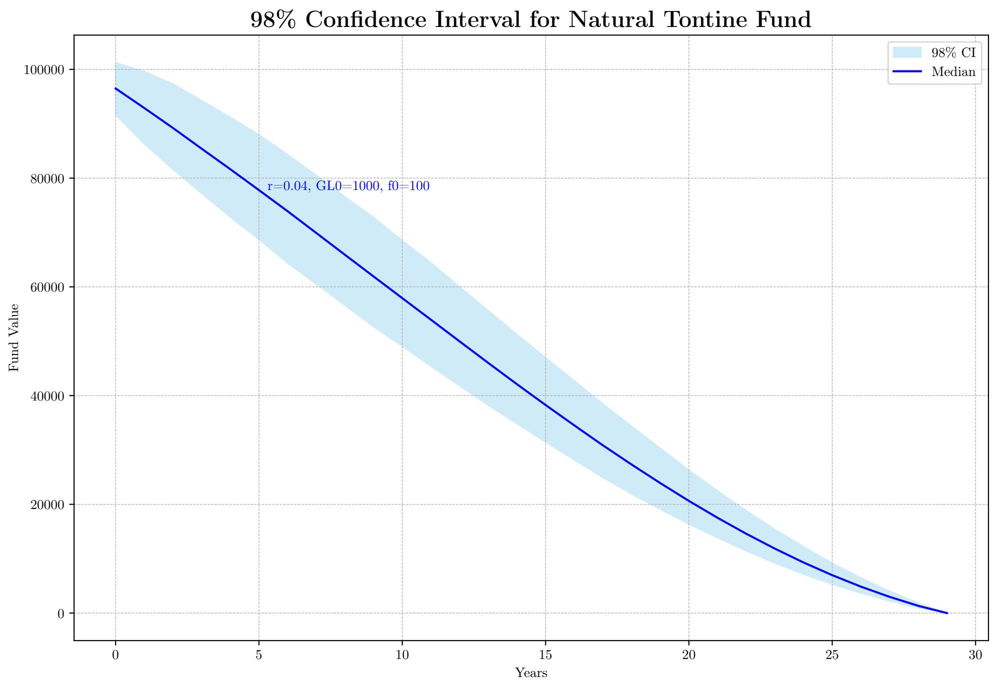

Figure 2.8: 98% confidence intervals for Natural Tontine Fund value, assuming that the number of survivors is generated with $m = 93$ and the annuities are priced with $m = 90$

This example demonstrates the importance of using correct values to price annuities. The state-of-the-art technique in the actuarial field is to use discrete mortality tables to price and value insurance products. This process involves selecting a suitable mortality table and adding improvement factors. Mortality tables contain various information, but for this purpose, the value q_x is relevant. This value represents the one-year mortality rate, given by

$$q_x = \frac{\text{Number dying between age } x \text{ and } x+1}{\text{Number of alive at age } x},$$

the fraction of individuals alive on their birthday who are expected to die before their next birthday. For this analysis, the 2020 Italian mortality table [4] will be used and it will be assumed an equal split between genders.

Given the year by year mortality rates it is possible to derive long-term survival rates under the assumption that the age-based mortality rates, q_x, remain the same over time. Survival probabilities are given by

$$p(x, i) = \prod_{j=0}^{j=i-1} (1 - q_{x+j}),$$

where $p(x, i)$ represents the probability that an x-year-old, who is alive, will survive to his $(x + i)$ birthday and, as described before, q_{x+j} denotes the probability that a $(x + j)$-year-old will die during the next year, before his or her next birthday.

For instance, $p(65, 3)$, the probability that a 65-years-old individual will survive for another 3 years, is the product of the quantities given by 1 minus the three individual mortality rates q_{65}, q_{66}, and q_{67}

By plotting the cumulative survival probabilities of both the Gompertz Mortality Law and the Italian Mortality Table, in Figure 2.9 shows that individuals tend to die younger than predicted by the Gompertz model with the arbitrary parameters $m = 90$ and $b = 10$. This suggests that the model parameters should be revisited to accurately price the annuity, and that higher Tontine Dividend Rates could have been afforded by the sponsor.

Indeed, it is possible to use the mortality table to determine the correct annuity price and show the difference in the prices estimated with the Gompertz Mortality Law. To do this, it must be computed the conditional survival probability from the starting age to the last year of payment and discount these payments at the valuation rate. Summing all the intermediate cash flows gives the annuity value, whose reciprocal is the initial Tontine Dividend Rate. From the script 2.7, using the mortality table, the Tontine Dividend Rate in the first year could have been 7.94% instead of 7.61% from the Gompertz Mortality Law.

[4]Data available at Human Mortality Database.

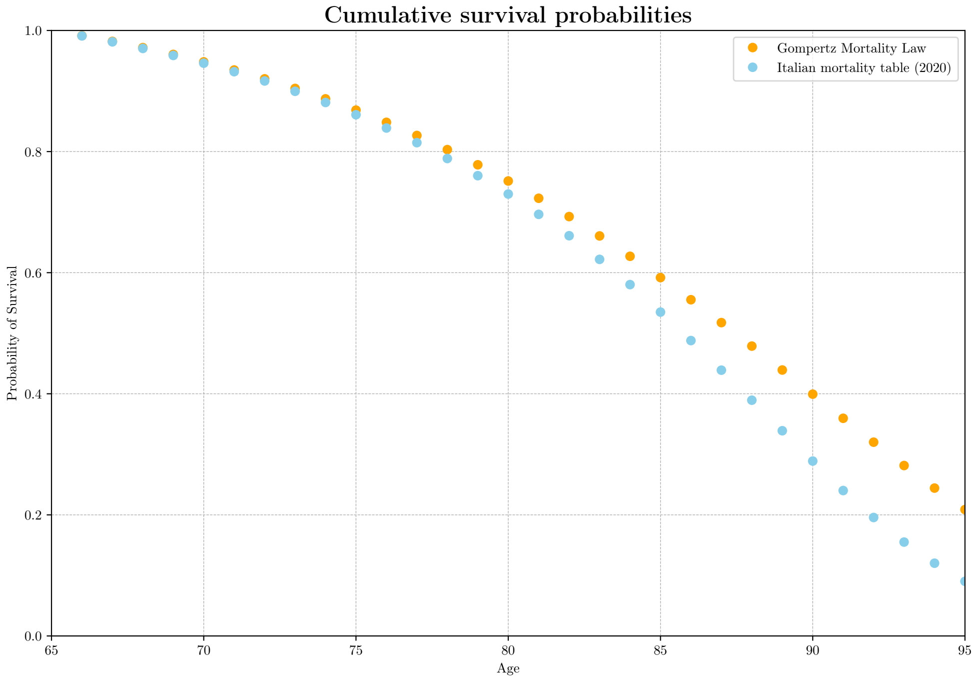

Figure 2.9: Cumulative survival probabilities, Gompertz Mortality model with $m = 90$ and $b = 10$ versus 2020 Italian mortality table

```
# Probability of survival
ps = np.cumprod(1-qx_u_2020)

# Discount rate
dr = np.array(np.cumprod(np.repeat(1/(1+r), len(ps))))

# value of the annuity
tdr_mortality_table = 1/np.sum(dr*ps)
tdr_gompertz = 1/TLIA(65,95, np.log(1+r), 90, 10)

print(f'Mortality table: {tdr_mortality_table*100:.04f}%')
> Mortality table: 7.9425%

print(f'Gompertz Mortality Law: {tdr_gompertz*100:.04f}%')
> Gompertz Mortality Law: 7.6101%
```

Listing 2.7: Python script that values an annuity based on the survival probabilities of a mortality table

So far, this example showed the importance of obtaining the correct parameters to value an annuity. However, another crucial point needs addressing. In a mortality table, q_{65} represents the fraction of 65-year-olds who will not survive through the year based on data from a specific year, such as 2020. However, using the q_x values from the same table to price a product can introduce biases.

