Individual Tontine Accounts

Richard K. Fullmer[*] and Michael J. Sabin[†]

July 17, 2018

Abstract

An individual tontine account (ITA) is an investment product similar to a conventional brokerage account, but with the added feature of mortality pooling through participation in an open-ended fair tontine. The ITA exploits the property that participants in a fair tontine need not be confined to a common investment portfolio or to a common payout method. Instead, participants are allowed to select and trade investments as they wish and to choose from a variety of payout methods, with each participant's results being largely unaffected by the investment and payout choices of others. We envision the ITA as being complementary to an individual retirement account (IRA), allowing retirees to derive extra income from savings without taking on additional investment risk and to obtain lifetime income at a lower cost than with comparable insurance products. The mortality-pooling features of ITAs compare favorably to those of insurance products. The cost per dollar of income is lower. The opportunity for individual choice is increased. Fees are transparent rather than opaque. Accounting is transparent and conveyed simply on account statements. The downside of ITAs is that income from mortality pooling is not guaranteed, and a participant might experience less income than hoped for if other participants live longer than expected.

ITAs represent a new arrow in the quiver for addressing global retirement needs and may address the "annuity puzzle" by giving retirees a more transparent, lower-cost alternative to insurance-based products.

_____

[*] Contact author. Founder, Nuova Longevità Research, Baltimore, MD, USA.

[†] Independent consultant, Sunnyvale, CA, USA.

Introduction

The phrase "the king is dead... long live the king!" is said to have originated in France upon the death of King Louis XII in the year 1515 to convey the notion that the country is never without a monarch because the transfer of sovereignty from one king to another occurs instantaneously the moment a king dies.[1] This notion carries over to tontines, in which there is a clear and instantaneous transfer of ownership rights in shares of the tontine scheme from dying members to surviving members.

A tontine is a financial arrangement with 17th century origins in which the parties involved agree to share the proceeds of a collective investment pool in a pre-described way. Notably, the arrangement mandates that the ownership of a member's share in the tontine is forfeited at death, with the proceeds apportioned among the surviving members. Members die, but the tontine pool lives on – perhaps forever if designed to run in perpetuity. Similar to the concept of annuitization, tontines offer investors a way to pool mortality risk, but directly among themselves as opposed to via an insurance company.

Tontines largely fell out of use early in the 20th century after self-dealing and fraud on the part of tontine providers prompted regulators to virtually ban them. Yet such problems can be addressed by proper design and oversight, and this recognition has led to renewed interest in tontines as a product.

Several authors have made contributions to specific tontine designs. Our focus here is not so much on any particular design, however. Rather, our intent is to explore the *boundaries* of tontine design – or more specifically, the boundaries of *fair tontine* design.

From their origins as lottery-like propositions to more recent innovations that engineer annuity-like payout streams, tontines have mostly been thought of as highly specialized offerings in which a prospectus would describe the specific terms of both the investment and the payout scheme. Most tontine designs in the literature (including tontine-like concepts such as pooled annuity funds and group self-annuitization schemes) have generally shared two common assumptions – that the scheme in question is associated with: 1) a particular investment fund in which all members invest, and 2) a specific payout method

---

[1] "*Le roi est mort!... Vive le roi!*" Variations of this phrase existed prior to this, dating as far back as 1422. A history of the phrase can be found in Kantorowicz (1957).

that applies to all.[2] In many ways, these tontines appear similar to managed payout mutual funds that have the added twist of mortality pooling.

While we acknowledge that such specialized solutions have appeal in specialized situations (tontine pensions, for example), there is no reason per se that tontines must be packaged in such a confined manner. As theorized in Sabin (2010) and Donnelly et al. (2014), the fair-tontine principle allows members to individually select their own investment portfolios and payout methods. We seek to advance this idea in practical terms by applying the fair-tontine principle to the concept of brokerage accounts that are perpetually open to new members, in which individuals are freely allowed to invest in virtually any investment of their choosing, to trade in their accounts as they wish, and to choose from a wide array of payout methods. Such tontine arrangements, which we call *individual tontine accounts*, or ITAs, can be fair to all members regardless of the investment and payout choices of the other members. Moreover, an individual member's results are largely unaffected by the investment choices of the other members. Among other potential uses, we envision that ITAs could serve as a special type of mortality-pooled Individual Retirement Accounts (IRAs), allowing retirees to derive extra income from their savings without taking on additional investment risk, and giving them the option to secure annuity-like lifetime income from their savings.

Fair Tontines

We define a "fair" tontine as one in which the expected value of the gain or loss that a member experiences as a result of mortality pooling is zero for each member. That is, each member receives a "fair" bet in the probabilistic sense.[3]

The total return of a tontine investment is a function of two components: 1) the amount of investment income that is earned and capital gains/losses that occur, and 2) the amount of mortality gains/losses that are credited. It is the second component that makes the tontine return different from that of a regular investment. We use the more consumer-

---

[2] See, for example: Donnelly (2015), Donnelly et al. (2013), Forman and Sabin (2014), Forman and Sabin (2016), Gründel and Wandt (2017), Milevsky and Salisbury (2015), Milevsky and Salisbury (2016), Piggott et al. (2005), and Stamos (2008).

[3] Milevsky and Salisbury (2016) make a technical distinction between the terms "fair" and "equitable." The tontine pool discussed in their paper has a finite life. As a result, there will unavoidably be a little money left over when the last member dies, which inures to the benefit of the tontine provider or perhaps some other party, but not to the benefit of the tontine members. Thus, the expected value of a member's mortality gains and losses is slightly negative and therefore not exactly "fair" as we (or they) define it. Nevertheless, the tontine is "equitable" in that no member or group of members has an advantage over any other. On the other hand, because our ITA pool is perpetual, no money is ever leftover. For this reason, the ITA is not only "equitable," it is also "fair."

friendly term "tontine gain" to refer to this second component – the tontine gain reflects the extra amount credited to a member's account due specifically to having invested in a tontine. Members suffer complete tontine losses when their account balances are forfeited upon death, but surviving members enjoy tontine gains when the proceeds of these forfeitures are shared among them.

For a tontine to be fair, the expected value of tontine gains and losses must be zero for each member. This equality, which may be described as a "fairness constraint," leads to another principle of fair tontines: The expected value of the total return of each member's investment is the same as it would be if the investment had instead been made outside of the tontine (in the latter case, this includes transfer of the balance at death to the investor's heirs). While the expected values are the same, investing in a tontine changes the conditional distribution of outcomes – those who live long lives do better by participating in the tontine, while those who die early do worse.

Satisfying the fairness constraint requires that the forfeited account balances of deceased members be transferred to the surviving members in an actuarially neutral (unbiased) way, taking into account each member's relative stake in the pool and probability of dying. As a simplified example, suppose an individual "Alice" wishes to make an investment over a given period. Assume that Alice's probability of dying during the period is q. Assume further that the value of Alice's investment at the end of the period is s. If Alice makes her investment within a regular account, her balance at the end of the period would simply be s, regardless of whether she lives or dies. But if Alice makes her investment within a tontine, her balance at the end of the period would be a random amount S that depends on who died during the period. If Alice dies, she forfeits her investment, and her balance at the end of the period is $S = 0$. But if she survives, her balance S depends on which of the other members may have died (if any). For the tontine to be fair, the expected value of her balance must equal s, the value it would be if she had invested outside the tontine. Thus:

$$s = q \times 0 + (1 - q) \times \mathrm{E}[S \mid \text{Alice survived}],$$

where $\mathrm{E}[S \mid \text{Alice survived}]$ is the expected value of Alice's balance S conditioned on her survival. Solving for it gives:

$$\mathrm{E}[S \mid \text{Alice survived}] = \frac{s}{1 - q} = s(1 + r),$$

where $r = q/(1 - q)$ is Alice's *nominal tontine yield*. Alice's *nominal tontine gain* is $s(1 + r) - s = rs$, meaning her nominal tontine yield r times her balance s. The actual tontine gain that Alice receives is a random amount that depends on who died, but the expected value of that gain, conditioned on her survival, is rs. Since q is bounded by the range $0 < q < 1$, r is a positive value. It is easy to see that the nominal yield r can be

significant. For example, if Alice's probability of dying over the period is 40%, her nominal yield is $r = 0.4/(1 - 0.4) = 66.67\%$. In other words, if Alice survives, her expected balance would be two-thirds higher by investing in the tontine than it would be by investing in a regular account.

In an ITA, tontine gains are computed and distributed periodically at some convenient interval, such as monthly, quarterly, or yearly. The fairness constraint is applied during each period and reflects the members' probabilities of dying during that period. Since mortality rates increase with age, each member's nominal tontine yield similarly increases with age, and thus the advantage of investing in the ITA compared to a regular account grows with age. Since the fairness constraint is applied to every period, the ITA can operate in an open-ended fashion, with new members allowed to join at the start of any period.

Forfeiture Allocation

The key to fair tontines lies in the allocation of forfeited balances, and several methods have been proposed in the literature. Sabin (2010) and Donnelly et al. (2014) each present methods that exactly meet the fairness constraint for every member. A key feature of these methods is that deaths are processed one at a time as they occur and become known. The transfer of power discussed in our opening quote ("the king is dead… long live the king!") makes a fit analogy in that it is, in theory, most accurate (as regards fairness) to transfer assets instantaneously as a death occurs. On the other hand, we are concerned that such a method may be difficult to explain to prospective customers because the calculations involved in the instantaneous transfer are complex. Because we are genuinely interested in the commercial success of tontines, we take such matters seriously.

For this reason, we elect a less complicated method of forfeiture distribution that we feel is much easier for providers to explain and investors to understand. This method, referred to as the "nominal-gain method" in Sabin and Forman (2016), is appealing in that all forfeiture allocations are easily decomposed into two transparent components: 1) a nominal tontine yield for each member, which is easily obtained from the tontine's publicly disclosed mortality table, and 2) a common adjustment factor that accounts for the difference between the amount of forfeitures actually experienced by the pool and the amount that was expected per the mortality table.

As discussed above, the first component – a member's nominal tontine yield – is $r = q/(1 - q)$, where q is the member's probability of dying during the period. The value of q depends on the member's age and gender and is specified in the tontine's publicly disclosed mortality table. In dollar terms, each member's nominal tontine gain is rs, where s is the member's balance at the end of the period.

We refer to the second component – the common adjustment factor – as the *group gain*, G. It is found by dividing the sum of all balances forfeited by those who died during the period by the sum of the nominal tontine gains of those who have survived. That is, if we let A denote the set of members who died during the period (the decedents) and A^c (i.e., the complement of A) the set of members who did not die (the survivors), then the group gain is:

$$G = \frac{\sum_{j \in A} s_j}{\sum_{j \in A^c} r_j s_j}. \tag{1}$$

In other words, the group gain G is the ratio of the total amount forfeited by the decedents to the total of the nominal gains of the survivors. This ratio is applied to the nominal gain of each survivor such that the actual tontine yield for surviving member j becomes Gr_j, and the actual tontine gain in dollars becomes $Gr_j s_j$. If the total amount forfeited by decedents is greater than the total nominal gains of survivors, G will be greater than 1, and the survivors will all receive more than the nominal gains they anticipated. But if the total amount forfeited by the decedents is less than the total of the nominal gains of survivors, G will be less than 1 and the survivors will receive less than the nominal gains they anticipated.

This nominal-gain method of forfeiture allocation is simple, transparent, and readily perceived as being fair to all members. Each member can look up her nominal tontine yield r in a table published in advance by the provider, and all members share the same group gain G which is published by the provider at the end of the period. Each member anticipates receiving a tontine gain which is close to her nominal amount rs. If her actual tontine gain is less than anticipated, say by 2% – meaning $G = 0.98$ – then she knows that every other surviving member's actual gain is also 2% less than anticipated. Or if she learns that another member has received an actual gain that is higher than anticipated, say by 2% – meaning $G = 1.02$ – then she knows that her actual gain is also 2% higher than anticipated. Thus, the perception is that all members are being treated fairly, since their actual tontine gains move in lockstep.

While the nominal-gain method is readily perceived as being fair to all members, it is not strictly fair in an actuarial sense, at least not exactly. When one looks carefully at the expected value of the tontine gain or loss that a member receives, one finds that some members have an expected value that is slightly positive, while other members have an expected value that is slightly negative. That is, there is a bias that favors some members over others. An analysis of this bias is provided in Sabin and Forman (2016). The analysis is complicated, but for our purposes the bottom line is that the bias is negligible in a tontine pool of the size and type we consider for an ITA. We can, for practical purposes, use the nominal-gain method and regard it as fair. We choose the nominal-gain method for its

advantages of simplicity, transparency, and perceived fairness, and in doing so we accept its theoretical imperfection.

Expanding the Boundaries of Fair Tontine Design

An important, and perhaps surprising, feature of fair tontines is that a surviving member's expected tontine gain depends only on her own balance s and her own likelihood of dying q. This follows from the simple observation that her expected tontine gain is $rs = sq/(1-q)$. It does not matter who else is participating: young or old, male or female, individual or couple, rich or poor. It also does not matter *how* others are participating: the size of their accounts, the investments they select, their trading activity, or the payout option they elect. All that matters is her own balance and her own probability of death.

Of course, the actual tontine gain will vary from its expected value since it is very unlikely that members will die in exactly the proportions and times given by the mortality assumptions. Actual tontine gains will exhibit randomness. For this reason, tontines require a large number of members to help ensure that actual mortality will be close to assumed mortality, which in turn will help ensure that actual tontine gains are close to expected tontine gains. This reflects the law of large numbers, which predicts that when the number of members of the tontine pool is sufficiently large, the actual payouts to surviving members will be close to the expected values.[4] This is one of the attractive features of ITAs – since they place relatively few restrictions on membership, the potential pool can potentially be very large indeed.

Individual Tontine Accounts

We envision ITAs as individually-owned investment brokerage accounts offered through a common tontine pool. ITAs could be opened as individual retirement accounts or standard taxable accounts. Like traditional brokerage offerings, the ITA offering is perpetually open-ended such that new members may open new accounts at any time, and current members may similarly make additional investments at any time. Thus, the individuals who make up the pool will change over time, and eventually newer generations will completely replace older generations. Members may be individuals or couples – thus, ownership may be either individual or joint.[5]

---

[4] Of course, mortality assumptions are uncertain. An individual's idiosyncratic mortality risk can be effectively diversified away by pooling. However, systematic mortality risk, the potential for aggregate mortality rates to differ from those assumed, cannot. We address this in a later section.

[5] When ownership is joint, the probability of dying, q, represents the probability that both owners will have died. This is readily determined by standard actuarial techniques.

Feature Summary

The basic features and value proposition of an ITA can be summarized as follows. An account holder can invest in whatever she wishes within some large collection of permitted liquid investments. If an account holder dies (or if the second/last account holder dies, for an account held jointly), her account balance is forfeited. But in each period that she survives, she receives a share of the balances forfeited by those who died during that period. An account holder can only withdraw according to a predetermined payout schedule. This restriction is necessary to prevent a type of "reverse moral hazard" in which members who become seriously ill could otherwise elect to cash in their account just before they die. The account holder chooses the payout schedule from some variety of options at the time she opens the ITA. Note that depending on the type of payout schedule selected, it may be possible for her to live long enough that she never suffers a forfeiture event. This will be true if she selects a fixed-term payout schedule and outlives the payout term. Conversely, if she selects lifetime payouts, she can never avoid eventual forfeiture – but of course, this means that she may also benefit from a greater share of the amounts forfeited by others if she lives a long life.

We envision that each ITA would include a subaccount that is external to the tontine pool to which a member's tontine payouts would be transferred. Thus, each ITA consists of a *tontine* subaccount and a *regular* (non-tontine) subaccount.

