Decumulation of Retirement Savings: *Are Modern Tontines the Solution?*[1]

Peter Forsyth[1] Ken Vetzal[2] Graham Westmacott[3]

[1]Cheriton School of Computer Science
University of Waterloo

[2]School of Accounting and Finance
University of Waterloo

[3]Richardson Wealth
Waterloo

ICCF
Wednesday April 3 16:30

[1]Gold Open Access: *"Optimal performance of a tontine overlay subject to withdrawal constraints,"* ASTIN Bulletin (2024)

Motivation

Defined Benefit Plans (DB) are disappearing

→ Corporations/governments no longer willing to take risk of DB plans

Recent survey[2] P7 countries[3]

- Defined Contribution (DC)[4] plan assets: 55% of all pension assets
- Some examples
  - → Australia 87% DC
  - → US 65% DC
  - → Canada 43% DC
  - → ...
  - → Japan 5% DC

Netherlands → *Collective* DC plan (2027)[5]

[2] Thinking Ahead Institute (2023)

[3] Australia, Canada, Japan, Netherlands, Switzerland, UK, US

[4] DC plan: retiree takes on all investment risk

[5] See *"Can DC participants trust the competence of Dutch pension funds,"* working paper, Georgetown University

The retiree dilemma (Defined Contribution (DC))

A retiree with savings in a DC plan[6] [7] has to decide on

- An investment strategy (stocks vs. bonds)
- A decumulation schedule

The retiree now has two major sources of risk

- Investment risk
- Longevity risk (running out of cash before death)

William Sharpe (Nobel Laureate in Economics) calls this

"The nastiest hardest problem in finance"

---

[6] In a DC plan, the retiree is responsible for investment/decumulation

[7] RRSP (Canada), SIPP (UK), 401(k)(US), Super Fund (Australia)

The Four per Cent Rule

Based on rolling 30-year historical periods, Bengen (1994) showed:

A retiree who

- Invested in a portfolio of 50% bonds, 50% stocks (US), rebalanced annually
- Withdrew 4% of initial capital (adjusted for inflation) annually

$\rightarrow$ Would never have run out of cash, over any rolling 30-year period (from 1926)

Criticism

- Simplistic asset allocation strategy
- Simplistic withdrawal strategy
- Rolling 30 year periods contain large overlaps

  $\rightarrow$ Underestimates risk of portfolio depletion

Bengen rule

> *"Play the long game. A retirement income plan should be based on planning to live, not planning to die. A long life will be expensive to support, and it should take precedence over death planning." Pfau (2018)*

Note that Bengen rule is based on assumption that 65-year old will live to be 95

- Should we mortality weight the cash flows (as in an annuity)?
- Example: median life expectancy of 65-year old male $\simeq$ 87.
  - $\rightarrow$ Effectively, mortality weighting will weight minimum cash flow of 87-year old by $1/2$
  - $\rightarrow$ If I am 87, and alive, I need 100% of my minimum cash flows
  - $\rightarrow$ If I am dead, I need zero dollars
- We will consider an individual investor, not averaging over a population
  - $\rightarrow$ 30 year retirement, no mortality weighting
  - $\rightarrow$ Consistent with Bengen approach

Fear of running out of cash

Recent survey[8]

- Majority of pre-retirees fear exhausting their savings in retirement more than death

In Canada, a 65-year old male

- Probability of 0.13 of living to be 95
- Probability of 0.02 of living to be 100

Conservative strategy:

$\rightarrow$ Assume 30 year retirement (as in Bengen (1994)).

Other assets can be used to hedge extreme longevity[9]

[8] 2017 Allianz Generations Ahead Study - Quick Facts #1. (2017), Allianz

[9] Real estate

Objective of this talk

Determine a decumulation strategy which has

- Variable withdrawals (minimum and maximum constraints)
- Minimizes risk of portfolio depletion
- Maximizes total expected withdrawals
- Allows for dynamic, non-deterministic asset allocation
- Pool longevity risk using a Modern Tontine overlay

We will treat this as a problem in optimal stochastic control

Modern Tontines (Individual Tontine Account)

DC members make irrevocable investment in a pooled fund

- If the member dies during a year, their assets distributed to the other members as longevity credits
- The sharing rule is actuarially fair, i.e. expected gain from participating is zero
  - If you are older or have more assets
    - → You get a larger share of longevity credits

Advantage:

- Transparent, peer-to-peer risk sharing: DeFi[10]
- Can decide your own investment strategy
- Expected withdrawals larger than a conventional TradFi[11] product
  - → Retiree bears investment risk, systematic mortality risk
  - → Assets forfeited on death (as in conventional DB plan, annuity)