It is reasonable to expect that life expectancies will improve over time, leading to higher survival probabilities in the future. For example, it would be incorrect to assume that the survival probability of a 65-year-old retiree in 2020 will remain the same when they turn 85 in 2040, as it is for an 85-year-old today. This improvement in survival probabilities needs to be accounted for in pricing models to avoid underestimating longevity risk, which is why improvement factors are included in mortality tables.

These improvement factors can be based on advanced techniques, but that is beyond the scope of this work. Instead, in this study some simple assumptions for the improvement in survival probabilities are made:

- Mortality rates for ages 65 to 75 will improve (decline) by 3% each year.
- Mortality rates for ages 75 to 85 will improve by 2% each year.
- Mortality rates for ages 85 to 95 will improve by 1% each year.

Figure 2.10 shows survival probabilities before and after applying the improvement factors.

After discussing the importance of selecting the correct parameters in the Gompertz Mortality Model and incorporating improvement factors to the discrete mortality table, the appropriate m and b parameters must be determined. The algorithm reported in Listing 2.8 linearizes the mortality rate q_x via a double log calculation and then regresses that number on age (see Milevsky 2020). The Gompertz parameters that approximate the 2020 Italian Mortality Table with the improvement factors are $m = 89.43$ and $b = 8.69$. As shown in Figure 2.11, the survival curve with these parameters closely resembles the actual mortality table.

```
# Define X (ages) and qx (probabilities)
X = np.arange(65, 95)

# Compute y
```

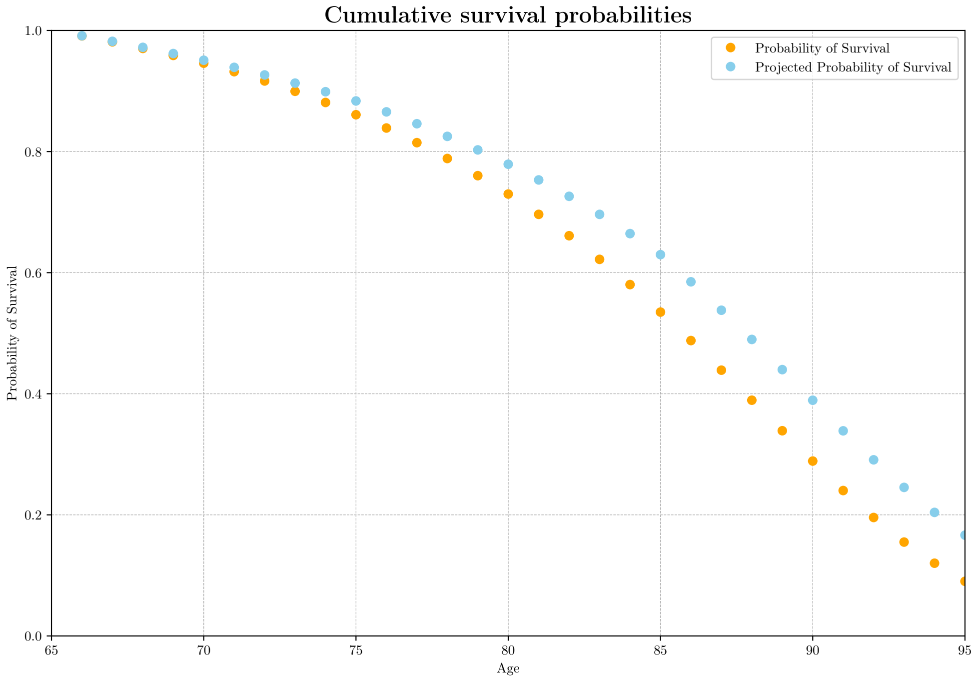

Figure 2.10: Cumulative survival probabilities, actual values versus projected values

```
y = np.log(np.log(1 / (1 - qx_u_2020_proj)))

# Fit a linear model
slope, intercept, r_value, p_value, std_err = stats.linregress(X, y)

# Coefficients from the linear model
g = slope
h = intercept

# Calculate m and b
m = np.log(g) / g - h / g
b = 1 / g
```

Listing 2.8: Python algorithms that fits a mortality table to the Gompertz Mortality Law parameters

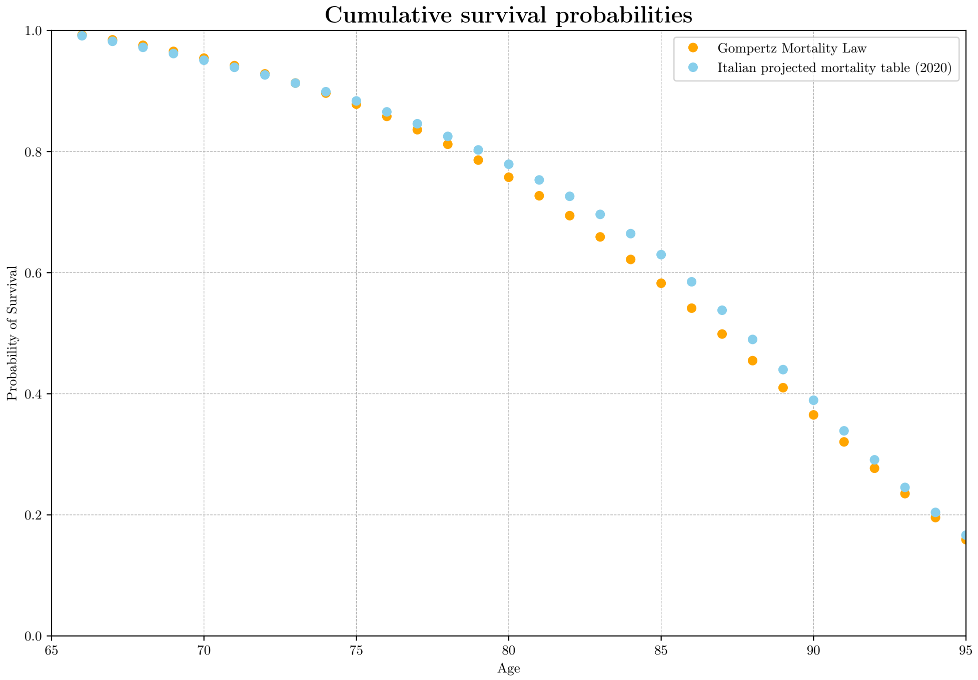

Figure 2.11: Cumulative survival probabilities, Fitted Gompertz Mortality Law versus projected mortality table

Bootstraping the returns

In the 'toy example' of the Natural Tontine developed in the previous chapter, it was assumed that returns were LogNormally distributed and time-independent, with a mean (ν) of 4% and a standard deviation (σ) of 3%. In this chapter, these simplistic assumptions will be relaxed and instead simulations with monthly returns bootstrapped from the S&P 500 data spanning from 1871 to 2023 will be used. Specifically, a block bootstrap technique will be implemented to preserve the autocorrelation in the simulated return paths by resampling with replacement from the original data.