Tontine Subaccount

Members can contribute to their tontine subaccounts at any time, subject to any restrictions that also apply to new members. These restrictions might include, for example, a maximum age beyond which members may not contribute or join, or a limit on the size of the balance that results from a contribution.[6]

Members are generally free to select whatever investments they wish and are free to change these investments at any time by trading within the tontine subaccount. We envision a range of permissible investments similar to those typically offered within IRAs, including stocks, bonds, exchange traded funds (ETFs), and mutual funds. Permissible investments should be sufficiently liquid to facilitate the timely redistribution of forfeited account balances at death.

Contributions to the tontine subaccount are irrevocable. Monies cannot be withdrawn at will. Rather, the member selects the payout schedule at the time the contribution is made, picking from a list of options offered by the ITA provider. Once selected, the member's payout option can never be accelerated, so as to prevent the "reverse moral hazard"

---

[6] Sabin and Forman (2016) discuss the technical considerations that might give rise to such restrictions.

issue mentioned previously. However, it is permissible to allow members to change to a new payout option that decelerates their payout schedule, subject to restrictions similar to those for making additional contributions. We envision that a wide variety of payout options could be made available, which could either be annuity-like or in the nature of simple term investments. Examples of the former include lifetime payouts similar to immediate annuities, deferred lifetime payouts similar to longevity insurance, or payouts over a specified period similar to term annuities.[7] An example of the latter is the election to receive a simple lump sum on some selected date in the future, in which case the member might also elect whether dividends or tontine gains earned over the term are to be paid out along the way or reinvested to the end of the term.

The tontine subaccount is credited with interest, dividends, and capital gain distributions as they occur. Upon death, the tontine subaccount is forfeited, liquidated, and redistributed in a prescribed way to the tontine subaccounts of surviving members. Surviving members receive these redistributed forfeitures as tontine gains. Since forfeitures are always fully redistributed, the aggregate amount of tontine gains paid to the survivors always equals the aggregate amount of mortality losses forfeited by the decedents.

Payouts from the tontine subaccount are always made to the member's regular subaccount.

Regular Subaccount

The regular subaccount is wholly separate from the tontine pool. Members are free to contribute to or withdraw from the regular subaccount as they please without restriction.[8] Beneficiaries may be named for the regular subaccount, who would inherit it upon the member's death.

Payouts from the tontine subaccount to the regular subaccount can be made in cash or as in-kind transfers of securities. For example, if a member is due to receive a $3,000 payout from the tontine subaccount, this could be carried out by liquidating $3,000 worth of securities within the tontine subaccount and transferring the resulting cash to the regular subaccount, or it could be carried out by transferring $3,000 worth of securities from the tontine subaccount to the regular subaccount. It is the member's choice which method, or

---

[7] Because tontines do not involve insurance and no guarantor is involved, promises of fixed payout amounts are not possible. Thus, any annuity-like payouts would necessarily be variable. Naturally, the degree of variability will depend significantly on the volatility of the selected investment portfolio.

[8] If the ITA is registered as an IRA, any withdrawal restrictions or requirements imposed by IRA regulations may still apply.

what combination of the two methods, is to be used. The member is likely to consider tax consequences and portfolio allocation when making the choice.

Illustrating the Member Experience

We illustrate the member experience using a simplified example of an ITA pool in which tontine gains are computed and distributed yearly. We illustrate yearly processing for convenience only, recognizing that in practice the processing period might be more frequent. Forfeiture redistributions are based on member survival status at the end of each calendar year but processed with a three-month lag to allow the ITA administrator time to discover who died during the previous calendar year. Thus, forfeiture processing occurs on March 31 of each year, at which time forfeitures are apportioned and redistributed as tontine gains to those who survived through the end of the prior calendar year.

Suppose the ITA administrator uses the mortality table shown in Table 1.

Table 1: IAM Mortality Rate Table for 2019 Under Projection Scale G2 (excerpt)

Age	Male	Female	Age	Male	Female
⋮	⋮	⋮	⋮	⋮	⋮
70	0.011352	0.009200	71	0.012413	0.010047
72	0.013670	0.010977	73	0.015144	0.012003
74	0.016852	0.013153	75	0.018806	0.014480
76	0.021021	0.016018	77	0.023529	0.017793
78	0.026364	0.019854	79	0.029559	0.022275
⋮	⋮	⋮	⋮	⋮	⋮

Using this table, the administrator publishes the nominal tontine yields for each age and gender cohort for the current year as shown in Table 2. These nominal tontine yields r are determined by looking up the corresponding mortality rate q for each cohort from Table 1 and applying the formula $r = q/(1-q)$.

Table 2: Nominal Tontine Yield Table for 2019 (excerpt)

Age	Male	Female	Age	Male	Female
⋮	⋮	⋮	⋮	⋮	⋮
70	0.011482	0.009285	71	0.012569	0.010149
72	0.013859	0.011099	73	0.015377	0.012149
74	0.017141	0.013328	75	0.019166	0.014693
76	0.021472	0.016279	77	0.024096	0.018115
78	0.027078	0.020256	79	0.030459	0.022782
⋮	⋮	⋮	⋮	⋮	⋮

Consider a 75-year old male, Samuel, who establishes an ITA account on January 1, 2019 with a contribution of $100,000. He elects to receive payouts as a lifetime income stream with annual payouts using an assumed interest rate of 4%. We use annual payouts for convenience only, recognizing that many retirees might prefer monthly payouts. Samuel

10

then directs his contribution to be invested in the securities and funds of his choice. The ITA provider administers his account throughout the year, keeping track of gains and losses, reinvesting dividends and interest, taking note of which members have died, etc. Samuel's tontine subaccount statement as of March 31, 2020, the date of his first eligible payout, appears in Figure 1. Let us step through it.

Figure 1: Sample Account Statement

Tontine Subaccount Statement As Of:	March 31, 2020			
Account Overview		**Tontine Gain (if alive as of December 31, 2019)**		
Value on March 31, 2019	$ 102,613.86	Your Nominal Tontine Yield for 2019		0.019166
Market appreciation/depreciation	$962.17	x Common Group Gain Factor for 2019		0.999644
Dividends, interest, and capital gains	$ 1,600.00	= Your Actual Tontine Yield		0.019159
Balance before tontine gain	$ 105,176.03	x Your Balance Before Tontine Gain		$ 105,176.03
		= Your Tontine Gain		$ 2,015.07
Tontine gain	$ 2,015.07			
Balance before payout	$ 107,191.10			
Tontine payout (to regular subaccount)	$ (10,017.44)	Payout Option: Single Life Annuity		9.3454%
Value on March 31, 2020	$ 97,173.66			

The statement period is April 1, 2019 through March 31, 2020. Samuel's contribution of $100,000.00 was made on January 1, 2019 and grew to a value of $102,613.86 on March 31, 2019, as shown on this statement. Details about how it grew to this amount would be shown on the statement for the prior period. This statement shows a set of summarized transaction entries for this period reflecting market appreciation or depreciation, credited investment income, and realized capital gains. Added together, these give an intermediate ending balance of $105,176.03 on March 31, 2020, before the effect of any tontine processing. Up to this point, the statement reflects only investment activity and looks like a regular investment account statement.

Samuel survived the calendar year 2019 and thereby is eligible to receive his share of the pool's yearly tontine gains. He also is eligible to receive his stipulated payout, which is automatically transferred out of his tontine subaccount into his regular subaccount. Note that his tontine gain and payout take place even if he died during the first quarter of 2020 – he earned them by having survived the year 2019 and therefore his regular subaccount beneficiaries would inherit his payout (although his remaining tontine subaccount balance would be forfeited the next time forfeitures are processed, on March 31, 2021).

The shaded box in Figure 1 shows how his tontine gain is computed. First, his nominal tontine yield of 0.019166 for 2019 is selected by looking up the appropriate value for his cohort from Table 2. This is multiplied by the group gain for 2019 of 0.999644 to give his actual tontine yield of 0.019159.

The group gain value was calculated using the accounting ledger shown in Table 3, in which the subscript j denotes the jth member in the ledger. First, the account balances s_j of those members who died during 2019 are summed and found to be \$15,200,088.61. These balances are forfeited. Next, the nominal tontine gain $r_j s_j$ is computed for each member who survived the year, and the resulting values are summed to give the total nominal gain of those who survived 2019. This sum is found to be \$15,205,501.77. All of these balances s_j are measured on March 31, 2020, while instead the nominal tontine yields r_j are for calendar year 2019. In other words, these quantities are the entries for "balance before tontine gain" and "your nominal tontine yield for 2019," respectively, that appear on each member's statement for this period. Samuel's entries in Table 3 appear in bold.[9]

Table 3: ITA Member Ledger for 2019 (excerpt)

	Died in 2019 ($j \in A$)		Survived 2019 ($j \in A^c$)		
	s_j		s_j	r_j	$r_j s_j$
	xxxxx.xx		xxxxx.xx	0.xxxxxx	xxxx.xx
	xxxxx.xx		xxxxx.xx	0.xxxxxx	xxxx.xx
	xxxxx.xx		**105,176.03**	**0.019166**	**1,965.17**
	xxxxx.xx		xxxxx.xx	0.xxxxxx	xxxx.xx
	$\vdots$		$\vdots$	$\vdots$	$\vdots$
	xxxxx.xx		xxxxx.xx	0.xxxxxx	xxxx.xx
Totals	15,200,088.61				15,205,501.77

The group gain G for the year is found by dividing the total amount forfeited by those who died by the total nominal gains of those who survived. In this example, $G = \$15,200,088.61/\$15,205,501.77 = 0.999644$. Applying this factor to each survivor's nominal gain ensures that the sum of the actual tontine gains credited to survivors matches the sum of the amounts forfeited by decedents. Thus, Samuel's actual tontine yield for the year is $0.019166 \times 0.999644 = 0.019159$ and he is credited with an actual tontine gain of $0.019159 \times \$105,176.03 = \$2,015.07$. His new balance becomes \$107,191.10.

His payout is then calculated based on his selected payout option of a single life annuity. His life annuity payout rate (discussed further in the following section) for the year is 9.3454%, making the payout amount $\$107,191.10 \times 0.093454 = \$10,017.44$. Subtracting this amount gives an end-of-period tontine subaccount balance of \$97,173.66.

[9] We are implicitly assuming that each member makes a one-time contribution on January 1 of the year of enrollment and makes no subsequent contributions. The general case of contributions made anytime requires a bit more bookkeeping, which we skip here for simplicity.

Simulating ITA Outcomes

With a basic understanding of how ITAs work, we turn now to how ITAs perform. Our goal is to illustrate that the ITA remains fair to all members regardless of their ages, genders, contribution amounts, contribution timing, investment choices, and payout choices, while also dutifully delivering payouts as prescribed. To do this, we simulate ITA activity and outcomes by randomly selecting members from a population and randomly selecting investment returns from a distribution.

In our simulation, the ITA opens for business on the first day of 2019. We simulate the first 82 years of operation, covering the years 2019 to 2100, inclusive. Forfeiture and payout processing occurs annually at the end of each year. New members join at the start of each year. We recognize that in practice an ITA might perform such processing and intake more frequently, perhaps monthly or quarterly, but we choose annual processing here for the sake of simplicity. We also recognize that in practice, some amount of lag is needed between the end of the year and the time at which forfeiture processing takes place (as illustrated in the account statement of Figure 1), but to keep the simulation simple we ignore this detail.

We use the 2012 IAM Basic mortality table with projection scale G2 (NAIC, 2013). This table determines each member's nominal tontine yield. The IAM table with projection scale is a generational table, meaning that an individual's probability of death depends not only on age and gender, but also on year of birth. The table projects decreasing probability of death (i.e., a longer life) as the birth year increases. As a result, the nominal tontine yield for individuals of a given age and gender decreases with each year of ITA operation, because the probability of such individuals dying during the year decreases. We envision that the ITA provider will publish, at the start of each year, a table that lists the nominal tontine yield for each age and gender for that year.

Member Enrollment

In our simulation, 1,000 new members are enrolled at the start of each year. Thus, the simulation begins with 1,000 members at the start of 2019, has slightly less than 2,000 members at the start of 2020 (2,000 less the number who died during 2019), has slightly less than 3,000 members at the start of 2021 (3,000 less than the number who died during 2019 and 2020), and so on. Eventually the number of members reaches a plateau, when it gets large enough that the number of members who die during a year offsets, on the average, the net inflow of new members.

Each member who joins is randomly assigned parameters as follows:

- **Age**: a randomly assigned integer in the range of 65 to 85, inclusive, with all ages equiprobable. By "age" we mean the age on January 1 of the year of entry.

- **Gender**: male or female with equal probability. For simplicity we assume all accounts are individually owned, meaning that we did not include any members with jointly-owned accounts in the simulation.
- **Initial balance**: a value ranging from $1,000 to $1,000,000 and selected according to a log-uniform distribution. That is, the initial balance was selected as 10^{3U+3}, where U is a uniform random number in the range of 0 to 1. The log-uniform distribution was used because it results in a large number of members that want to contribute relatively small amounts. For example, roughly one third of the new members wish to contribute less than $10,000, and roughly two thirds wish to contribute less than $100,000. Only a small fraction wish to contribute amounts near $1,000,000. It is our guess that something comparable to this will occur in practice – many members desiring small amounts, a few members desiring large amounts. Of course, in practice we would not expect a nice structure like log-uniform; we would expect it to be much more irregular. We use the log-uniform distribution here for convenience.
- **Portfolio**: one of three portfolios, each equiprobable: 100% stock; 100% bond; 50% stock and 50% bond.
- **Payout schedule**: one of two payout schedules, each equiprobable: 10-year lump sum; a life annuity. Details of each schedule are given below.

Lump sum. A member who has the 10-year-lump-sum payout schedule receives no payouts until the end of the 10$^{\text{th}}$ year. If she is alive at the end of the 10$^{\text{th}}$ year, she is paid the entire accumulated balance, which consists of the original contribution, the accumulated investment gains (or losses) over the 10 years, and the accumulated tontine gains over the 10 years. If she dies before the end of the 10$^{\text{th}}$ year, she receives nothing.

The appeal of the lump-sum payout is that it maximizes the tontine gain that can be realized over the 10-year period. Since it is an all-or-nothing bet on surviving 10 years, it pays more at the 10-year mark than any schedule that would provide earlier payouts. We believe this would make it an appealing option for some members.

We choose a 10-year term for simplicity in the simulation. We envision the ITA provider allowing each member to select the term from a range of options, perhaps any term between 5 and 30 years. We also envision providers allowing members to hold multiple lump-sum payout contracts, each with its own terminal date and balance, allowing members to set up multiple lump-sum payouts that occur at different ages of life contingent upon surviving to those ages. Such payouts would allow the ITA provider to offer the benefits of "survivor funds" as proposed by Forman and Sabin (2016) within the single ITA, rather than having to operate a plethora of standalone survivor funds as proposed there.

Life annuity. A member who has the life-annuity payout schedule receives an annual payout in the amount of s/a, where s is his balance at the end of the year including his tontine gain for the year, and a is his current "annuity factor." His annuity factor a is the expected present value of $1 paid this year and every subsequent year for the duration of his lifetime, with future payments discounted to the present using an assumed annual interest rate of 4%. Calculating the value of a is a standard exercise.[10] If it happens that the member's investments earn exactly 4% in every subsequent year, and that the tontine's group gain is exactly 1 in every subsequent year, then the member's payout will have the same value s/a in each subsequent year as it does this year. Of course, future investment returns will not be exactly 4% each year, and future group gains will not be exactly 1. Thus, future payouts will not be constant, but instead will fluctuate according to actual investment return and group gain.