---

[10] Decentralized Finance

[11] Traditional Finance

Longevity Credits: Example

CPM2014 Life table: theoretical longevity credit

- Yearly credit for 76-year old male: 2%
- Yearly credit for 86-year old male: 8%
- Yearly credit for 96-year old male: 33%

Example:

- 85 year-old, living member of pool on January 1, 2024
- Total wealth W in account (December 31,2024)
- If he is still alive on January 1, 2025 (now 86 year old)
  - $\rightarrow$ He will earn longevity credit of $0.08W$

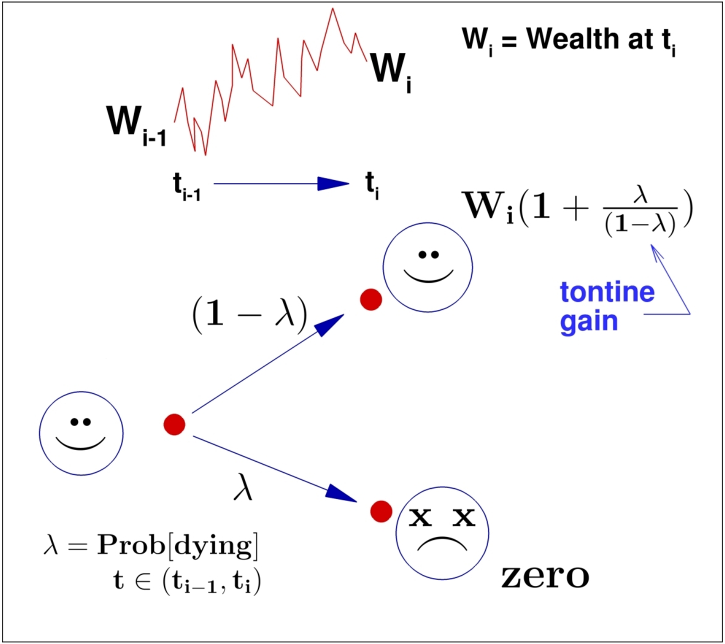
Theoretical credit depends only on

- Your age and your account balance

Does not depend on how anyone else invests, their age, or their account balances![12].

[12] This is a counterintuitive result, see R. Fullmer, Tontines: a practitioner's guide, CFA Institute (2019), and references therein

Tontines at a glance



How does this work in practice?

Define for each year:

$$\left\{ \begin{array}{l} \textit{Group} \\ \textit{Gain} \end{array} \right\} = G \quad = \quad \frac{\text{Total actual assets forfeited due to deaths in pool}}{\text{Total expected longevity credits for survivors}}$$

$$\begin{aligned} &\text{Actual longevity credit for each investor} \\ &= \text{Theoretical Credit} \times G \end{aligned}$$

This ensures that total longevity credits handed out

$\rightarrow$ Equals total assets forfeited

Can show that $E[G] = 1$, $Var[G] \rightarrow 0$ if

- Pool is sufficiently large
- Diversity condition holds [13]

[13] The expected total longevity credits must be large compared to any members expected credit. Simulations: perpetual pool size $\simeq$ 5,000-15,000.

Formulation

Investor has access to two funds

- A broad stock market index fund
  - *Amount* in stock index S_t
- A constant maturity bond index fund
  - *Amount* in bond index B_t

$$\text{Total Wealth } W_t \;=\; S_t + B_t \tag{1}$$

Model the returns of both indexes

- Parametric, jump diffusion
- Non-zero stock-bond correlation

• Fit parameters to market data 1926:1-2020:12
$\hookrightarrow$ All returns adjusted for inflation

Notation

Withdraw/rebalance at discrete times $t_i \in [0, T]$
The investor has two controls at each rebalancing time

$$
\begin{aligned}
q_i &= \text{Amount of withdrawal} \\
p_i &= \text{Fraction in stocks after withdrawal}
\end{aligned}
$$

$$
\begin{aligned}
W_i^- &= \text{wealth after tontine gains and fees} \\
&\quad \text{before withdrawals} \\
&= \underbrace{(S_i^- + B_i^-)}_{\textit{Before gains/fees}} \underbrace{(1 + \text{tontine gain})(1 - \text{fees})}_{\textit{tontine gains and fees}}
\end{aligned}
$$

At t_i^+, the investor withdraws q_i

$$
W_i^+ = W_i^- - q_i
$$

Rebalancing

Recall that

$$
\begin{aligned}
W_i^- &= \text{wealth after tontine gains and fees} \\
&\quad\ \text{before withdrawals} \\
W_i^+ &= \text{wealth after withdrawals}
\end{aligned}
$$