The first step is to assess the time dependence in the S&P 500 monthly returns by performing an autocorrelation analysis. The autocorrelation plot reported in Figure 2.12 reveals that the only statistically significant time lag is at $\rho = 1$, indicating that returns are correlated from one month to the next. This suggests that the appropriate block size for the resampling should be two months.

Next, a random date is selected from the dataset and the returns for that month and the following month are extracted, as determined by the autocorrelation analysis. This

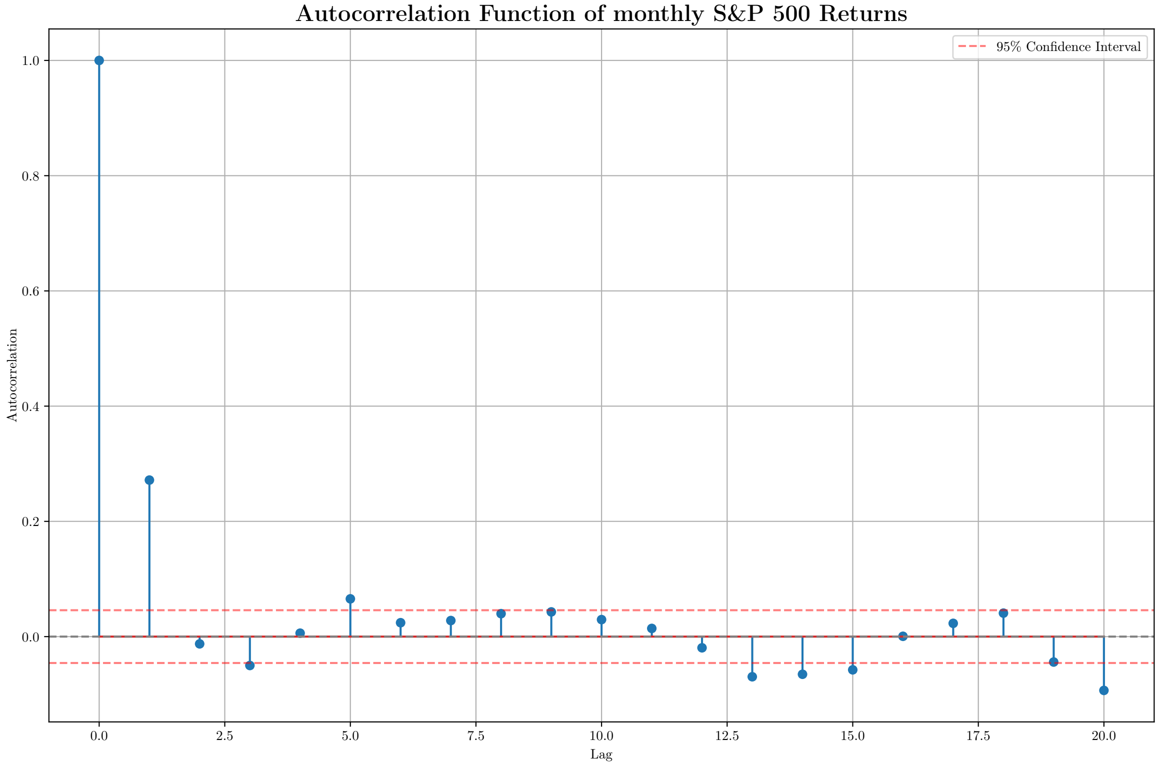

Figure 2.12: Autocorrelation plot for the monthly returns of the S&P500 from 1871 to 2023

process is repeated until enough returns are collected to cover the entire simulation period of interest. For instance, if the fund's horizon is 30 years, as assumed, 360 monthly returns are needed.

Finally, since the Natural Tontine Fund pays dividends annually, the monthly returns are converted to yearly returns, thus generating a single return path. This procedure is repeated `N` times to create all the necessary return simulations.

The 10,000 simulations resulted in yearly returns with a mean of 10.50% and a standard deviation of 17.53%. The returns are positively skewed (0.69), indicating that there are more extreme high values in the dataset than extreme lows. The kurtosis is 2.95, which is just below the value of a normal distribution. The Kolmogorov-Smirnov test is performed to check if the generated returns approximate a normal distribution. However, the test results in a p-value of 0, leading to the rejection of the null hypothesis that these returns are normally distributed.

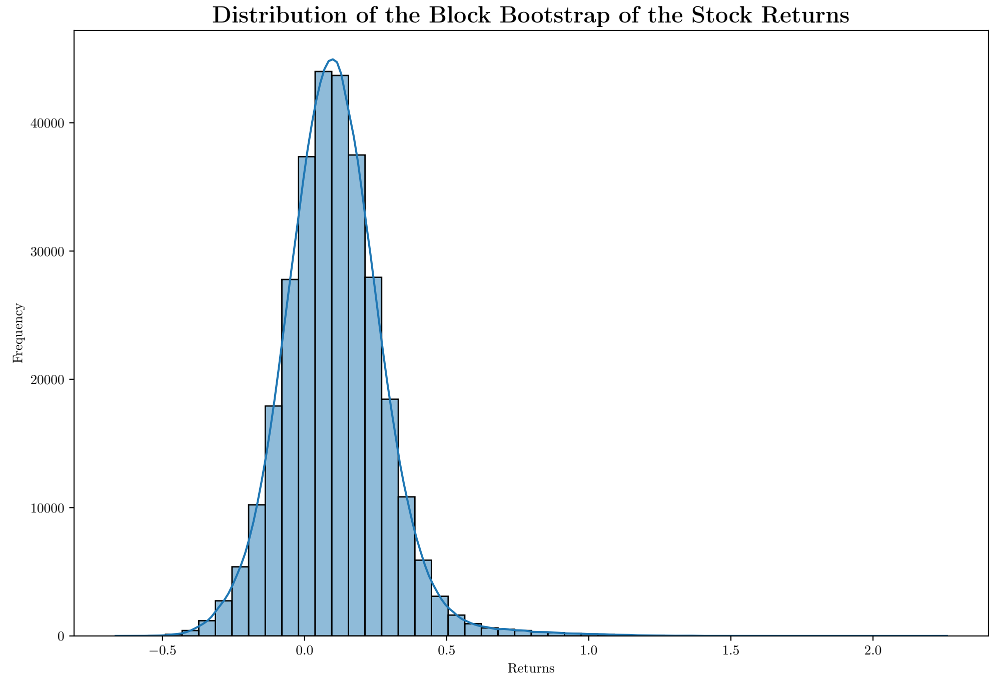

Figure 2.13: Returns' distribution of the Block Bootstrap simulation

The evolution of the Natural Tontine Fund with real data

As a final step, the improved mortality model with the simulated returns to understand the evolution of the tontine dividend and the fund value using real data.