The appeal of the life-annuity payout option is that it mimics an immediate variable-income annuity that could be purchased from an insurer – meaning, an annuity that makes lifetime payouts that vary according to the value of some underlying investment. The advantage of the ITA over the insurer product is that the ITA offers an actuarially fair payout, as opposed to the insurer product which must offer a less-than-fair payout to cover its exposure to mortality risk. Thus, an ITA member with a life-annuity payout option hopes to receive a higher payout over his lifetime than he would in an insurer annuity (assuming same amount invested in each, and the same portfolio). Of course, the higher payout in the ITA is not guaranteed because it depends on the actual mortality experience of the tontine pool relative to that projected by the mortality table – that is, it depends on the group gain not being significantly below 1 on average (explained further in a later section).

Operation

The number of members is initially set to zero. Beginning with year 2019, the logic for each year of a simulation run is as follows:

At the start of each year:

1. New members for the year are enrolled.
2. Member portfolios are rebalanced (applicable only to those who selected the 50/50 blend of stocks/bonds).

---

[10] The formula for the annuity factor at age x is $a = \ddot{a}_x = 1 + \sum_{t=1}^{\infty} v^t \, {}_tp_x$, where ${}_tp_x$ is the probability of surviving to age $x + t$ given that the member is alive at age x, and $v = 1/(1 + i)$ is the discount factor, with i the assumed interest rate (e.g., 4%). The value of ${}_tp_x$ is calculated from the mortality table. The IAM mortality table used here has a terminal age of 120, meaning there is zero probability of surviving to ages greater than 120, and so the sum in the formula has a finite number of terms.

At the end of each year:

1. The balance s_j for each member j is calculated based on the investment market outcome for the member's portfolio that year.
2. The group gain G is calculated using formula (1).
3. Each surviving member's account is credited with a tontine gain equal to $Gr_j s_j$, where r_j is the member's nominal tontine yield for the year. The member's balance is updated to reflect the credited tontine gain.
4. Each surviving member's payout is deducted from the member's balance, according to the member's payout contract.
5. Each decedent's balance is set to zero.
6. Members with zero balance are deleted. Zero balance is the case for decedents and for survivors whose payout equals a full withdrawal of the balance.

We performed 10,000 simulation runs, each run spanning the 82 years from 2019 to 2100. In each simulation run, random investment returns for the stock and bond asset classes were generated for each of the 82 years. Each year's investment return of the two asset classes was generated using a jointly log-normal distribution with parameters as shown in Table 4. A total of $82 \times 10{,}000 = 820{,}000$ random investment returns were generated for each asset class.

Table 4: Investment Return Parameters (annual)

	Arithmetic Mean	Standard Deviation	Correlation	
Stock	9.0%	18.0%	1.0	
Bond	5.5%	6.5%	0.3	1.0

At the start of each year, 1,000 new members were enrolled with randomly selected parameters (age, gender, amount, portfolio, payout schedule) as described above. These parameters are the same on every run, meaning they were randomly selected one time and then used on every run. Thus, we simulate a single, randomly selected population of $82 \times 1{,}000 = 82{,}000$ members. Each member of the population was assigned a random year of death, which changes with every run; that is, a total of $82{,}000 \times 10{,}000 = 820{,}000{,}000$ random years of death were simulated.

Each member's year of death was randomly generated using the member's probability of death as defined by the IAM Basic mortality table. This is the same table that the ITA uses to calculate the member's nominal tontine yield. Thus, the underlying distribution that governs member death times is an exact match to what is assumed by the tontine. In other words, we are simulating only the effect of idiosyncratic risk, and are ignoring the effect of systematic risk for now. We discuss systematic risk in a later section.

In each simulation run, forfeiture allocation was executed on the last day of each year using the nominal-gain method, and the corresponding group gain value was calculated. Thus, a total of $82 \times 10{,}000 = 820{,}000$ group gain values were calculated.

Assessing Fairness

For as long as they survive, fair tontine investors will hope to achieve the same return as if they had invested outside of a tontine, plus a nominal tontine yield. The latter reflects the expected value of the actual tontine yield that results from fairly pooling mortality risk with other members of the tontine.

We postulate that members will believe their tontine investment to be successful and perceive the tontine to be fair if they consistently receive tontine yields that are close to their nominal values. We say "close to" because members should understand that their actual tontine gains will naturally vary from their nominal values to some degree because although mortality pooling effectively mitigates mortality risk, it does not eliminate it completely. It is easy to see that if the group gain G is close to 1, members will receive actual tontine yields of rG that are close to their nominal values of r. Thus, an assessment of fairness boils down to the characteristics of the group gain.

Simulation Results

Preliminaries

Pool size. The number of members in the ITA pool changes over time. Figure 2 plots the number of members at the start of each year, averaged over the 10,000 simulation runs.

Figure 2: Average Number of Members

Average number of members at start of year
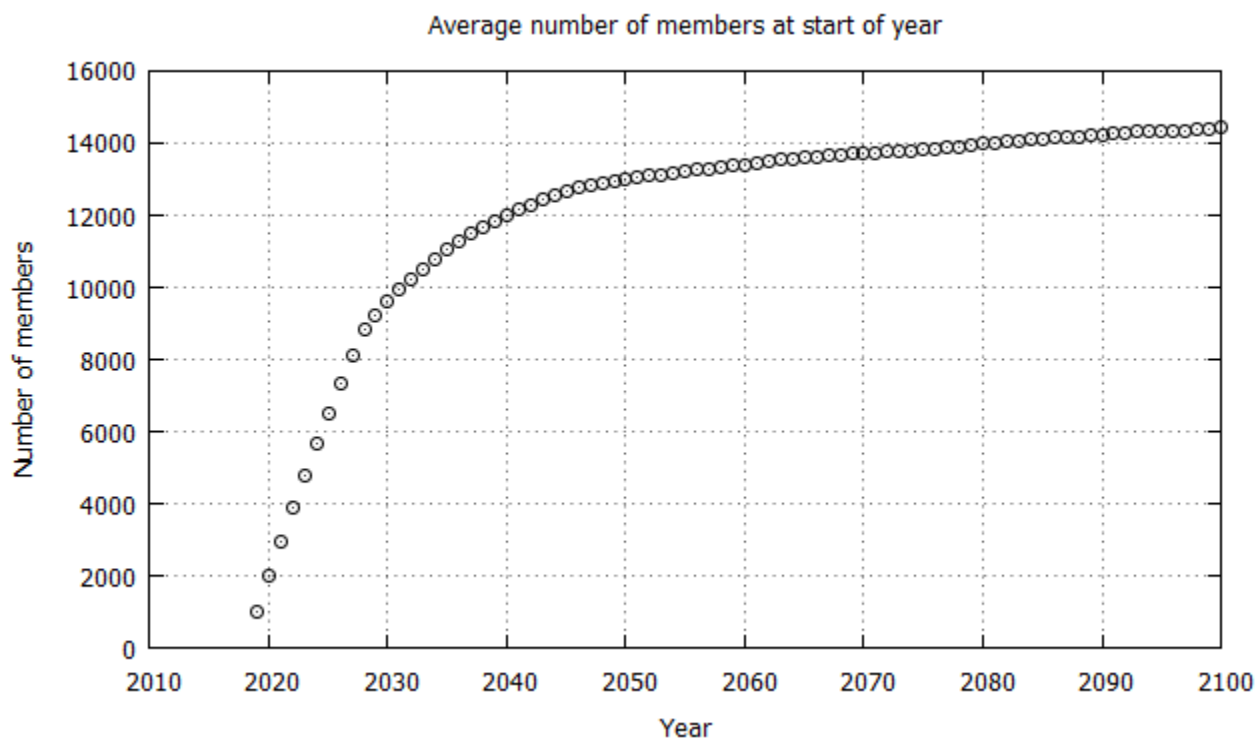
For the first ten years, 2019 through 2028, the plot roughly follows a straight line. This is because 1,000 new members join at the start of each year, while the number who die is relatively low. At the start of the tenth year, 2028, a total of 10,000 people have joined, and on the average about 1,160 of those have died during the preceding nine years, for an average of about 8,840 members at the start of 2028.

Beginning in 2029, the plot shows a slower growth in the average number of members. This is due to the expiration of 10-year lump sum contracts. Approximately half of the members who joined in 2019 have lump-sum contracts, and all of them will have exited the tontine by the end of 2028 – some by dying, but most because they outlived the 10-year term. The same is true at the start of subsequent years – about half the members who joined ten years earlier have lump-sum contracts, and those who did not die will now be terming out. In short, the growth rate after 2029 is slower than before because of lump-sum term-outs.

Eventually the growth rate slows, when the pool becomes large enough that a near-equilibrium is reached where, on the average, the number of deaths plus the number of lump-sum term-outs matches the number of new entrants (1,000 per year). However, the equilibrium is not static, because mortality rates decrease as time goes by. As a result, the growth rate does not completely halt, because decreasing mortality pushes up the pool size needed to get the offsetting number of deaths.

We emphasize that this is only one example, an example that reflects the way we chose to do the simulation. In practice the average number of members will be very different than this, because there will not be exactly 1,000 members joining each year, and their parameters (age, gender, initial amount, portfolio, payout contract) will not be so nicely assigned. Our goal here is merely to create a dynamic membership that we can use to illustrate ITA behavior.

Group gain. Figure 3 plots the mean and standard deviation of the yearly group gain as it occurred over the simulation runs. For each year, the simulation produced 10,000 sample values of the group gain, one sample from each of the 10,000 runs. The mean μ and the standard deviation σ of these 10,000 samples were calculated, and the three points μ, $\mu + \sigma$, and $\mu - \sigma$ are plotted. For example, for the year 2019, the mean and standard deviation of the 10,000 samples were calculated as $\mu = 1.00998$ and $\sigma = 0.39094$, so the plotted values for 2019 are 1.00998, 1.40092, and 0.61904.

Figure 3: Group Gain
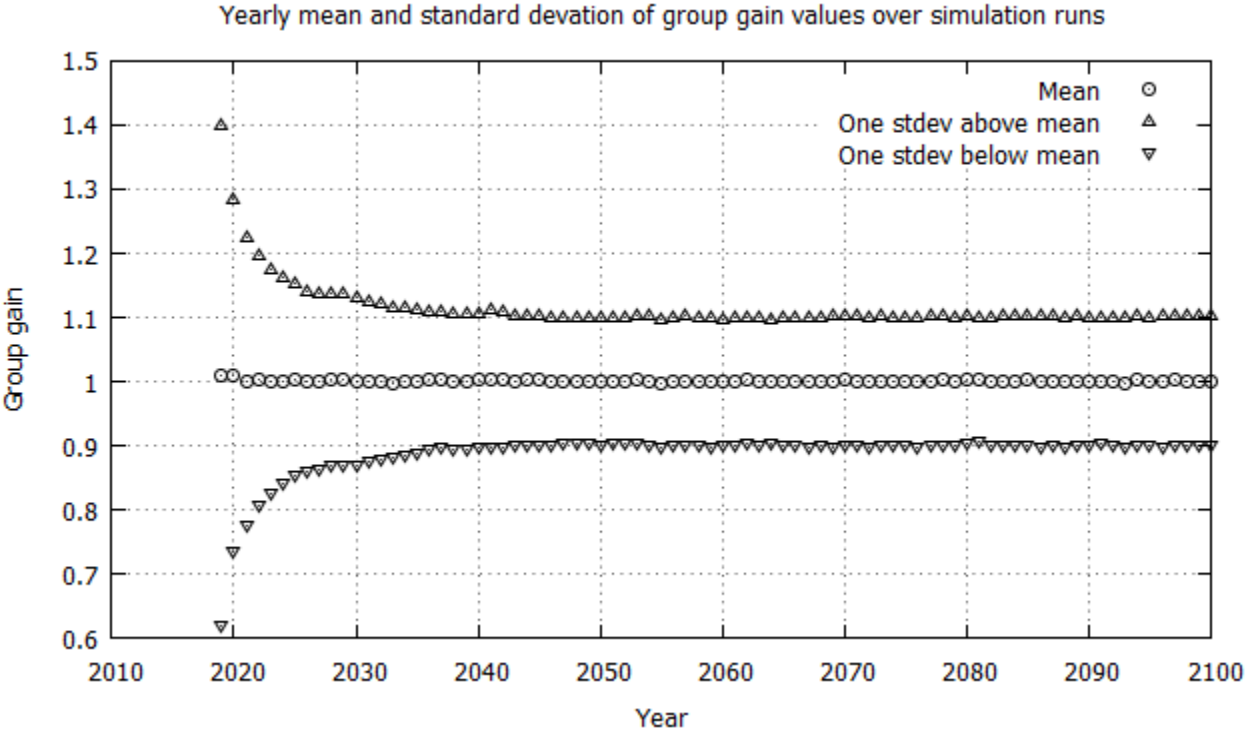
Importantly, the plot shows that the mean value of the group gain is very close to 1 in every year. This validates that the tontine design is fair, or at least approximately so, because on the average each survivor receives an actual tontine gain Grs that is close to his nominal tontine gain rs.

The actual group gain varies randomly about 1 over the simulation runs, and the extent of that variation is quantified by the standard deviation. The standard deviation is greatest

19

in early years, when the number of members is small, and it decreases as the number of members grows. This can be understood by referring to the group-gain formula (1). The numerator and denominator in the formula are each random sums. The expected value of each sum is identical.[11] Intuitively, by the law of large numbers, each sum becomes more tightly distributed about its expected value as the number of members grows, and thus the ratio of the two becomes more tightly distributed about 1. This intuition is validated by the plot.[12]

Figure 4 illustrates that the group gain is uncorrelated with portfolio returns. The figure is a scatter plot of 820,000 points, one point for each of the 82 years in each of the 10,000 simulation runs. The y-coordinate of a point is the value of the group gain for a particular year in a particular run. The x-coordinate is the difference between the returns of the stock and bond portfolios for that year and that run. A positive x-coordinate means stocks outperformed bonds, and a negative coordinate means bonds outperformed stocks. For example, suppose that in year 2030 of the fifth simulation run, the group gain value was 1.1, the stock portfolio returned 10%, and the bond portfolio returned 1%; then stocks outperformed bonds by $10 - 1 = 9$%, and the plotted point for that year and run would be (9, 1.1). And if in the year 2050 of the tenth simulation run, the group gain was 0.8, the stock portfolio returned 2%, and the bond portfolio returned 5%, then stocks underperformed bonds by $5 - 2 = 3$%, and the plotted point for that year and run would be (–3, 0.8).

---

[11] The expected value of the numerator is $\sum_j q_j s_j$, and the expected value of the denominator is

$$\sum_j (1 - q_j) r_j s_j = \sum_j q_j s_j.$$

[12] For an analytical discussion, see Sabin and Forman (2016).

Figure 4: Scatter Plot of Group Gain and Stock Return Premium

Scatter plot of yearly group gain versus market return, from simulation runs

Visually, the scatter plot shows no obvious correlation between the x and y coordinates. That is, the group gain values (the y values) appear evenly distributed about 1 regardless of the relative return of stocks versus bonds (the x values). To quantify the correlation, the figure includes a trend line, meaning a straight line fitted to the plotted points using linear regression. The formula for the trend line is:

$$y = 1.0009074 - 0.0000013x.$$

This is an excellent match to the horizontal line $y = 1$, validating the uncorrelation.