Then, the investor rebalances the portfolio

$$
\begin{aligned}
S_i^+ &= p_i W_i^+ \\
B_i^+ &= (1 - p_i) W_i^+
\end{aligned}
$$

Controls

Constraints on controls

$$
\begin{aligned}
q_i &\in [q_{\min}, q_{\max}] \quad ; \quad \text{withdrawal amount} \\
p_i &\in [0, 1] \quad ; \quad \text{fraction in stocks} \\
&\Rightarrow \text{ no shorting, no leverage}
\end{aligned}
$$

Set of controls

$$
\mathcal{P} = \{(\ q_i(\cdot), p_i(\cdot)\ ) : i = 0, \ldots, M\} \tag{2}
$$

Reward and Risk

Reward: Expected total (real) withdrawals (EW)

$$\text{EW} = E\Big[\overbrace{\sum_i q_i}^{\textit{total withdrawals}} \Big]$$

$$E[\cdot] = \text{ Expectation}$$

Risk measure: Expected Shortfall *ES*

$$ES(5\%) \equiv \Big\{ \text{Mean of worst 5\% of } W_T \Big\}$$

$$W_T = \text{ terminal wealth at } t = T$$

ES defined in terms of final wealth, *not losses*[14]

→ Larger is better

[14] ES is basically the negative of CVAR

Objective Function

Multi-objective problem $\rightarrow$ scalarization approach for Pareto points

Find controls $\mathcal{P}$ which maximize (scalarization parameter $\kappa > 0$)[15]

$$\sup_{\mathcal{P}} \left\{ EW + \kappa \; ES \right\}$$

$$\sup_{\mathcal{P}} \left\{ \overbrace{E_{\mathcal{P}}[\sum_i q_i]}^{total \;\; withdrawals} + \kappa \overbrace{\left(\frac{E_{\mathcal{P}}[W_T \; 1_{W_T \leq W^*}]}{.05} \right)}^{mean \;\; worst \;\; 5\% \;\; outcomes} \right\}$$

$$\text{s.t.} \quad Prob[W_T \leq W^*] = .05$$

Varying κ traces out the efficient frontier in the (EW, ES) plane

[15] $E_{\mathcal{P}}[\cdot] \equiv$ expectation under control $\mathcal{P}$.

EW-ES Objective Function

Given an expectation under control $E_{\mathcal{P}}[\cdot]$ (Rockafellar and Uryasev, 2000)

$$\text{ES}_{5\%} = \sup_{W^*} E_{\mathcal{P}}\Big[H(W_T, W^*)\Big]$$

$$H(W_T, W^*) = \left(W^* + \frac{1}{.05}\left[\min(W_T - W^*, 0)\right]\right)$$

Reformulate objective function:

$$\sup_{\mathcal{P}} \sup_{W^*} E_{\mathcal{P}}\left\{ \overbrace{\sum_i q_i}^{total\ \ withdrawals} + \kappa \overbrace{H(W_T, W^*)}^{mean\ \ worst\ \ 5\%\ \ W_T} \right\}$$

Under above assumptions: can show that[16]

$$q_i = q_i(W_i^-) \quad ; \quad p_i = p_i(W_i^+)$$

[16] q_i withdrawal, p_i fraction in stocks, $W^{\pm}$ wealth before/after withdrawals

Time Consistency

The EW-ES objective function is not formally *time consistent*

Time inconsistency

$\Rightarrow$ Investor has incentive to deviate from initial optimal policy at later times

EW-ES policy computed at time zero
$\hookrightarrow$ Pre-commitment policy

Induced time consistent policy

At t_0 we compute the pre-commitment EW-ES control

- For $t > t_0$ we assume that the investor follows the *induced time consistent* control (Strub et al (2019))
- This control is identical to the pre-commitment control at t_0
- No incentive to deviate from this control at $t > t_0$

Induced time consistent control determined from (fixed W^*)

$$\sup_{\mathcal{P}} \; E_{\mathcal{P}} \left\{ \sum_i q_i + \kappa H(W_T, W^*) \right\}$$

W^* from pre-commitment solution at time zero

Alternative: equilibrium mean-ES control
$\hookrightarrow$ Does not actually control tail risk! (Forsyth(2020)) [17]

[17] For more discussion of time consistency, induced time consistency, pre-commitment, see Bjork et al (2021), Vigna (2020, 2022), Strub et al (2019), Forsyth (2020)

Scenario: all amounts indexed to inflation

- DC account at $t = 0$ (age 65) \$1,000K (one million)
  - Minimum withdrawal from DC account \$40K per year[18]
  - Maximum withdrawal from DC \$80K per year
  - Fees: 50bps per year
- No shorting, no leverage, annual rebalancing
- Investment Horizon: $T = 30$ years, i.e. from age 65 to 95

  $\hookrightarrow$ Tontine gains: CPM 2014 mortality table
- Assume pool is very large, diverse so that

  $\hookrightarrow$ Group Gain $G \equiv 1$
- Retiree owns mortgage-free real estate worth \$400K

  $\hookrightarrow$ Hedge of last resort (reverse mortgage)

[18] Never less than Bengen rule: $40K/1000K = 4\%$

Scenario II

Why do we include real estate in the scenario?