In the 'toy example', returns were generated using a data-generating process with specific parameters. Specifically, it was assumed assumed that the returns were log-normally distributed with a mean value of EXR = 4%, which was equivalent to the Assumed Rate of Return (r). This equivalence between the mean value of the returns and r allowed to maintain a constant tontine dividend over time, as shown in Figure 2.5.

However, in the real-world example, the underlying data-generating process of the returns is not known. Consequently, it is not possible to determine the correct discount rate to guarantee the stability of the dividends a priori. In this chapter, the mean return of the simulated returns is used as the discount rate (r). While this approach is convenient, its correctness is debatable due to its susceptibility to look-ahead bias.

The initial tontine dividend is set at \$11,781.48, and although Figure 2.15 suggests that the tontine dividend remains constant over time, running the regression described

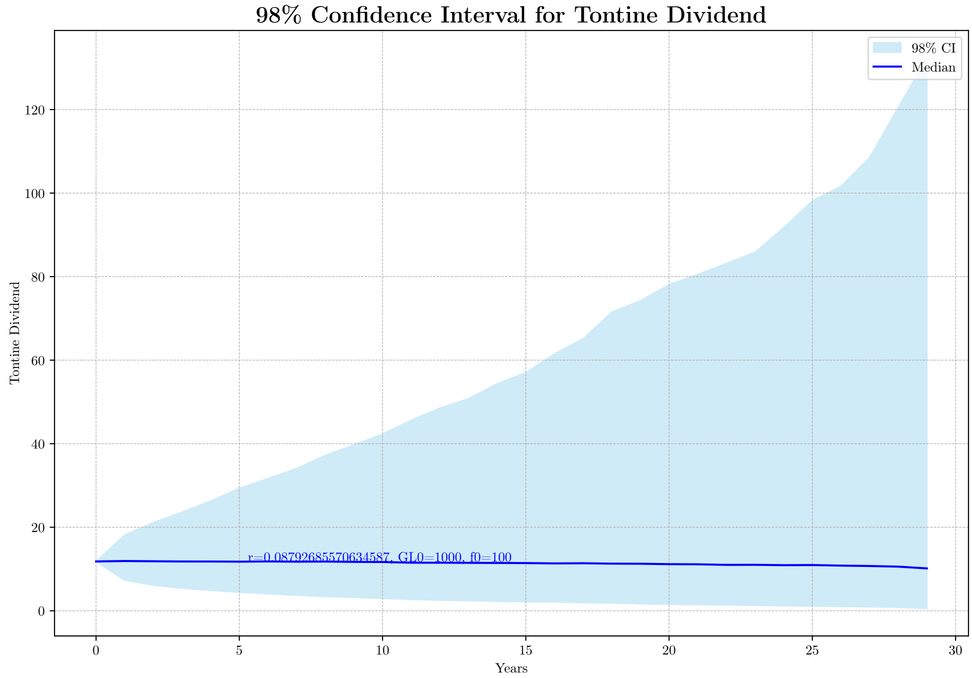

Figure 2.14: 98% confidence intervals for Tontine Dividend, with Gompertz Mortality Law with fitted parameters, Block Bootstraped S&P500 returns

in 2.6 reveals a slight downward trend (slope: -0.0362).

Comparing Figure 2.15 (the Natural Tontine Fund with real data) with Figure 2.5 (the 'toy example' of the Natural Tontine Fund), some interesting differences can be observed. In the real-data example, the tontine dividend is highly asymmetric, with significant upward potential. The mean dividend is $2,796.96 higher than the median, whereas, in the 'toy example', the difference was just $27. This disparity is likely due to the positively skewed nature of the real data, where positive returns are more probable than negative ones. In contrast, the returns in the 'toy example' were lognormally distributed with zero skewness.

For the same reason, the confidence interval for the Fund Value in Figure 2.15 does not remain within a tight and stable band, as it did in the 'toy example', but varies greatly from simulation to simulation.

Despite these differences, when simulating the evolution of the Fund Value and the tontine dividend with real data, the main features of the Natural Tontine Fund are still verified: the tontine dividend remains (fairly) stable, and the fund allows for perfect

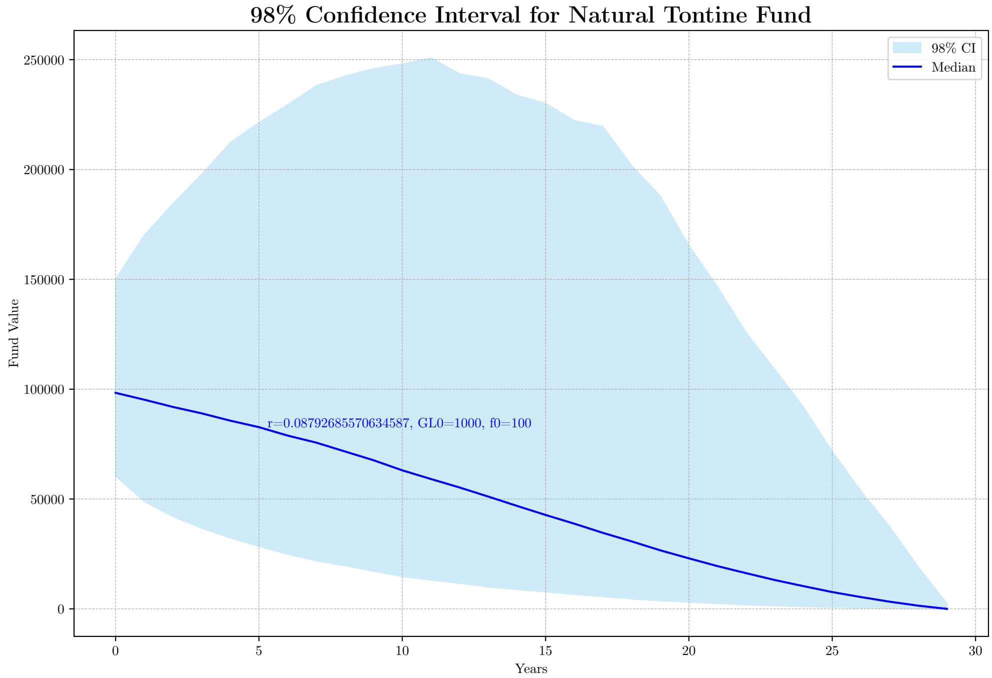

Figure 2.15: 98% confidence intervals for Natural Tontine Fund Value, with Gompertz Mortality Law with fitted parameters, Block Bootrstraped S&P500 returns

decumulation.