The uncorrelation can be understood by again referring to the group-gain formula (1). The expected value of the numerator matches the expected value of the denominator. This is true regardless of the values of the s_j terms, meaning regardless of the balances of members. Those balances are, of course, affected by portfolio returns, but the expected value of numerator and denominator remain identical regardless. Intuitively, then, the effect of portfolio return is self-cancelling as far as the average group gain is concerned. The plot validates this intuition.[13]

These are perhaps the most significant properties of the tontine, that the average value of the group gain is 1, and that the group gain is uncorrelated with market return. They

---

[13] For an analytical discussion, see Sabin and Forman (2016).

are what allow investors to select their own portfolios and payout schedules, in whatever amounts they wish, without affecting the expected value of each other's tontine gain. The properties are exactly true under the idealization that the tontine is exactly fair; they are approximately true here in our tontine design that uses the nominal-gain method. In this simulation, the approximation is indistinguishable from exact.

Yearly Tontine Yield

It is worth noting that all members of a given cohort – meaning all members of a given birth year and gender – experience the same yearly tontine yield. A member's yearly tontine yield has value rG, where r is the member's nominal yield during for the year, and G is the tontine's group gain for the year. All members of his cohort share the same value of nominal yield r since they have the same probability of dying during the year, and all members of the tontine regardless of cohort share the same value of group gain G. Thus, the yearly tontine yield rG is the same for all members of his cohort. This is true even if cohort members invest in different portfolios, or have different payout contracts, or join the tontine at different times, and it is true even if actual mortality varies widely from what is projected (G much different than 1). Thus, when discussing yearly tontine yield, we do so in terms of cohorts rather than in terms of individual members.

Figure 5 illustrates the yearly tontine yield for the male cohort who turns 65 in the year 2019, meaning the male cohort born in 1954.

Figure 5: Yearly Tontine Yield, by Age

Yearly tontine yield (male, age 65 in 2019)
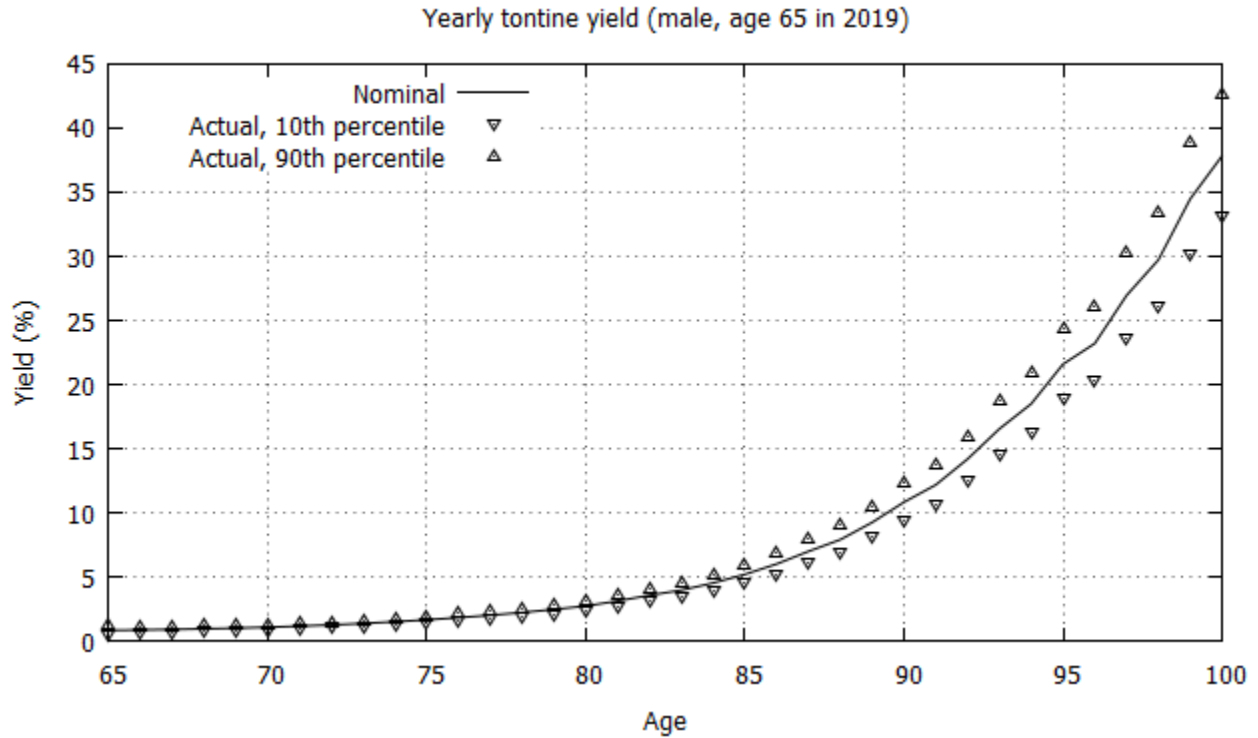
The nominal value of the cohort's yearly tontine yield during age x is $r_x = q_x/(1 - q_x)$, where q_x is the probability of dying during the year of age x. This is the curve labelled "nominal" in the figure. The nominal yield increases with age, since the probability of dying increases. It is less than 1% at age 65, is about 5% at age 85, and is upwards of 20% after age 95.

The actual value of the cohort's yearly tontine yield is $r_x G_x$, where G_x is the tontine's group gain during the cohort's year of age x. The figure shows two percentile curves that represent the actual yield as it occurred in the simulation. The curve labelled "10th percentile" means that, at each age, the cohort's actual yield was less than the plotted point in 1,000 simulation runs and greater than the plotted point in 9,000 runs. Similarly, the curve labelled "90th percentile" means that, at each age, the cohort's actual yield was less than the plotted point in 9,000 simulation runs and greater than the plotted point in 1,000 runs. Note that the percentiles are determined for each age separately – meaning, for example, the 1,000 runs that lie below the 10th percentile curve at age 70 are different than the 1,000 runs that lie below the curve at age 80. The percentile curves straddle the nominal curve in approximate symmetry with deviation from nominal increasing with age. But the scale of the figure makes it hard to discern much about the deviation.

Figure 6: Yearly Tontine Yield, Relative to Nominal, by Age
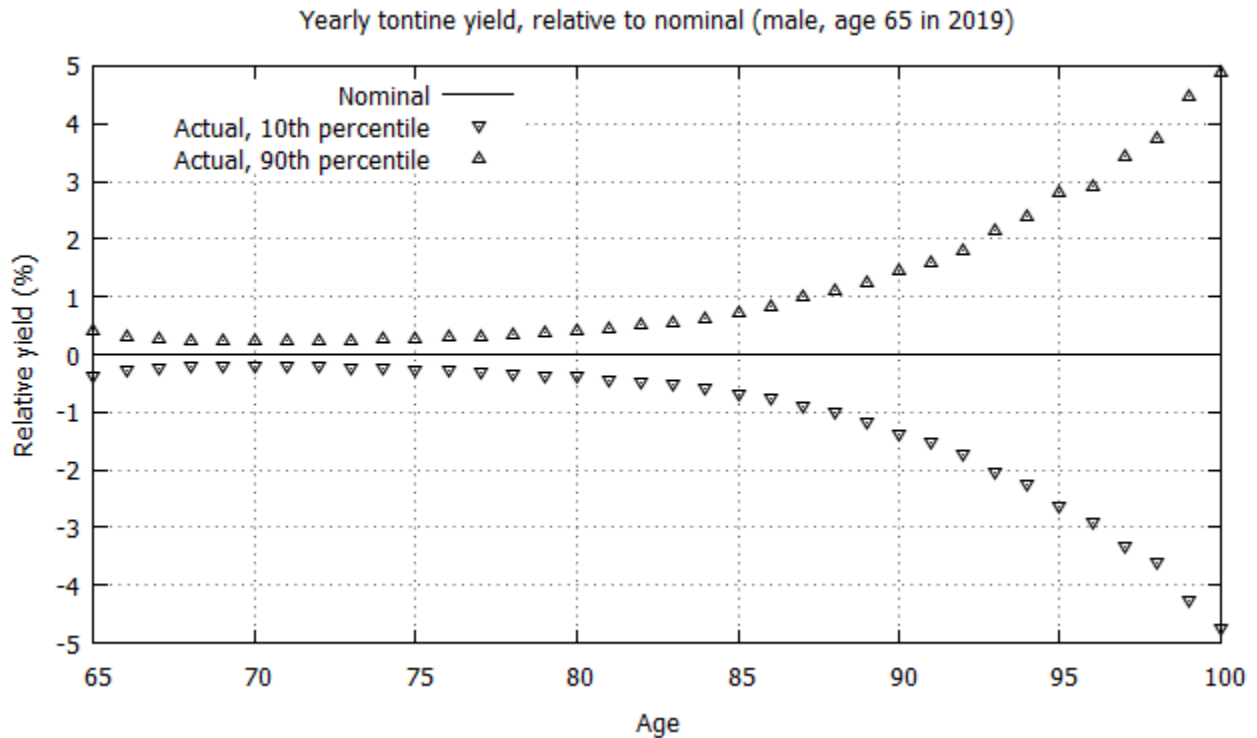
For a closer look at the deviation, Figure 6 shows the actual yield relative to the nominal yield. That is, it plots the percentiles of the values $r_x(G_x - 1)$ that occurred during the cohort's age x over the simulation runs.[14] At age 65, the 10th and 90th percentiles are roughly $\pm 0.4\%$. That is, 80% of the simulation runs resulted in an actual yearly yield within 0.4% of nominal, with 10% of the runs having an actual yield worse than 0.4% below nominal, and 10% of the runs having an actual yield better than 0.4% above nominal. The deviation of the percentile curves initially decreases, until about age 72. This is because these are the first years of operation of the tontine (it began in 2019, when this cohort was age 65), and the number of participants grows rapidly each year, such that the variance in the group gain decreases (see Figure 3). Thereafter, the deviation of the percentile curves increases with age, because the variance of the group gain stabilizes, while the value of r_x increases with age. The increasing value of r_x amplifies the effect of group-gain variance on the actual yield $r_x G_x$, resulting in increasing deviation of the percentile curves. At age 87 the percentile curves deviate by roughly $\pm 1\%$ from nominal, and the deviation increases rapidly from there to about $\pm 5\%$ at age 100.[15]

---

[14] An equivalent description is that Figure 6 subtracts the nominal curve from each curve in Figure 5.

[15] We remark that the plots in Figure 5, Figure 6, Figure 7, and Figure 8 are strictly correct only if, in every simulation run, there is at least one member of the cohort who survived to age 100. This was not always the case – in some simulation runs, all members of the cohort died earlier than age 100. However, the plots

24

Annualized Tontine Yield

We can express a member's cumulative tontine gain in terms of an annualized tontine yield by computing the geometric mean of the member's yearly tontine yields. Expressed over the period from age m to age n, the annualized tontine yield is:

$$Y_{m:n} = \left(\prod_{x=m}^{n} (1 + r_x G_x) \right)^{\frac{1}{(n-m+1)}} - 1.$$

The annualized tontine yield is the average value of the member's yearly tontine yields, in the sense that if the yearly tontine yield from age m to age n had been the fixed value $Y_{m:n}$ every year, this would have resulted in the same balance and payout at age n as the member's actual balance and payout. For a detailed discussion, see Appendix A.

The nominal value of the annualized tontine yield is:

$$r_{m:n} = \left(\prod_{x=m}^{n} (1 + r_x) \right)^{\frac{1}{(n-m+1)}} - 1 . \tag{2}$$

If the group gain G_x equals 1 during each year of the member's age m to age n, then the actual value of the annualized tontine yield $Y_{m:n}$ matches the nominal value $r_{m:n}$. In general, of course, the group gain does not equal 1, so the actual value of the annualized yield differs from its nominal value. Since the group gain is non-negative, the actual value of the annualized yield is never less than 0.

Since the yearly tontine yields are identical for all members of a given cohort, the annualized tontine yields $r_{m:n}$ are the same for all members who belong to the same cohort and who are alive and participating in the tontine from age m to age n. This is true even if they invest in different portfolios, or have different payout contracts, or join the tontine at different times, and it is true even if actual mortality varies widely from what is projected. Thus, when discussing annualized tontine yield, we do so in terms of cohorts and age ranges.

Figure 7 illustrates the annualized tontine yield for the male cohort who turns 65 in the year 2019 (i.e., birth year 1954) for those who joined the tontine in 2019.

---

are still correct if interpreted as representing what a cohort member would receive if he survived to age 100, assuming that his survival has negligible effect on the group gain.

Figure 7: Annualized Tontine Yield, by Age
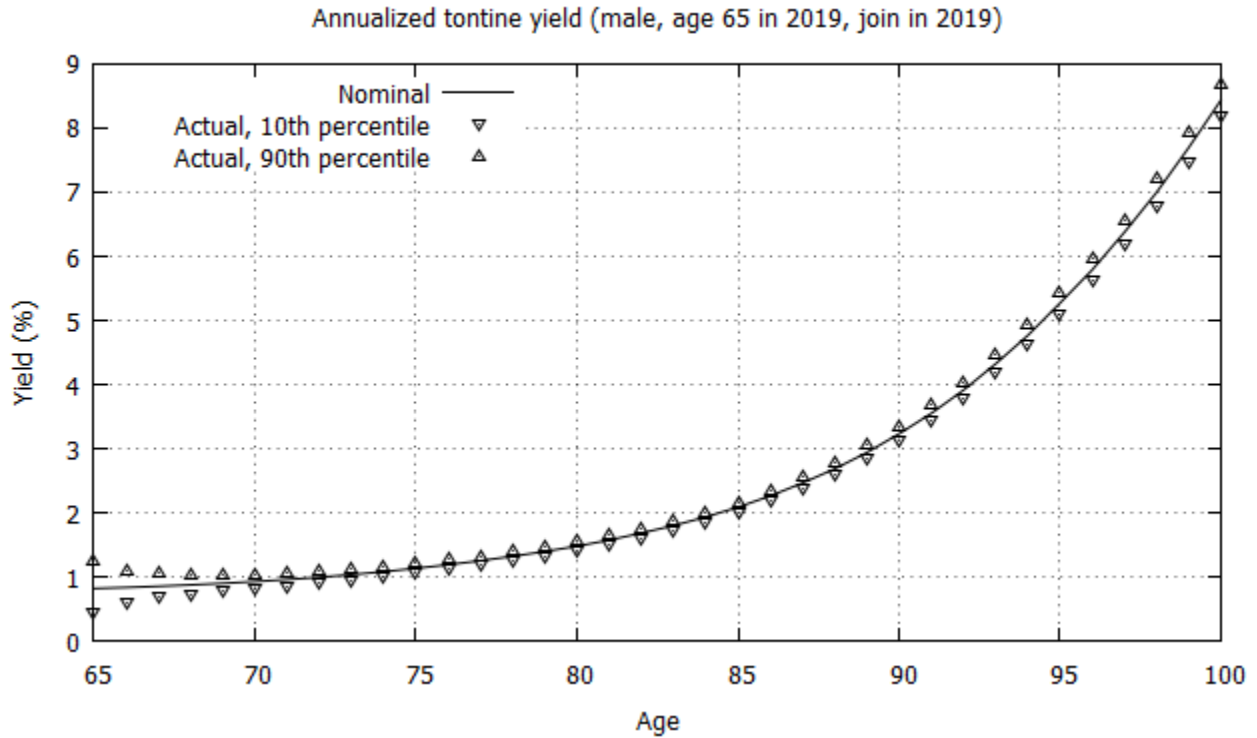
The nominal curve in Figure 7 plots the nominal value $r_{65:n}$ of the annualized tontine yield for these members. The nominal annualized yield increases with age because the nominal yearly yield r_x increases with age, as seen in Figure 5, and thus the geometric mean in formula (2) increases with age. However, the averaging effect of the geometric mean causes the annualized yield in Figure 7 to grow more slowly than the yearly yield in Figure 5.