Chapter 3

The Natural Tontine Fund for Early Retirement

The purpose of this chapter is to evaluate whether a Natural Tontine Fund could be a desirable investment product for early retirees. Indeed, there is a clear parallel between withdrawal rates, introduced in 1.1 and Tontine Dividend Rates, discussed in 2.3.3, as the tontine dividend can be seen as the amount that is withdrawn each year from the investment portfolio.

The same data generated in the previous chapter for the investment returns and the number of survivors will be used, but in this scenario, it will be assumed that all retirees are 45 years old and the fund's horizon is 50 years.

3.1 Imposing a floor to the tontine dividend

Early retirees want to maintain constant dollar spending in real terms over time. Therefore, it is crucial for this purpose to move from the nominal dividends and fund values to real terms, to maintain stable purchasing power over time. This can be achieved by simply dividing the fund value at the end of the year by the observed rate of inflation during the same year. Furthermore, it is necessary to assume that the discount rate r is now equal to the difference between the nominal mean stock return and the mean inflation rate.

By implementing this update to the algorithm, it is possible to see that the fund will pay an initial dividend of \$7,965.89. The mechanism of the Natural Tontine still allows the tontine dividend to remain fairly constant in real terms (slope coefficient 0.0058) and to perfectly decumulate the wealth.

Even though the median dividend is projected to remain constant over the entire horizon of the fund, in 50% of the cases, retirees will receive a tontine dividend lower than the initial amount. This feature is undesirable for early retirees who need to receive and spend a specific sum each year to cover their needs and sustain their lifestyle. The issue lies in the structure of the Natural Tontine itself: it optimizes for the median case, while early retirees are concerned with being secure even in the worst-case scenario.

To overcome this limitation of the Natural Tontine Fund, it is possible to set a floor for the tontine dividend amount and assess the impact of this additional feature. This would ensure that fund members receive at least the amount they need to cover their needs. This can be achieved by adding a condition to the script: whenever the computed tontine dividend falls below a certain floor in dollar terms, the sponsor will pay the remaining members the minimum agreed amount.

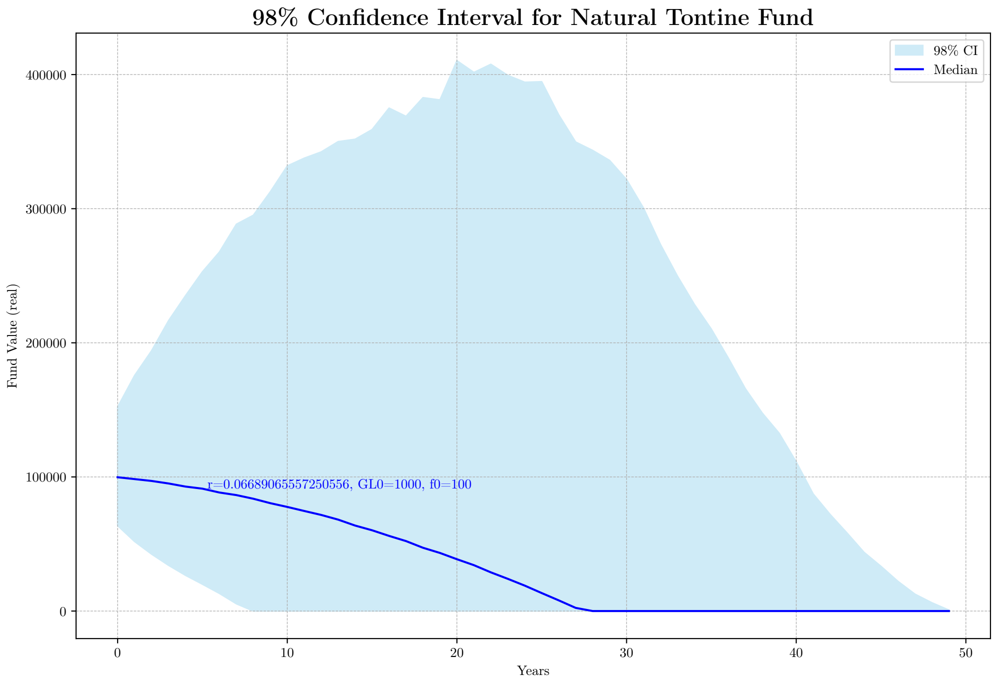

Figure 3.1: 98% confidence intervals for Tontine Fund value, imposing a floor to the tontine dividend

Figure 3.1 shows the consequences of adding such a floor to the tontine dividend.

With this feature, there is a chance that the fund will run out of money at some point. Specifically, if the floor is set equal to the initial dividend payout in dollar terms, there is a 72.8% chance that the fund will run out of money before the last year. The intuitive explanation is that a higher payout is being enforced than what the Natural Tontine structure calculated to maintain the fund's solvency.

At this point, it is not premature to conclude that it is not possible to build a Natural Tontine Fund where the sponsor also guarantees a specific dollar dividend each year. There is a clear trade-off: either variability in the tontine dividend is accepted, or there is a significant risk of running out of money before the last year. It is not possible to have both with the Natural Tontine structure.

However, the Natural Tontine structure can still be used as the backbone of the fund. To make it a desirable investment instrument for early retirees while adding a floor, strategies must be experimented with to mitigate the risk of failure, although this will modify the nature of the Natural Tontine itself.

3.2 Managing failures

Setting a Cap

The first strategy to limit the ruin probability when adding an artificial floor to the natural tontine is to impose a cap. This implementation would limit the range of payments that a retiree could receive in a year, thus creating a buffer for the lean years when the tontine dividend might fall below the floor.

To assess the effectiveness of this strategy, it can be first tried the most extreme case: imposing a cap equal to the floor and hence equal to the initial dividend, effectively forcing a fixed tontine dividend over time. By imposing this tight constraint, the risk of failure can be reduced to 50.76% in the simulation. However, this is still far from acceptable for retirees. This strategy completely ignores the Natural Tontine mechanism, which was designed to keep the fund solvent and decumulate wealth based on experienced returns and the number of survivors. As a result, in cases where the fund remained solvent, it was often shut down with a substantial amount of untouched wealth.

While setting a cap on the dividend seemed promising in theory, it did not significantly reduce the failure risk and introduced other dilemmas for the sponsor, such as what to do with the leftovers.

Alternative Asset Allocation

Previously, it was explained that setting an artificial floor for the tontine dividend introduces some risk of failure because it forces a higher payout than what the Natural Tontine structure calculated to maintain the fund solvent. Another way to frame this issue is that the fund is paying out too much during years of bad returns or when there are more survivors than expected. This issue reminds of the sequence of return risk introduced in Section 1.1, the risk that a series of low or negative returns can significantly impact the long-term sustainability of a withdrawal plan

A possible solution could be to use an alternative asset allocation, for example by increasing the exposure to bonds, to reduce the variability of returns and see if this could mitigate the failure rate.