The actual value $Y_{65:n}$ of the annualized tontine yield for these members as it occurred in the simulation is represented in the figure by two percentile curves, the 10th and 90th percentiles, similar to those in Figure 5. The two percentile curves straddle the nominal curve in approximate symmetry, with deviation from nominal initially decreasing with age and then increasing with age. But the scale of the figure makes it hard to discern much about the deviation.

Figure 8: Annualized Tontine Yield, Relative to Nominal, by Age
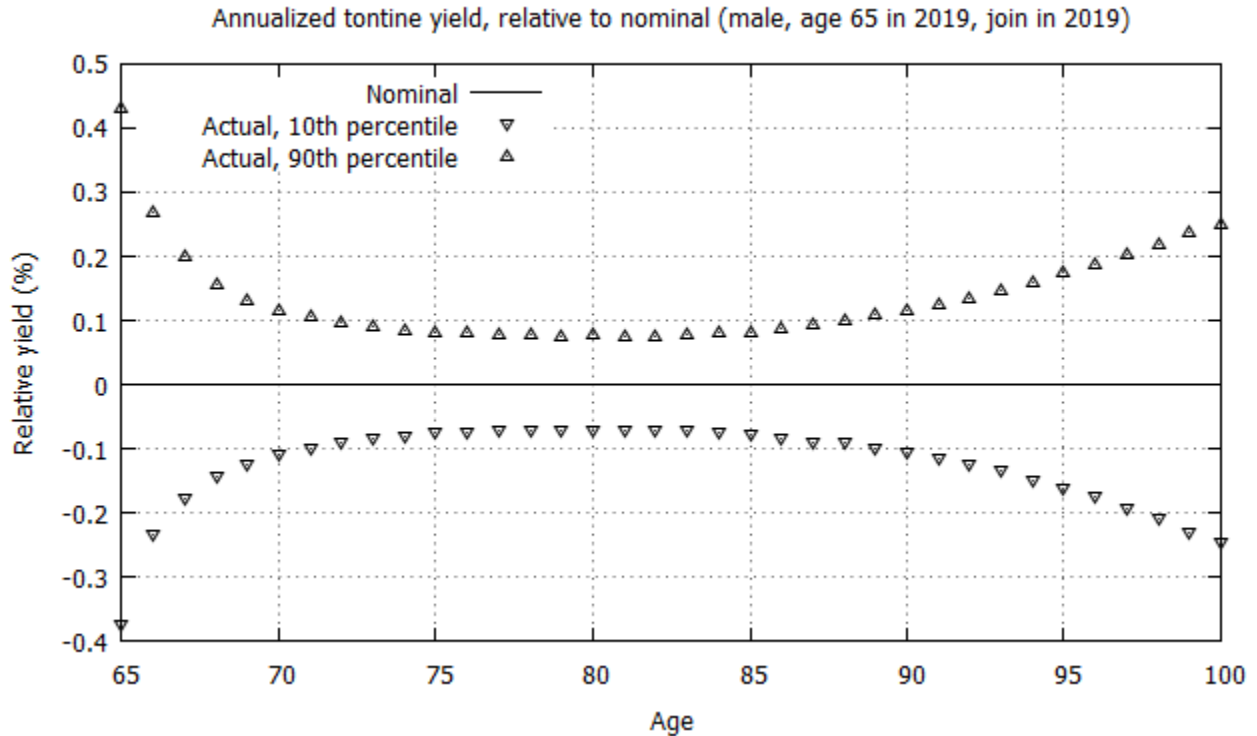
For a closer look at the deviation, Figure 8 shows the actual annualized yield relative to the nominal annualized yield. That is, it plots percentiles of the values $Y_{65:n} - r_{65:n}$ that occurred in the simulation during age n for these members.[16] At age 65, the 10th and 90th percentiles are roughly $\pm 0.4\%$ – the same as in Figure 6, since only a single year is involved. Also as in Figure 6, the deviation of the percentile curves decreases during early years, because the number of participants grows and the variance in the group gain decreases. The deviation starts growing in later years, beyond roughly age 82, but at a much slower rate than the deviation growth in Figure 6. For example, at age 95, the percentile curves are only about $\pm 0.2\%$ in Figure 8, compared to $\pm 2.5\%$ in Figure 6. The deviation in Figure 8 is lower at advanced ages because the annualized yield is an average of the yearly yields, and the averaging process tends to smooth out the up-and-down deviations in the yearly yields.

Persistence

The desired characteristics of the relative tontine yield need to be persistent since ITAs are meant as open-ended offerings operated in perpetuity. To illustrate that this is indeed the case in our simulation, Figure 9 shows the relative tontine yield for selected age/gender

---

[16] An equivalent description is that Figure 8 subtracts the nominal curve from each curve in Figure 7.

pairings by calendar year since the ITA is first incepted. That is, it plots the percentiles of the values

$$r_{[age,gender,year]}(G_{year} - 1)$$

versus year, where $r_{[age,gender,year]}$ is the nominal yield of a member of the specified age and gender for that year, and G_{year} is the group gain for that year.[17]

Figure 9: Yearly Tontine Yield, Relative to Nominal, by Calendar Year
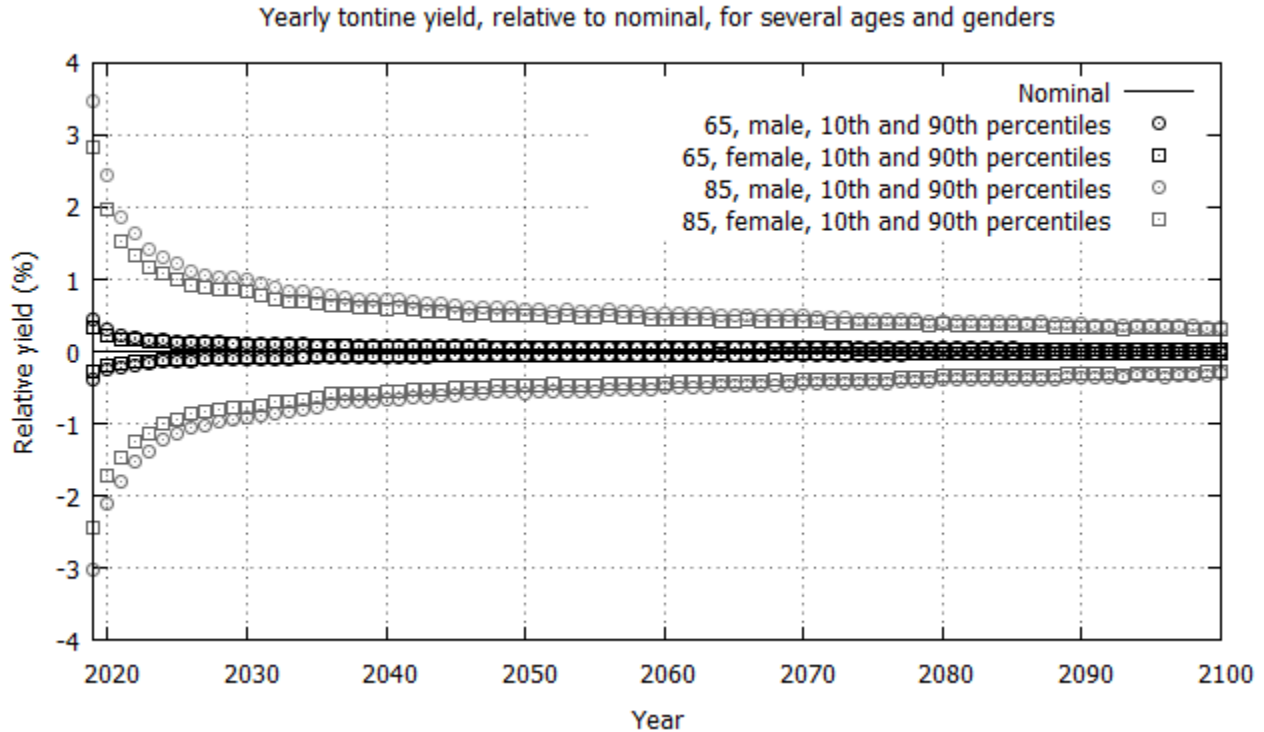
Note that the data points in Figure 9 for the age-65 male in 2019 match the data points in Figure 6 for age 65. Similarly, the data points in Figure 9 for the age-85 male in 2039 match the data points in Figure 6 for age 85.

In the early years of the ITA, the 10th/90th percentiles deviate from nominal by a relatively large amount due to the relatively low enrollment (see Figure 2). This deviation decreases as enrollment increases and has mostly settled down by about 2030, when the ITA's enrollment has reached about 10,000 members. Thereafter there is a slow decrease in the percentile deviation over time due to decreasing mortality, which causes a slow decrease in nominal yield over time.[18] The deviation remains well behaved in that it is

---

[17] For comparison, Figure 6 plots $r_{[age,male,2019+age-65]}(G_{2019+age-65} - 1)$ versus age.

[18] Recall that we apply an improvement scale to the nominal mortality rates each year.

symmetric around a mean value of 0, meaning that on the average members continue to receive actual tontine yields that are close to their nominal yields.

Simulation Insights

Our simulation illustrates how the fair-tontine principle can be applied to the concept of mortality-pooled brokerage accounts that are perpetually open to new members. In the same way as IRA brokerage accounts, individuals can invest and trade as they choose. They may also choose from a wide array of payout methods. ITAs are fair to all members regardless of the demographics, balances, investment choices, and payout choices of the other members. Furthermore, this fairness is persistent – ITAs remain fair in perpetuity, regardless of when a member elects to join.

The degree of variation in ITA outcomes depends on the size of the membership pool. To provide additional context, recall that each member's realized tontine yield is a function of his predetermined nominal tontine yield r scaled by a common group gain G. Since the nominal tontine yield for each member is given, the effect of pool size boils down to its effect on the group gain. The effect is complex since it depends on the composition of ages, genders, and balances of the members, but in general we can say that the variance of the group gain decreases with pool size. The underlying principle here is the effect of the law of large numbers in diversifying away mortality risk.

With regards to the diversifying effect of pool size, we can make the following statement. Suppose we start with a given tontine pool and then enlarge that pool by a factor of k by replicating each member $k-1$ times (meaning k members of each age, gender, balance, portfolio, and payout schedule specified in the original pool). The variance of the group gain in the enlarged pool will be reduced by the factor $1/k$, and the standard deviation by $1/\sqrt{k}$. For a derivation of this result, see Sabin and Forman (2016).

So, if we were to double the number of new members added each year in our simulation, we would expect the standard deviation to be reduced by a factor of $1/\sqrt{2}$. And if we were to increase the number of members one-hundredfold, we would expect the standard deviation to be reduced by a factor of $1/10$. These factors are only approximate – their actual values depend on the ages, genders, and balances of the members. In our simulation, where each member's parameters are randomly chosen from fixed distributions, we would expect the actual factors to be close to the values here. But in practice, the parameters of new members will not be so nicely distributed, and we cannot expect the standard deviation to follow the $1/\sqrt{k}$ formula too precisely. Nevertheless, the trend applies – the larger the pool size, the smaller the standard deviation. Thus, if ITAs were to become widely adopted, we would expect to see a lower group gain variance than illustrated in our simulation.

Effect of the Investment Decisions of Others

We opened by stating that an individual member's results are largely unaffected by the investment choices of the other members. Despite the preceding analysis that demonstrates this, it may still not be intuitive to the reader why a successful aggressive investor is not disadvantaged by being in a pool with conservative investors who are likely to die with lower balances than would be the case if they had invested more aggressively. Wouldn't members want other members to die with larger balances?

While it is true that investment decisions made by members may result in either smaller or larger balances and therefore smaller or larger forfeiture amounts, what matters is the *reallocation* of those forfeiture amounts. The key is to ensure that reallocations remain actuarially fair to each member based on his own mortality rate q and account balance s. By ensuring that the reallocation is fair, each member will have an expected nominal tontine gain of $rs = sq/(1-q)$ regardless of how anyone else invests – even the aforementioned aggressive investor who is in a pool dominated by conservative investors. Three concepts are fundamental to understanding this. First, by successfully investing aggressively his balance s will be higher than if he had invested conservatively. Second, his share of forfeiture reallocations will also be higher because a higher value of s results in a higher value of rs. Third, while the investment choices of the members will of course have an impact on their balances and therefore will also have an impact on the group gain, it will do so for both the numerator and the denominator of the group gain formula (1). Most importantly, the expected value of the group gain will remain very close to 1, and thus the expected tontine gain rsG received by each member will remain very close to its fair value of rs.

To illustrate this, we reran our simulation using the same assumptions as before except for the selection of investment portfolios. This time, each member randomly chose either a 100% stock portfolio or a 100% bond portfolio, with a 5% probability of selecting the stock portfolio and a 95% probability of selecting the bond portfolio. Recall that each member's realized tontine yield is simply his nominal tontine yield scaled by the group gain, and so fairness boils down to the characteristics of the group gain. For this reason, we focus on the new simulation's group gain as evidence that the tontine pool remains fair.

Group gain. Figure 10 plots the mean and standard deviation of the yearly group gain over the new simulation runs. It looks very similar to the original simulation result shown in Figure 3.