For this purpose the returns of the GS10 bond, which represents the yield on U.S. Treasury securities with 10-year constant maturities from 1871-2024 [1] will be used, and different return paths using a block bootstrap resampling technique will be simulated. In each year and in each simulation, the bond return associated with the previously extracted stock return will be used to maintain the structural relationships between stocks and bonds. The extracted bond returns appear to be less risky than stocks, with a lower standard deviation of returns (5.34% versus 17.54% for stocks) and also uncorrelated with S&P 500 returns, as demonstrated by a correlation coefficient of -0.079.

By trying different combinations of stock and bond weights, the lowest failure rate is achieved with 10% stocks and 90% bonds, but the failure rate is still 61.3%. Additionally, changing the asset allocation also modifies the mean return of the portfolio, hence the discount rate r, and consequently, the Tontine Dividend Rate κ.

[1] Data available at Shiller Data.

Reducing the Tontine Dividend Rate κ

Another strategy to reduce the failure rate is to hold back some of the tontine dividend by subtracting a few basis points from κ. Intuitively, this would create a cushion for the bad years and reduce the failure rate without affecting the decumulation process. For example, reducing the Tontine Dividend Rate by 300 basis points results in an 18.71% risk of failure, significantly reducing it from the initial case, but it still seems above what retirees might consider acceptable.

Since the insolvency issue is closely related to the sequence of return risk, the Tontine Dividend Rate could be reduced just in the first 20 years when the negative impacts of the sequence of return risk are more pronounced, or only in the years of negative returns.On paper, these strategies seem promising, and failure rates are not expected to be significantly higher than in the case where there was a reduction each year. However, testing a 300 basis point reduction according to these rules results in failure rates of 32.49% and 33.05%, respectively.

Combining different strategies

None of the proposed strategies alone seems to be the solution. However, it is possible to experiment with different combinations of stock-bond allocation and Tontine Dividend Rate reduction to assess if this could reduce the failure risk to an acceptable level.

Figure 3.2 plots the different combinations of asset allocation and Tontine Dividend Rate reduction. On the x-axis, the ruin probability is plotted, and on the y-axis, the initial Tontine Dividend Rate that is also used to determine the floor for the tontine dividend. The solid blue line can be interpreted as an ‘efficient frontier’: all the points below it are inferior. Hence, for the same ruin probability, it is possible to obtain a higher initial Tontine Dividend Rate or conversely, for the same Tontine Dividend Rate, a lower ruin probability can be obtained. Therefore, it is rational to consider only the points on the frontier, and the decision between the different combinations depends on how much risk the retiree is willing to bear. If an acceptable level of risk can be considered a ruin probability of less than 5%, the optimal combination would be a 200 basis points reduction in the Tontine Dividend Rate each year and a 40% stock allocation.

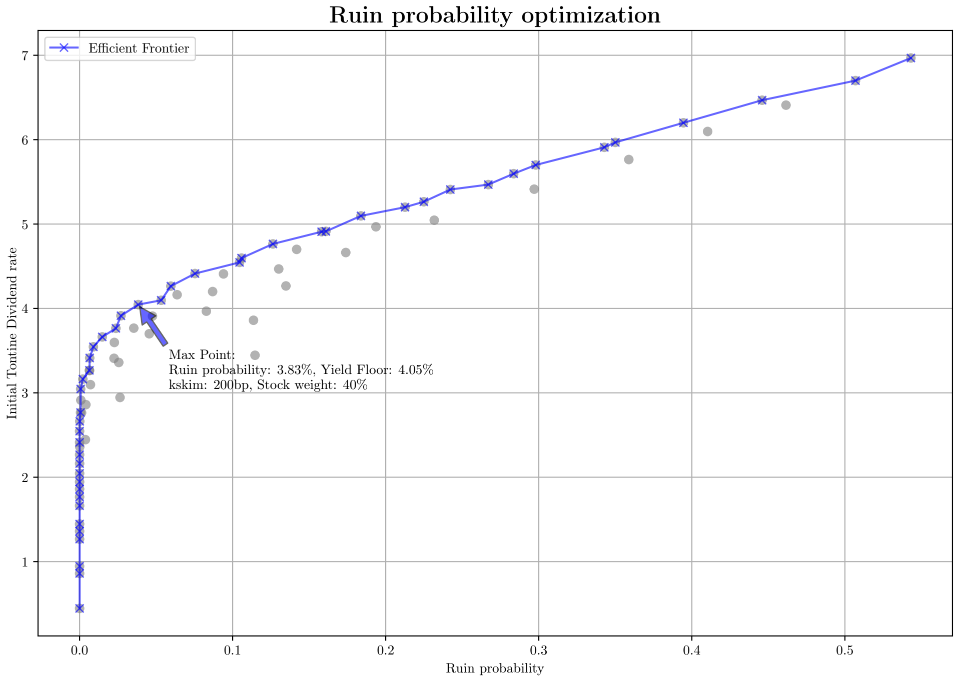

Figure 3.2: Different combinations of stock-bond weights and Tontine Dividend Rate reduction

Figure 3.3 shows the behaviour of this optimized fund. This combination allows for an initial Tontine Dividend Rate of 4.05%, therefore, assuming an initial investment of \$100,000, the retiree will receive exactly \$4,051.39 in the first year and at least this amount in all subsequent years, as this is the imposed floor. However, in the median case, the retiree would receive \$7,845.95 each year, with payments potentially growing larger in more positive scenarios. This is due to the Natural Tontine structure's ability to pay larger dividends in years with better returns or fewer survivors than expected, allowing for perfect decumulation of the fund. On the other hand, as previously noted, the sponsor cannot guarantee payments to annuitants because imposing a floor introduces the risk of insolvency, which in this simulations is 3.31%.

Early retirees set their desired annual spending and derive the necessary capital by multiplying the initial withdrawal by a factor, which is the reciprocal of the initial withdrawal rate. For a Natural Tontine Fund with an initial 4.05% Tontine Dividend Rate, the multiplier is $\frac{1}{4.05\%} = 24.69$. Therefore, a retiree wanting to receive at least \$40,000 annually would need to invest $\$40,000 \times 24.69 = \$987,600$ in the Natural