Figure 10: Group Gain When Portfolio Selection is Skewed Conservatively
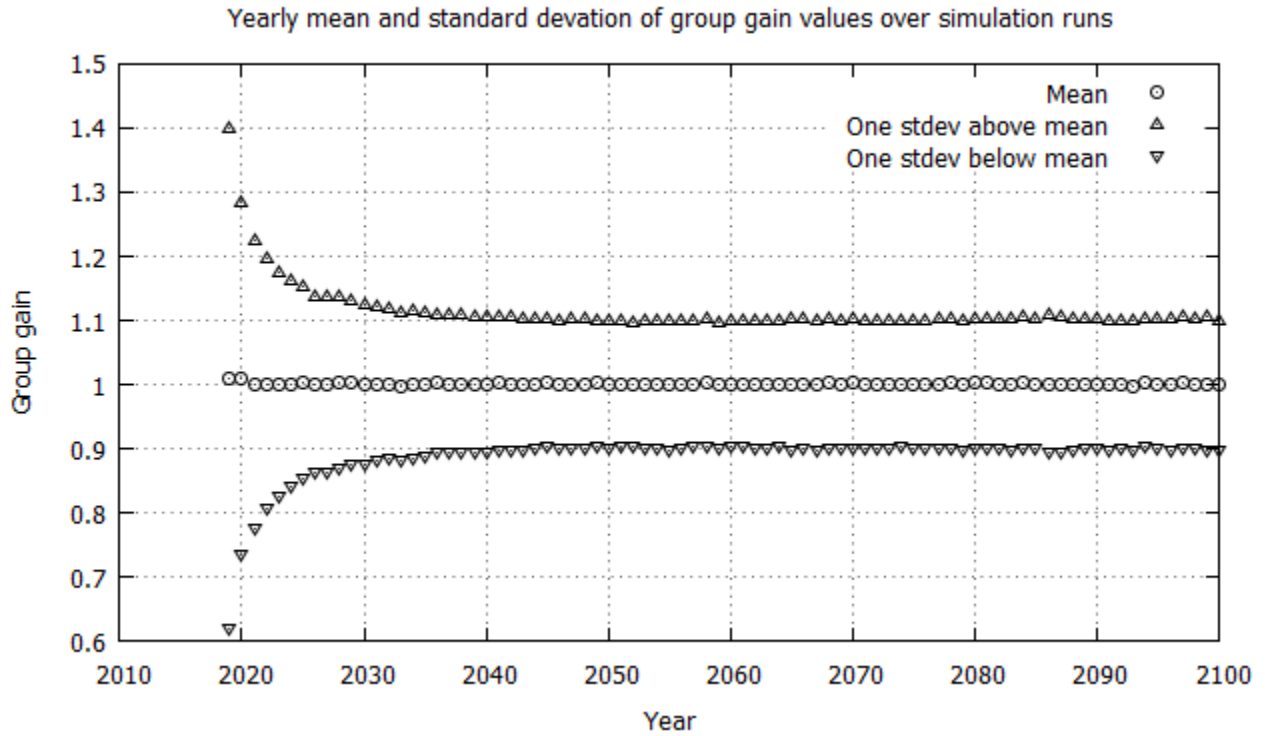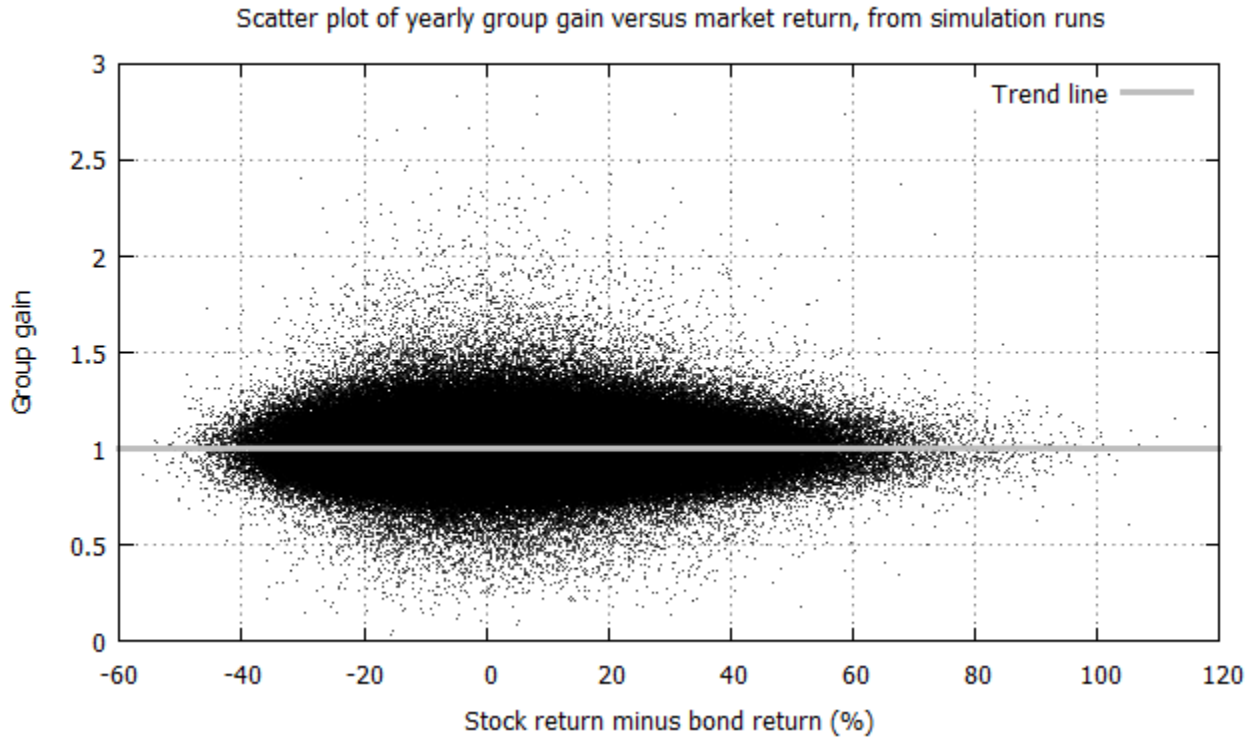

Figure 11 illustrates that the group gain remains uncorrelated with portfolio returns. It likewise looks similar to the original simulation results shown in Figure 4.

Figure 11: Scatter Plot of Group Gain and Stock Return Premium

Scatter plot of yearly group gain versus market return, from simulation runs

This time the formula for the trend line is:

$$y = 1.0008250 + 0.0000093x.$$

Importantly, these plots again reveal that the mean value of the group gain remains very close to 1 in every year. This validates that the tontine design remains fair, or at least approximately so, because on the average each survivor receives an actual tontine gain Grs that is close to his nominal gain rs.

Benefits of Individual Tontine Accounts

ITAs represent a unique value proposition in that they give their members an extensive amount of choice similar to IRAs, with the added benefit of earning a tontine yield on top of any underlying investment returns. Tontine yields represent the best kind of portfolio returns in that they are 1) uncorrelated to the returns of the investments, and 2) always nonnegative. Furthermore, because ITAs pool mortality risks, they offer a way to generate lifetime payouts in a way not possible with traditional IRA investment accounts, which offer no such pooling.

Of course, insurance companies offer mortality-pooled lifetime income in the form of payout annuities. The comparable annuity product is a variable-income annuity (VIA). VIAs typically make payouts similar to the life annuity payout used in our simulation, discussed previously. The annuitant selects an underlying benchmark portfolio from some

allowable set of options. He receives an initial payout based on his annuity factor and the initial value of his investment. The annuity factor represents the expected present value of $1 paid annually (or on some other selected frequency) for the duration of his lifetime, with future payments discounted to the present using an assumed interest rate that reflects the anticipated return of his benchmark portfolio. If the actual return on the portfolio exactly matches the assumed interest rate every year, then the payout will have the same value every year. Otherwise the payout will fluctuate up or down based on the actual return of the portfolio relative to the assumed interest rate.

Both ITAs and VIAs offer payouts that vary with investment performance in a similar manner. Both can effectively diversify away the idiosyncratic component of mortality risk. The primary difference between the products lies in who bears the undiversifiable, systematic component of mortality risk, which represents the risk that the overall population of annuitants/members may die at an aggregate rate that is different than anticipated. In other words, it represents the risk that the underlying mortality table and improvement scale used by the product designer turns out to be wrong. With insurance annuities such as VIAs, the insurer bears this systematic risk. In ITAs, the members bear it.

Because insurers bear this risk and back it with a guarantee, they are required to ensure their solvency by pricing in a suitable risk premium. Such risk premiums do not apply to ITAs since they offer no such risk transfer or guarantee. Theoretically, then, the payout of ITAs should be higher than that of VIAs invested in the same portfolio of assets.

We examine this by comparing the "mortality yields" of ITAs versus VIAs. Mortality yields represent the amount credited as a result of mortality pooling. For ITAs, which operate quite transparently, this is quantified by the tontine yield. VIAs, which operate much more opaquely, also provide a similar type of mortality yield that arises from the fact that annuitants who die relatively early essentially subsidize those who outlive them.

To measure this mortality yield, we model VIAs using a different mortality table – this time, the 2012 IAR Period mortality table, again using projection scale G2 (NAIC, 2013). The IAR table is a suitable choice because it is derived by applying a reserve margin to the IAM table to provide a cushion against the systematic mortality risk that an insurer bears in selling individual annuities. The IAM table represents *expected* mortality rates for the segment of the population who buy annuities (or tontines), while the IAR table is more conservative due to its reserve margin. IAR mortality rates are lower than IAM mortality rates – approximately 90% lower from birth through age 100, then gradually increasing to 100% of the IAM mortality rate at higher ages.

We compute "no load" VIA mortality yields $\hat{r}$ using the formula $\hat{r} = \hat{q}/(1 - \hat{q})$, where $\hat{q}$ represents the corresponding cohort mortality rate from the projected IAR table. By no

load, we mean a VIA with no mortality and expense fee, no administrative fees, etc. In other words, the VIA mortality yield $\hat{r}$ is similar to our nominal tontine yield r except that it is computed from the IAR table rather than the IAM table.[19] Net mortality yields on VIAs that include an additional fee loading can also be computed using the formula $\hat{r}_{net} = \hat{r} - e(1 + \hat{r})$, where e represents the fee loading as an expense ratio against assets (see Appendix B for details).

Yearly Yield

As shown previously, an ITA's tontine yield each year will depend on actual mortality experience, which is quantified on a relative-to-nominal basis by the scaling factor we call the group gain. Conversely, since the VIA transfers *all* mortality risk to the insurer, its mortality yield (though opaque and undisclosed to the annuitant) each year is guaranteed. Figure 12 compares the yearly tontine yield of an ITA to the yearly mortality yield of a VIA for the male cohort who turns 65 in the year 2019.

Figure 12: Yearly Tontine/Mortality Yield, by Age
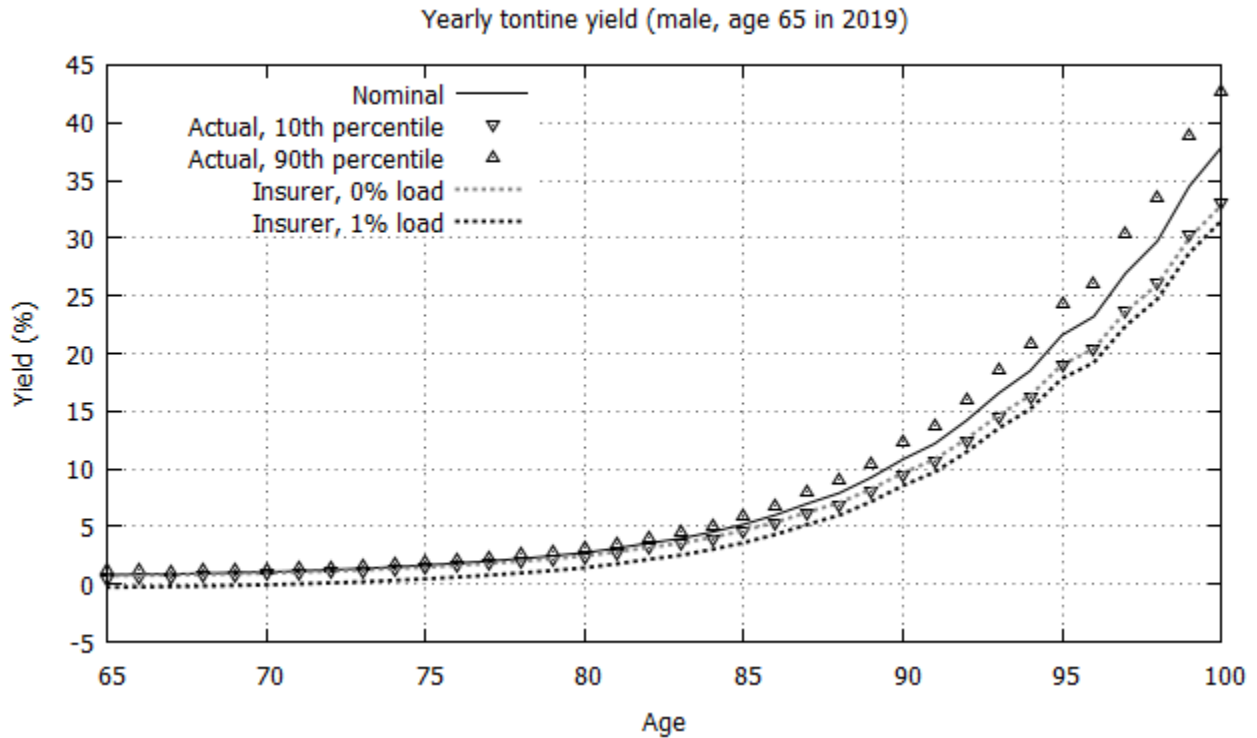
The plot shows the nominal tontine yield along with the 10th and 90th percentiles of the simulated actual tontine yield, from Figure 5. The curves labeled with the prefix "Insurer"

---

[19] Stated another way, the value $\hat{r}$ would be equivalent to the nominal tontine yield r of an ITA if the ITA had been designed using the IAR table instead of the IAM table.

illustrate the mortality yield he would get from an insurer offering a VIA. These are explained next.

Insurer, 0% load. This is the yearly mortality yield an insurer would pay if 1) the insurer calculates the annuity payment using the IAR table, and 2) the insurer charges no load (meaning no mortality/expense fee, no administrative fee, etc.). In other words, it is the mortality yield of a fair annuity where fairness is based on the IAR table. Theoretically, this represents a best-case example of what an individual might hope to buy.

Quite coincidentally, the yield curve for this zero-load annuity is a very close match to the 10th percentile actual yield in our tontine simulation. That is, in 90% of the simulation runs, the member did better by participating in the tontine than he would have by purchasing a VIA even if the insurer offered a 0% load. The exactness of the match is a mere coincidence of our simulation -- had we simulated a larger (or smaller) population, we would have found more (or less) than 90% of the runs to favor the tontine.

Insurer, 1% load. This is the yearly yield an insurer would pay if it added a 1% expense load to the VIA. Not surprisingly, the yield is less than the zero-load case by about 1%. The difference is a function of both the expense load and the mortality yield $\hat{r}$, meaning that it increases with age. See Appendix B for discussion.

For a closer look, Figure 13 illustrates these yields relative to the nominal tontine yield.

Figure 13: Yearly Tontine/Mortality Yield, Relative to Nominal, by Age
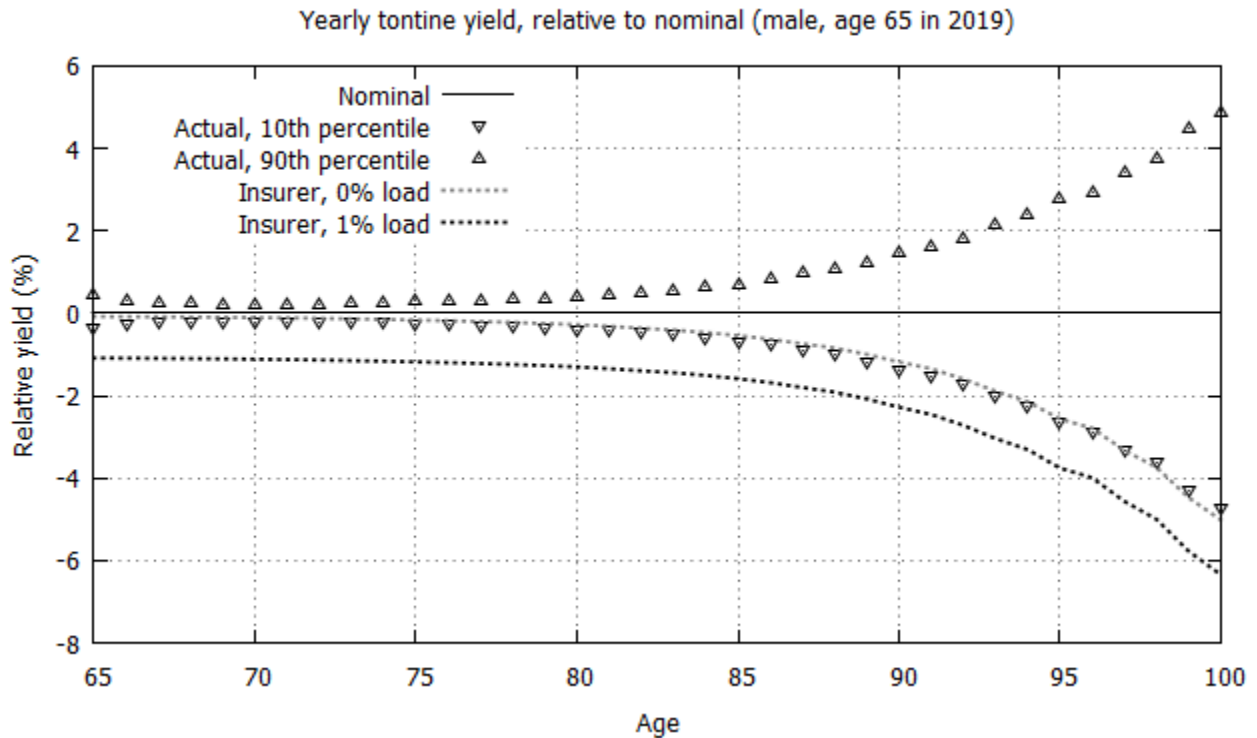

Relative to the nominal tontine yield, the insurer annuity offers a significantly lower yield at advanced ages, even with zero load. This is because the IAR table used by the insurer has lower mortality rates than the IAM table used by the ITA. The gap between the insurer yield and the nominal tontine yield represents the insurer's reserve and profit margin. Note that for ages greater than about 90, the zero-load annuity provides a margin greater than 1%; thus, for such advanced ages, the conservative rates of the IAR table contribute more to the insurer's margin than does a 1% load. Of course, the insurer bears the risk that actual mortality will be less favorable than the IAM table predicts, and the reserve and profit margin is a buffer to absorb losses that might arise from unfavorable systematic mortality outcomes.

Annualized Yield

Figure 14 represents the annualized yield for the male cohort who turns 65 in the year 2019.

Figure 14: Annualized Yield, by Age

Annualized tontine yield (male, age 65 in 2019, join in 2019)
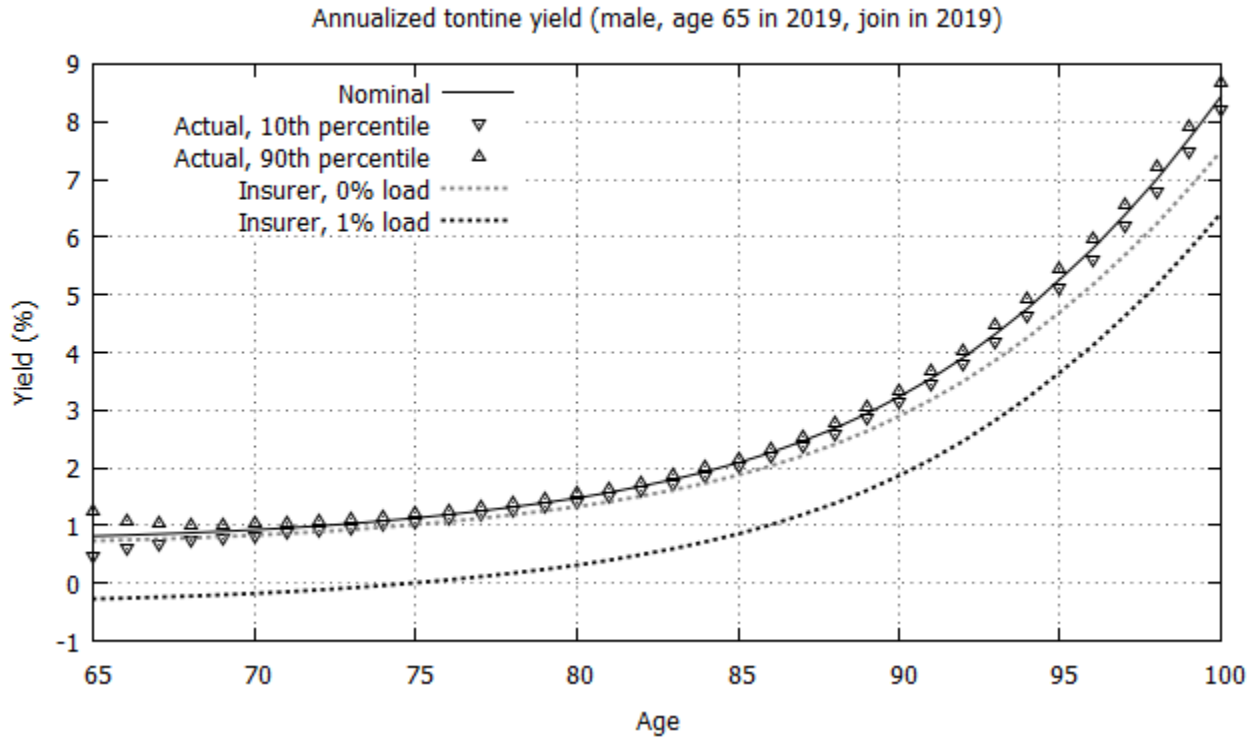
Recall that the annualized yield is an average of the yearly yields, and therefore the annualized yield tends to deviate less from nominal than do the individual yearly yields. The plot illustrates that the tontine is very likely to deliver a higher yield, and therefore a higher payout. The likelihood of this performance advantage increases with age.

For a closer look, Figure 15 illustrates the annualized yield relative to nominal for this cohort.

Figure 15: Annualized Yield, Relative to Nominal, by Age
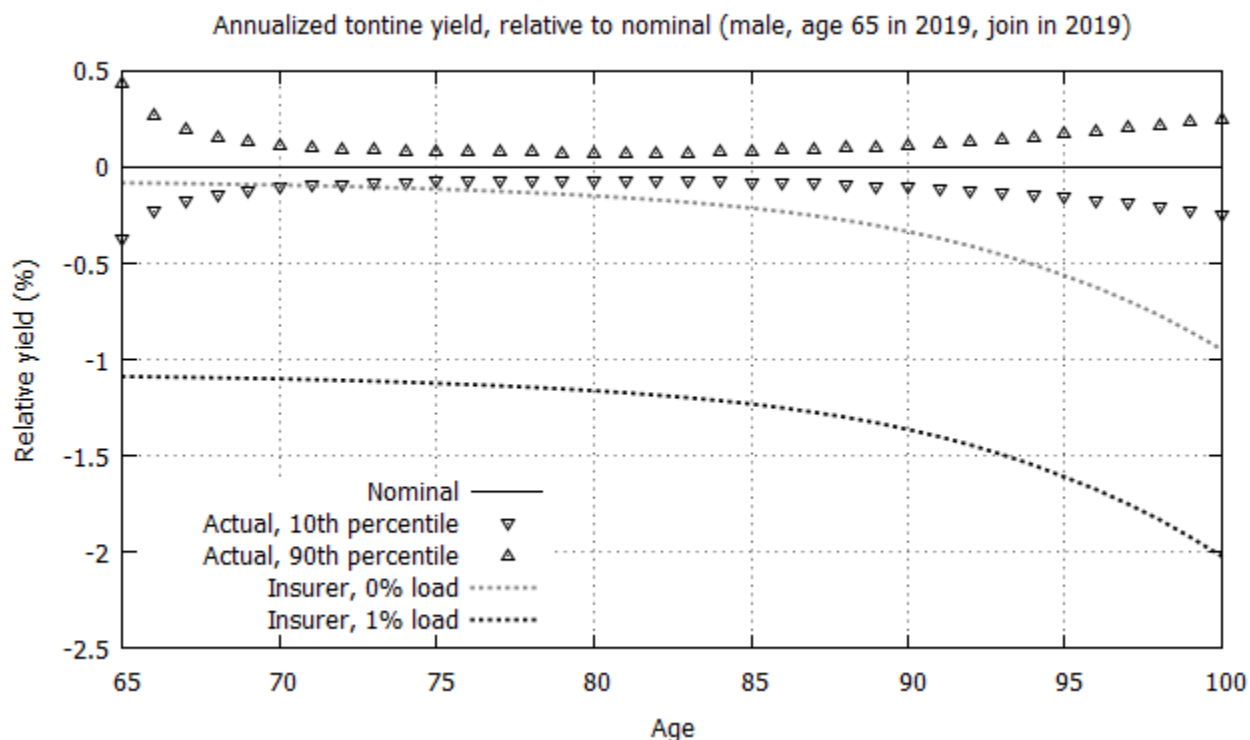
This plot reveals that a long-lived person is virtually certain to do better in the tontine simulation than in either insurer annuity. This effect is important because no one buys a lifetime payout stream with the expectation of dying soon. Life annuities are protection against living a long time. For those who want longevity protection, the simulation suggests that tontines are a more efficient vehicle.

ITAs effectively diversify idiosyncratic mortality risk and eliminate the cost of protecting against systematic mortality risk in the aggregate. To be sure, ITA members bear the systematic mortality risk themselves collectively, whereas annuity purchasers do not. Annuitants sacrifice a significant yield as the price for transferring the systematic component of mortality risk to an insurer rather than bearing it themselves.

We do not mean to overstate the likelihood that an ITA will outperform a similarly-structured VIA. Our tontine simulations measure only the idiosyncratic component of mortality risk. A more complete analysis of the probability distribution of the performance differences between ITAs and VIAs would include a model of systematic mortality risk as well. That is a subject for future research.

The Effect of Systematic Mortality Risk

The subject of systematic mortality risk is an important one. Although modeling it is a challenge that we choose not to tackle here, it is worthwhile nevertheless to show how

ITAs behave in the presence of systematic mortality risk. An analytical discussion of the topic is presented in Sabin and Forman (2016). Here, we illustrate it by rerunning our simulation using the same assumptions we made originally except for the rate at which individuals actually die. As before, the ITA is designed using the IAM table with projection scale G2, meaning that the nominal tontine yields are computed exactly as before, derived from IAM table rates. This time, however, the year of death for each simulated member is drawn from the IAR table with projection scale G2. Since mortality rates are lower for the IAR table due to its reserve margin, the effect is that ITA members die at a slower rate than expected.

Group gain. Since the IAR mortality rates used to model actual death rates are only about 0.9 times the IAM mortality rates used to design the ITA, we would guess the mean value of the group gain to now be about 0.9, compared to 1.0 in the original case. Figure 16 reveals that this is indeed the case.

Figure 16: Group Gain When Mortality Evolves According to IAR Table Rates
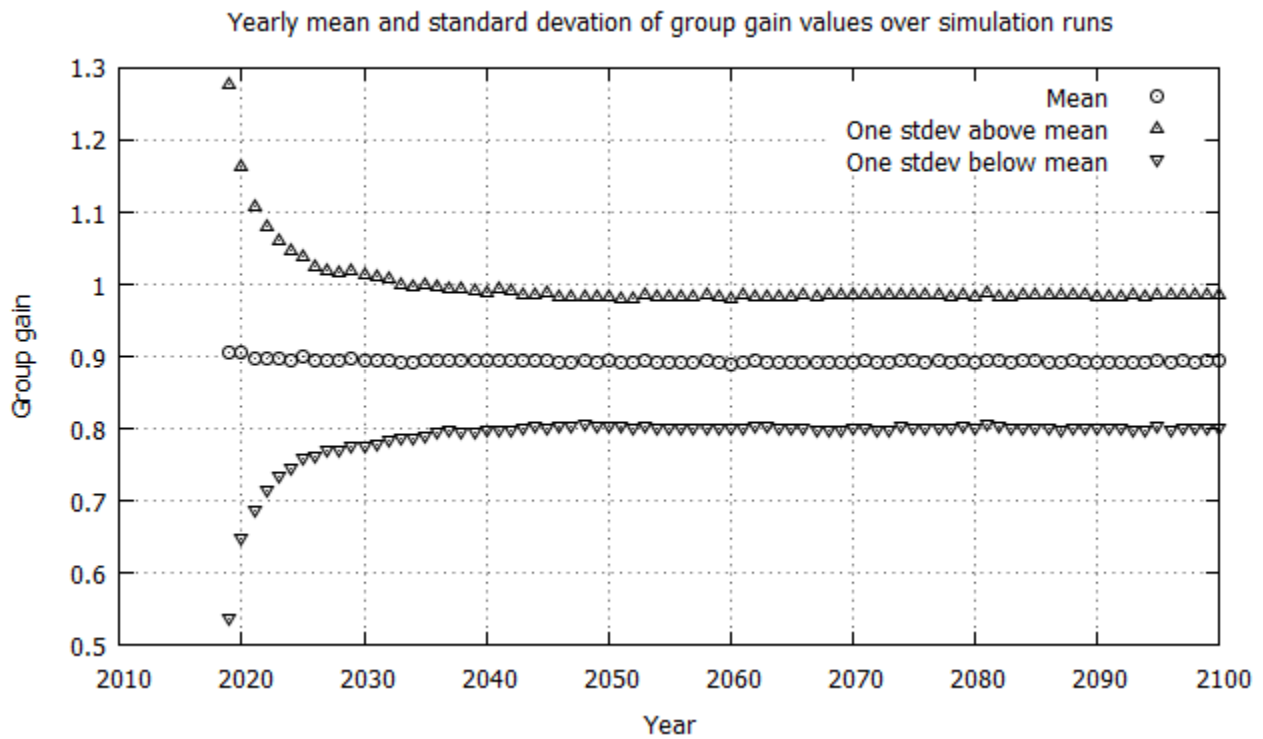
This result can be explained by examining the group gain formula (1). If members die at a rate that is 0.9 times the rate that was anticipated, the expected value of the numerator will decrease by a factor of about 0.9, while the expected value of the denominator will not change much. The standard deviation of the group gain is largely unchanged, having decreased only slightly. See Sabin and Forman (2016) for a more thorough discussion of these statistics.

Tontine yield. Because the membership in aggregate lives longer than expected, the actual tontine yield that an individual member experiences in any given year (based on IAR) is likely to be less than the nominal yield that the member hopes for (based on IAM). The member's actual tontine yield in year of age x is $r_x G_x$, where r_x is the member's nominal yield and G_x is the tontine's group gain for that year. Since the average group gain is about 0.9, the average actual yield is about 0.9 times nominal. Since the nominal gain increases with age, the yield relative to nominal, $r_x(G_x - 1)$, becomes (on average) more negative with age.

Figure 17 and Figure 18 illustrate this for members in the age-65 male cohort who join in 2019.