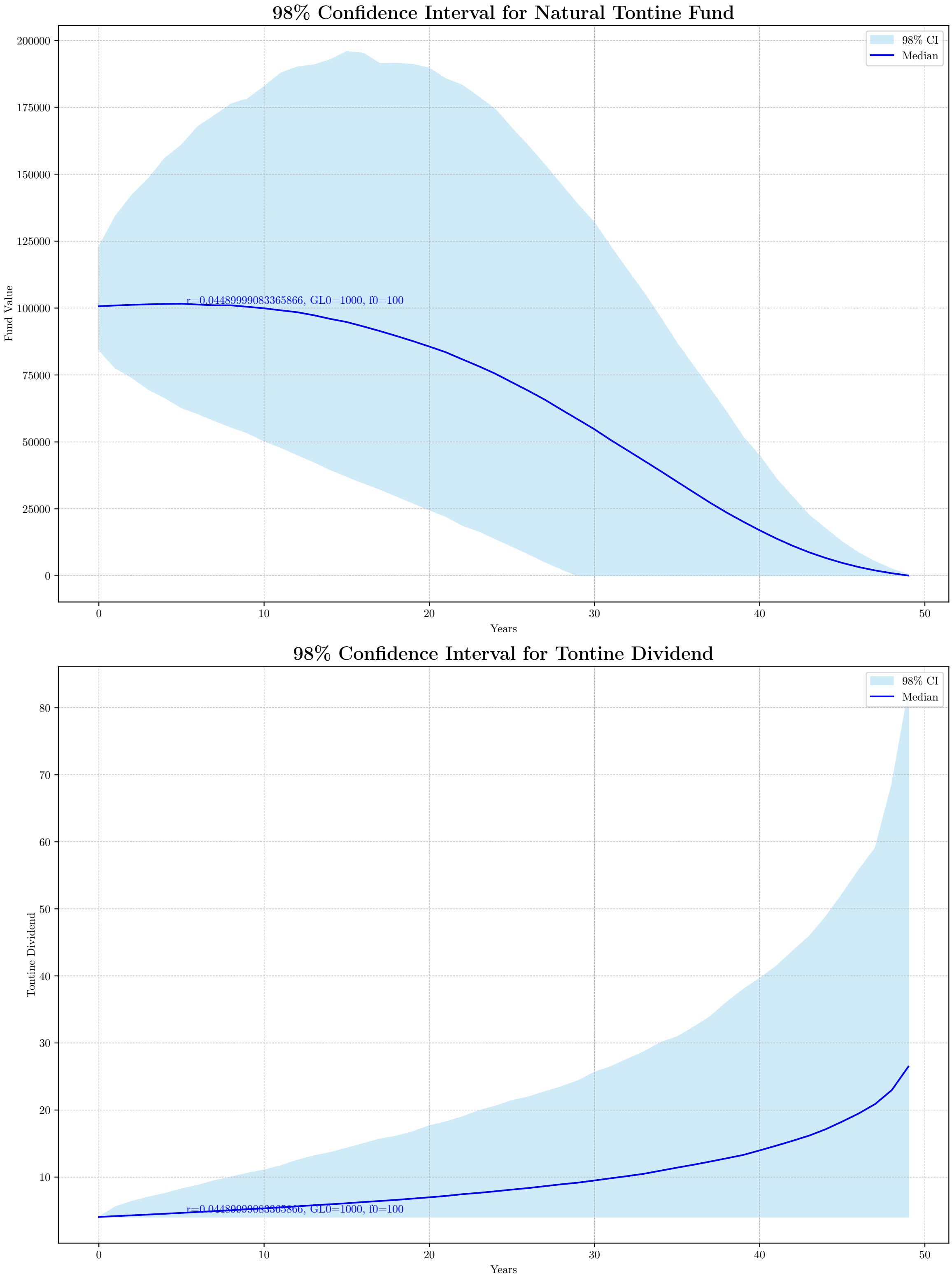

Figure 3.3: 98% confidence intervals for Tontine Fund Value and tontine dividend with optimized parameters

Tontine Fund.

3.3 Is this fund desirable for early retirees?

To evaluate whether the Natural Tontine Fund is an attractive investment product for early retirees, this must be compared to what a single retiree could achieve by managing and withdrawing funds independently. A single retiree aiming to withdraw an initial 4.05% and maintaining the same dollar amount in real terms for 50 years would face a 14.88% risk of depleting their entire wealth (assuming a 60% stock and 40% bond allocation), a risk that most retirees are unwilling to bear. Single retirees can reduce this risk to 4.42% by lowering the withdrawal rate to 3%, but this requires an initial investment 35% larger than the investment in the Natural Tontine Fund [2].

The Natural Tontine Fund ensures a minimum annual payment of $4,051.39 (assuming a $100,000 initial investment), with a median tontine dividend of $7,845.95. Conversely, the single retiree consistently withdraws the same amount in real terms each year. Thus, in the worst-case scenario, both groups receive the same amount annually, but in the median case, fund participants receive 93.66% more thanks to the Natural Tontine mechanism and pooled longevity risk.

In Section 1.3 it was argued that retirees using a constant withdrawal strategy in real terms overpay for the potential of investment gains that they do not need to meet their retirement income goals as they may accumulate wasted surpluses when their investments outperform (Scott, Sharpe, and Watson 2008). Indeed, in the simulation single retirees in the median case have $197,282.51 per $100,000 in their investment portfolio after 50 years, almost twice as the principal in real terms. On the other hand, the median final value of the Natural Tontine Fund is $83.80 per $100,000 initially invested. Hence, while single retirees end up wasting a large part of their savings on unnecessary surpluses, a Natural Tontine fund allows for perfect decumulation of the wealth.

Considering the median tontine dividend is almost double the floor, low-risk-averse investors could benefit from some flexibility. Allowing a floor that is 80% of the initial tontine dividend and reducing the Tontine Dividend Rate by 120 basis points results in an initial Tontine Dividend Rate of 4.85%, a median tontine dividend of $7,240.11

[2] Specifically, to finance $40,000 annual spending, single retirees would need $\$40,000 \times \frac{1}{3\%} = \$1,333,333.33$, compared to $987,600 required for investors in the Natural Tontine Fund, a 35% increase.

(assuming a \$100,000 initial investment), and a 3.61% risk of failure. Thus, by allowing some flexibility, participants in a Natural Tontine Fund can achieve a similar risk of failure with a lower initial investment[3].

Sponsors of a Natural Tontine Fund cannot promise participants a minimum dividend payment every year, as adding this constraint introduces a risk of insolvency. However, it is possible to implement strategies that allow for a small risk of failure in the fund even when a floor on dividend payments is imposed. Consequently, the payment profile is no longer flat, violating one of the main features of the Natural Tontine. Despite this, the modified Natural Tontine Fund structure could be a desirable investment product for early retirees. Compared to an early retiree managing his wealth independently, the fund, for comparable levels of risk, allows for a smaller initial investment, larger median payouts, and perfect decumulation of the funds.

[3] Specifically, an investor wanting to receive at least \$40,000 annually would need $\$40,000 \times \frac{1}{4.85\%} = \$824,742$, a 16.49% reduction compared to the case where the floor equals the initial payment.

Conclusions

The purpose of this thesis was twofold. First, to analyze the safety and optimality of spending plans based on constant real withdrawals, such as the one proposed by the '4% rule'. In particular, the analysis assessed the risks of extending the retirement period beyond 30 years and whether the '4% rule' overstates safe withdrawal rates. Secondly, this thesis evaluated whether a Natural Tontine Fund can be structured as a suitable investment tool for early retirees.