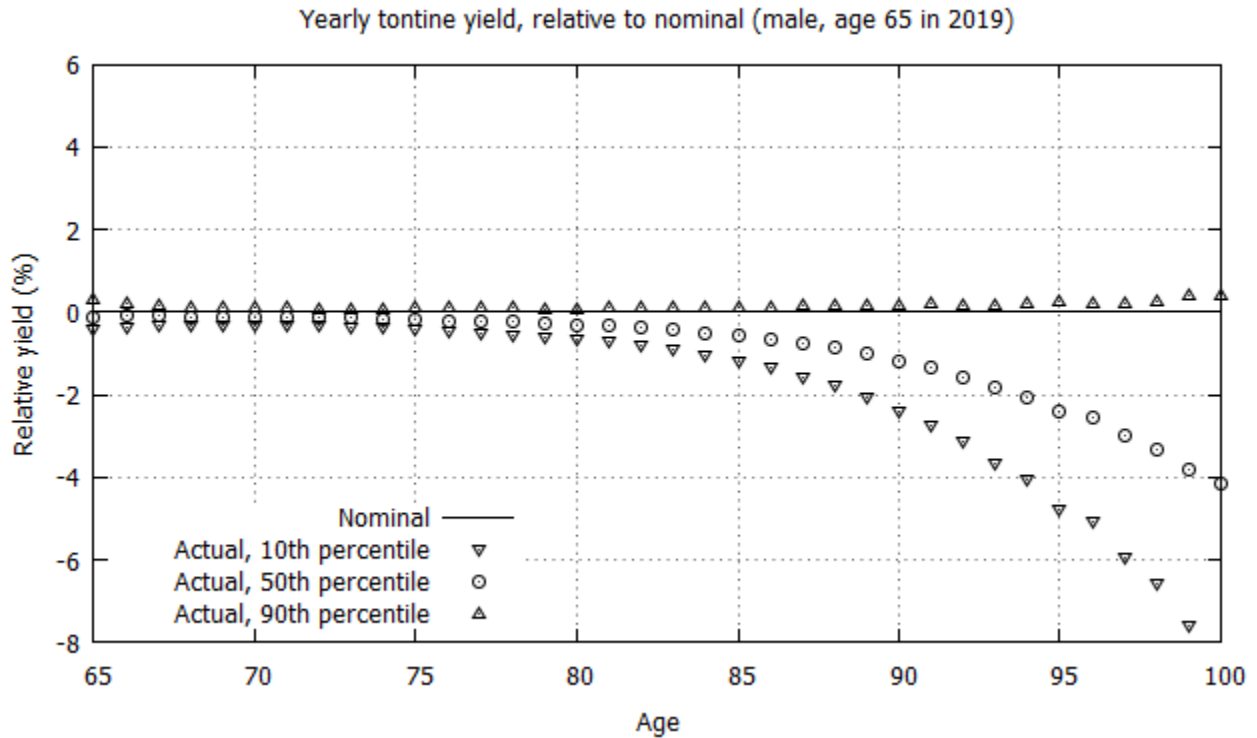**Figure 17: Yearly Tontine Yield, Relative to Nominal, by Age**

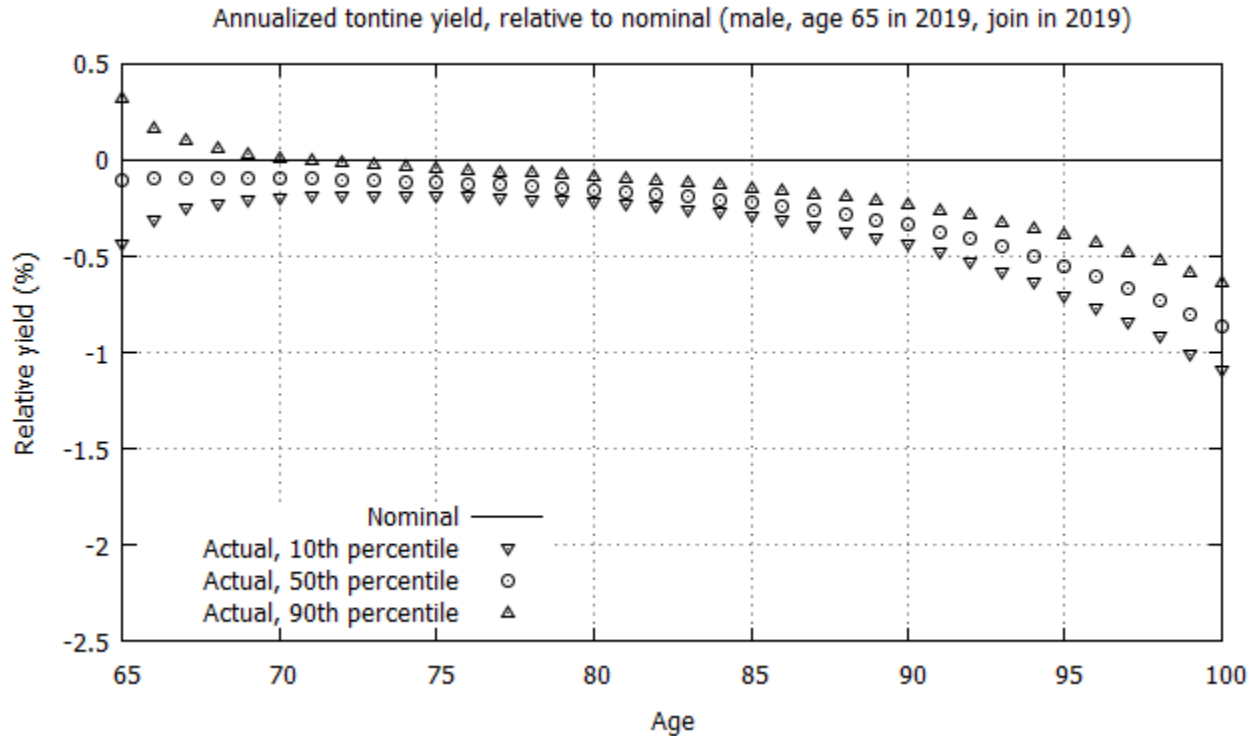**Figure 18: Annualized Tontine Yield, Relative to Nominal, by Age**

Figure 18 shows that a long-lived member in this cohort is virtually certain to do worse than nominal in this simulation. This is not surprising, since on the average his actual yearly yield is only 0.9 times nominal, so over a long lifetime the actual annualized yield will tend close to 0.9 times nominal.

It is noteworthy that members of this cohort experienced an average yearly yield similar to that of a zero-load VIA, which is a fair annuity based on the IAR mortality rates. This can be seen by comparing the 50th percentile curve in Figure 17 to the zero-load curve in Figure 13, and by similarly comparing Figure 18 to Figure 15. This is what we would hope – even though there is a mismatch between the mortality rates assumed in the tontine design and the true mortality rates that the tontine experienced, the ITA behaved for this cohort as if it were designed using the true mortality rates.

Unfortunately, this will not happen in general. Instead, the ITA will behave more favorably for some cohorts than it would have if designed with the true mortality rates, and less favorably for other cohorts, according to how the true mortality rates for each cohort differ from the design mortality rates. For a discussion of this effect, see Sabin and Forman (2016). Nevertheless, we believe that all members will perceive the tontine as fair. "True" mortality rates are theoretical parameters unseen by members, not known by even the provider until after-the-fact (if then). What members do see is that they receive lower tontine yields than they had hoped for compared to the nominal yields that the provider

41

had projected. Though they are undoubtedly disappointed, we believe they will perceive the tontine as fair because everyone's yields are reduced by the same percentage and the calculations that determine yields are fully transparent.

Conclusion

This paper illustrates in a practical way what has previously been theorized: that fair tontines are not bounded by any restriction that members must invest in a common portfolio or enroll in the same payout option.

Individual tontine accounts represent an attractive alternative solution to the retirement income problem. They operate very much like IRA brokerage accounts, but with the added benefit of providing a nonnegative and uncorrelated tontine yield on top of a member's underlying investment returns. Individuals can invest and trade when and as they choose. They may also choose from a wide array of payout contracts based on their particular preferences and needs. ITAs are fair to all members regardless of when they open their accounts and regardless of the demographics, balances, investment choices, and payout choices of the other members.

Another benefit of ITAs lies in their cost efficiency and transparency. Fees are plainly disclosed and all-in costs to account holders should be very low when low-cost investments are selected. The accounting is simple to report, and member statements are easy to understand.

Economists and public policy makers have long pondered the so-called annuity puzzle; namely, why do so few people annuitize when it seems to be in their interest to do so? To the extent that the answer involves the perceived high costs and lack of transparency of annuity products, ITAs represent an attractive remedy to these problems. Because ITAs pool and diversify idiosyncratic mortality risk, they can offer payout options similar to insured annuity products, but at lower cost since the tontine structure obviates the added costs of guarantees. Although (and because) ITA members collectively bear systematic mortality risk whereas annuity buyers do not, ITAs are likely (but not guaranteed) to deliver higher yields.

ITAs give retirees a low-cost way to derive extra income from their savings without taking on additional investment risk. Of course, since account holders cannot withdraw freely from their accounts whenever they wish but rather only per a payout schedule selected at the time of contribution, ITAs are not a complete replacement to traditional IRAs. But they could be a very useful complement, and one with unique benefits not otherwise available from traditional investment and annuity products.

References

Donnelly, C. (2015). Actuarial fairness and solidarity in pooled annuity funds. *ASTIN Bulletin*, 45(01):49–74.

Donnelly, C., Guillén, M., and Nielsen, J. P. (2013). Exchanging uncertain mortality for a cost. *Insurance: Mathematics and Economics*, 52(1):65–76.

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

Forman, J. B. and Sabin, M. J. (2014). Tontine pensions: A solution to the state and local pension underfunding crisis. *University of Pennsylvania Law Review*, 163(3).

Forman, J. B. and Sabin, M. J. (2016). Survivor funds. *Pace Law Review*, 37(1). https://digitalcommons.pace.edu/plr/vol37/iss1.

Gründel, H. and Wandt, M. (July 6, 2017). The modern tontine: An innovative instrument for longevity risk management in an aging society. ICIR Working Paper Series No. 22/2016.

Kantorowicz, E. H. (1957). *The King's Two Bodies: A Study in Mediaeval Political Theology*, Princeton University Press, Princeton, NJ, 411-412.

Milevsky, M. A. and Salisbury, T. S. (2015). Optimal Retirement Income Tontines. *Insurance: Mathematics and Economics*, 64:91–105.

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

NAIC (2013). *NAIC model rule (regulation) for recognizing a new annuity mortality table for use in determining reserve liabilities for annuities*. National Association of Insurance Commissioners. Available at http://www.naic.org/store/free/MDL-821.pdf, last accessed 9 Nov 2016.

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

Sabin, M. J. (March 26, 2010). Fair tontine annuity. Available at SSRN: https://ssrn.com/abstract=1579932 or https://dx.doi.org/10.2139/ssrn.1579932.

Sabin, M. J. and Forman, J. B. (November 9, 2016). The analytics of a single-period tontine. Available at SSRN: https://ssrn.com/abstract=2874160 or https://dx.doi.org/10.2139/ssrn.2874160.

Stamos, M. Z. (2008) Optimal consumption and portfolio choice for pooled annuity funds. Insurance: Mathematics and Economics, 43(1):56–68.

Appendix A: Annualized Tontine Yield

Consider a member who joins the tontine with an initial contribution of s_0 dollars. If she survives to the end of the first year, her balance at the end of the year, prior to her payout, is:

$$S_1 = s_0(1 + X_1)(1 + r_1 G_1),$$

where X_1 is the net investment return earned by her portfolio during the year, r_1 is her nominal tontine yield, and G_1 is the tontine's group gain for the year. Her payout is $D_1 = d_1 S_1$, where d_1 is her payout rate, which is some value between 0 and 1 that depends on her payout contract. For example, if she has contracted for an annuity payout, then $d_1 = 1/a_1$, where a_1 is her annuity factor for the first year. Or if she has contracted for a lump-payout, then $d_1 = 0$ if the contract term is greater than one year, or $d_1 = 1$ if the contract term is one year. Her balance after the payout is $S_1 - D_1 = S_1(1 - d_1)$. This is her balance at the start of the second year.

If she survives to the end of the second year, her balance at the end of the year, prior to her payout, is:

$$\begin{aligned} S_2 &= S_1(1 - d_1)(1 + X_2)(1 + r_2 G_2) \\ &= s_0(1 - d_1)(1 + X_1)(1 + X_2)(1 + r_1 G_1)(1 + r_2 G_2). \end{aligned}$$

Her payout is $D_2 = d_2 S_2$, and the balance after her payout is $S_2(1 - d_2)$.

More generally, if she survives to the end of year n, her balance at the end of the year, prior to her payout, is

$$S_n = S_{n-1}(1 - d_{n-1})(1 + X_n)(1 + r_n G_n) \tag{3}$$

$$= s_0 \prod_{t=1}^{n-1}(1 - d_t) \prod_{t=1}^{n}(1 - X_t) \prod_{t=1}^{n}(1 + r_t G_t). \tag{4}$$

Her payout is $D_n = d_n S_n$ and the balance after her payout is $S_n(1 - d_n)$. For completeness we define $S_0 = s_0$ and $d_0 = 0$ so that formula (3) holds for the first year $n = 1$ as well as for subsequent years.

Formula (3) is the yearly version, showing how the current balance S_n evolves from the previous year's balance S_{n-1}. Formula (4) is the cumulative version, showing how the current balance S_n evolves from the starting contribution s_0. We consider each of these separately.

The payout rate d_n is related to the balance S_n by the formula $D_n = d_n S_n$. The quantity d_n is a parameter of the contract that specifies what percentage of the member's balance gets paid out in year n. It is a parameter that does not depend on the balance. Thus, it does not matter much whether we study the payout or the balance, because the results for one will apply immediately to the other. We choose to study the payout.

Regarding the d_n parameters, we have considered two possible contracts, lump sum and annuity. However, the analysis here applies more generally to any contract which can be specified in terms of d_n parameters. For example, it would apply to an N-year contract that each year pays in inverse proportion to the remaining number of years, by specifying $d_1 = 1/N$, $d_2 = 1/(N-1)$, ..., $d_{N-1} = 1/2$, $d_N = 1$. These are just examples; many other contracts are possible under the d_n-parameter framework.

Yearly tontine gain. Referring to (3), we can write the yearly formula for the nth payout $D_n = d_n S_n$ in a way that separates the effect of investment return from tontine yield, as follows:

$$D'_n = d_n(1 - d_{n-1})S_{n-1}(1 + X_n)$$
$$D_n = D'_n(1 + r_n G_n).$$

D'_n is the payout a member would receive for the year if he were not participating in the tontine, assuming he invested in the same portfolio (same X_n), had the same prior balance (same S_{n-1}), and withdrew under the same terms as the payout contract (same d_n, d_{n-1}). By participating in the tontine, and surviving, he boosts his payout by the factor $1 + r_n G_n$. We call this factor his yearly tontine gain.

Cumulative tontine gain. Referring to (4), we can write the cumulative formula for the payout in year n in a way that separates the effect of investment return from tontine yield, as follows:

$$D'_n = s_0 \prod_{t=1}^{n-1}(1 - d_t) \prod_{t=1}^{n}(1 + X_t)$$
$$D_n = D'_n \prod_{t=1}^{n}(1 + r_t G_t).$$

D'_n is the payout the member would receive if she were not participating in the tontine, assuming she invested in the same portfolio (same X_t values), with the same initial contribution (same s_0), and adhered to the same payout contract (same d_t values). By participating in the tontine, she boosts her payout in year n by the factor $\prod_t(1 + r_t G_t)$. We call this factor her cumulative tontine gain for year n. It is the product of her yearly tontine gains for each of the first n years. If the group gain G_t equals 1 for each of those n years, then her cumulative tontine gain equals its nominal value of:

$$\prod_{t=1}^{n}(1 + r_t) = \prod_{t=1}^{n}\frac{1}{1 - q_t} = \frac{1}{1 - q_{1:n}},$$

where $q_{1:n} = 1 - \prod_t(1 - q_t)$ is her probability of dying during the first n years. In general, of course, G_t does not equal 1, and thus her cumulative gain differs from its nominal value. Since each G_t is non-negative, the cumulative gain is never less than 1.

Annualized tontine yield. We can express the cumulative tontine gain in terms of an annualized yield by using the geometric means of the yearly tontine gains, as follows:

$$Y_n = \prod_{t=1}^{n} (1 + r_t G_t)^{1/n} - 1.$$

We call Y_n the member's *annualized tontine yield* for year n. The annualized tontine yield is the average value of the yearly tontine yields, in the sense that if each yearly tontine yield for the first n years had value Y_n, then the payout in year n would equal D_n; that is, $D'_n (1 + Y_n)^n = D_n$.

Appendix B: Computing the Net Mortality Yield

Suppose that the end-of-year balance of a VIA before the deduction of fees and before making its annual payout is s. The insurer first deducts fees in the amount of es, where e is the expense load in percentage terms. The insurer then credits a mortality yield of $\hat{r}$ on the remaining balance $s(1 - e)$, where $\hat{r}$ is the nominal mortality yield calculated using a mortality table that includes a reserve margin (i.e., the IAR table) since the insurer must price in the systematic mortality risk that it takes on in offering these products. The net mortality gain to the annuitant is $\hat{r}s(1 - e) - es = s(\hat{r} - e(1 + \hat{r}))$, and thus the net morality yield, net of fees, is $\hat{r} - e\,(1 + \hat{r})$.

The difference between the two insurer curves shown in Figure 13, in which the expense load is 0% and 1% respectively, is $(\hat{r} - 0(1 + \hat{r})) - (\hat{r} - 0.01(1 + \hat{r})) = 0.01(1 + \hat{r})$.