The research finds that extending the retirement horizon beyond 30 years reduces the chances of success for all withdrawal rates, especially if a large portion of the portfolio is allocated to bonds. Furthermore, it was found that the '4% rule' overestimates the safe withdrawal rate by about 75 basis points due to the modeling assumptions of the most influential studies on safe withdrawal rates. However, it was also found that a Natural Tontine Fund could be structured to provide retirees with a lifetime income stream with less risk of failure, higher median payments, and perfect decumulation of wealth compared to a constant-amount spending plan.

While this thesis provides a foundational exploration of the Natural Tontine Fund concept, one particular assumption underlying the analysis warrants further investigation to fully understand its real-world applicability. In the 'toy example' in Section 2.3, returns were generated assuming a lognormal distribution with a mean value of 4%, which was equivalent to the Assumed Rate of Return (r). This equivalence between the mean return and the discount rate allows to maintain a constant tontine dividend over time. However, in the real world, the underlying data-generating process of returns is not known, and consequently, it is not possible to determine the correct discount rate to guarantee the stability of the dividends a priori. In Section 2.3.5, the mean of the simulated returns was used as the discount rate (r), but this approach suffers from look-ahead bias. Using the mean return of the last 150 years of financial history may prove to be a flawed approach to derive the Tontine Dividend Rate that will maintain a constant tontine dividend in future years. More advanced techniques to derive forward-looking discount rates are necessary; however, this goes beyond the scope of this thesis.

This thesis aimed to discuss how progress in the asset management and insurance industries could help early retirees achieve their goals. However, to bridge the gap between theory and practice, it is also important to reflect on some of the implications and challenges of launching a Natural Tontine Fund on the market.

In Section 3.1, a floor on the tontine dividend was imposed so that fund members receive an amount each year that is greater than or equal to the initial payment in real terms. This floor forced the dividend to be higher than what was computed by the Natural Tontine structure calculated to maintain the fund's solvency. Therefore, when introducing this floor, there is a non-trivial chance of the fund's failure, and the sponsor of the fund cannot guarantee payments to its participants. The risk of insolvency for the fund will most likely impact its credit rating, and therefore, before launching this venture on the market, it would be extremely important to consider how this would influence participation in the fund.

Tontines are investment products that allow participants to speculate on their life expectancy, introducing two issues: adverse selection and moral hazard. Tontines suffer from adverse selection as they tend to be purchased by investors with longer life expectancies than the average person. Imagine a scenario in which a group of individuals could choose between an annuity and a tontine—two products that guarantee a lifetime stream of income. If an individual believes that they will live longer than the other members of the group, they will likely choose to invest in the tontine as they expect to receive larger mortality credits. On the other hand, a person who believes they have a life expectancy lower than the other members of the group will probably pick the annuity. This would lead to a situation in which people with above-average life expectancies buy the tontine, while others opt for the annuity. Regarding moral hazard, in a scenario where only three participants remain in the fund, individuals might be incentivized to engage in unethical behaviours, such as trying to eliminate the other survivors to obtain larger coupons.

If individuals do not know the identities of the other potential investors, the issue of adverse selection is drastically reduced, as no one can make an investment decision based on the health status of others. Moreover, if participants do not know who the other members of the fund are, the problem of moral hazard is also eliminated. Anonymity seems to be a solution to these issues, and institutions considering launching a Natural Tontine Fund should explore potential methods to anonymize their participants.

The evidence provided in this thesis suggests that an ancient tool, the tontine, could solve the modern problem of decumulation. The Natural Tontine Fund aims to provide early retirees with an alternative investment approach to the simple constant withdrawal spending strategies. However, significant efforts by institutions will likely be necessary to motivate insurance companies to launch this product on the market and to incentivize people to consider the Natural Tontine Fund as an efficient solution to decumulating their wealth during retirement.

Bibliography

Anarkulova, Aizhan et al. (2022). *The Safe Withdrawal Rate: Evidence from a Broad Sample of Developed Markets. SSRN Electronic Journal.*

Bengen, William P (1994). *Determining Withdrawal Rates Using Historical Data.*

Choi, James J. (2022). *Popular Personal Financial Advice versus the Professors. Journal of Economic Perspectives.*

Cooley, Philip L, Hubbard, Carl M, and Walz, Daniel T (1998). *Retirement Savings: Choosing a Withdrawal Rate That Is Sustainable.*

Davidoff, Thomas, Brown, Jeffrey, and Diamond, Peter (2003). *Annuities and Individual Welfare.* National Bureau of Economic Research.

Donnelly, Catherine, Guillén, Montserrat, and Nielsen, Jens Perch (2014). *Bringing cost transparency to the life annuity market. Insurance: Mathematics and Economics.*

Hakansson, Nils H (1970). *Optimal Investment and Consumption Strategies Under Risk for a Class of Utility Functions. Econometrica.*

Jorion, Philippe and Goetzmann, William N. (1999). *Global Stock Markets in the Twentieth Century. The Journal of Finance.*

Li, Hong et al. (July 2021). *Gompertz law revisited: Forecasting mortality with a multi-factor exponential model. Insurance: Mathematics and Economics.*

McKeever, Kent (2010). *A Short History of Tontines.*

Mehra, Rajnish and Prescott, Edward C. (1985). *The equity premium: A puzzle. Journal of Monetary Economics.*

Merton, Robert C. (1969). *Lifetime Portfolio Selection under Uncertainty: The Continuous-Time Case. The Review of Economics and Statistics.*

Milevsky, Moshe A. (2006). *The Calculus of Retirement Income: Financial Models for Pension Annuities and Life Insurance.* 1st ed.

— (2020). *Retirement Income Recipes in R: From Ruin Probabilities to Intelligent Drawdowns.*

Milevsky, Moshe A. (2022). *How to Build a Modern Tontine: Algorithms, Scripts and Tips.* Future of Business and Finance.

Milevsky, Moshe A. and Salisbury, Thomas S. (2015). *Optimal retirement income tontines. Insurance: Mathematics and Economics.*

Pfau, Wade D. (2017). *How Much Can I Spend in Retirement?: A Guide to Investment-Based Retirement Income Strategies.*

Piggott, John, Valdez, Emiliano A., and Detzel, Bettina (2005). *The Simple Analytics of a Pooled Annuity Fund. Journal of Risk and Insurance.*

Poterba, James, Venti, Steven, and Wise, David (2011). *The Composition and Drawdown of Wealth in Retirement.*

Promislow, S. David (2014). *Fundamentals of actuarial mathematics.* Third edition.

Pye, Gordon B. (2000). *Sustainable Investment Withdrawals. The Journal of Portfolio Management.*

Sabin, Michael J. (2010). *Fair Tontine Annuity. SSRN Electronic Journal.*

Samuelson, Paul A. (1969). *Lifetime Portfolio Selection By Dynamic Stochastic Programming. The Review of Economics and Statistics.*

Scott, Jason S, Sharpe, William F, and Watson, John G (2008). *The 4% Rule—At What Price?*

Stamos, Michael Z. (2008). *Optimal consumption and portfolio choice for pooled annuity funds. Insurance: Mathematics and Economics.*