Since $q_{\min} = 40K$ per year, W_t can become negative

- When $W_t < 0$, assume retiree is borrowing, using a reverse mortgage[19]
  - Reverse mortgages allow borrowing of 50% of home value
  - In our case: $200K
- Once $W_t < 0$
  - All stocks are liquidated
  - Debt accumulates at borrowing rate
- If $W_T > 0$, then real-estate is a bequest
- Real estate is a hedge of last resort: not fungible with other wealth
  - This mental bucketing of real estate is a well-known behavioral finance result.[20]

[19] See Pfeiffer et al, Journal of Financial Planning (2013)

[20] I also observe this with my fellow retirees: real-estate is a separate bucket

Numerical Method I

Pre-commitment control at t_0 (same as induced time consistent control)

Interchange $\sup \sup (\ldots)$

$$\underbrace{\sup_{W^*} \qquad \overbrace{\sup_{\mathcal{P}} \; E_{\mathcal{P}} \left\{ \sum_i q_i + \kappa H(W_T, W^*) \right\}}^{Solve\;\; using\;\; Dynamic\;\; Programming\;\; (fixed\;\; W^*)}}_{maximize\;\; over\;\; W*}$$

Solve inner DP problem using PIDE methods

Numerical Method II

Inner maximization: dynamic programming

- Conditional expectations at t_i^+
  - Solve linear 2-d PIDE
  - Use ϵ-monotone Fourier method (Forsyth and Labahn (2019))
- Optimal controls at each rebalancing time
  - Discretize controls
  - Find maximum by exhaustive search
- Guaranteed to converge to the solution as discretization parameters $\rightarrow 0$

Outer maximization over W^*

- Discretize W^*, use coarse PIDE grid
  - $\rightarrow$ Find optimal W^* by exhaustive search
- Use coarse grid W^* as starting point for 1-d optimization on finer grids

Data

Center for Research in Security Prices (CRSP) US

- Cap weighted index, all stocks on all major US exchanges 1926:1-2020:12
- US 30-day T-bill[21]
- Monthly data, inflation adjusted by CPI

Synthetic Market

- Stock/bond returns driven by parametric jump-diffusion model, calibrated to data
- Optimal controls computed in the *synthetic* market

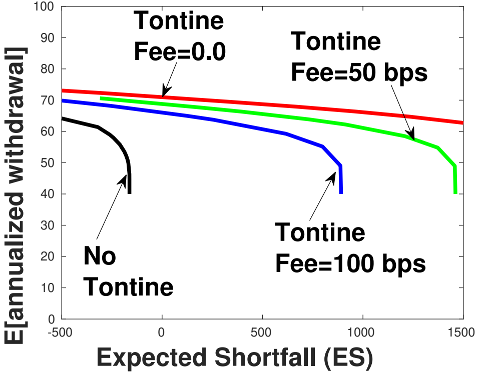
Historical market

- Stock/bond returns from stationary block bootstrap resampling of actual data[22]
- No assumptions about stock/bond processes
- Used to test control robustness computed in the synthetic market

[21] Slightly better results with 10 year Treasuries

[22] Dichtl et al (2016, Appl. Econ.), Anarkulova et al (JFE,2022)

Synthetic Market Results (parametric model)



Efficient frontier: fixed ES
↪ Largest possible expected withdrawal
↪ Pareto optimal points

•Farther to right is better
•Farther up is better

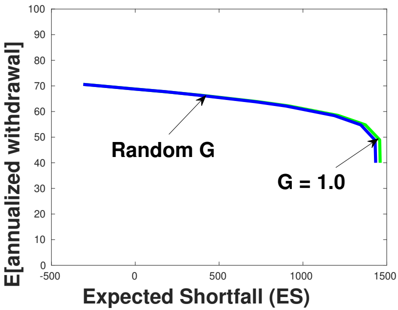
Base case fee 50bps

X-axis: expected shortfall (larger better)

Y-axis: Expected annualized withdrawal (larger better)

No Tontine: Optimal strategy, no Tontine overlay

Effect of Random G (group gain)



Compute control with $G \equiv 1$

$\hookrightarrow$ Test with random G (Monte Carlo simulation)

Random G: simulation of pool with 15K members

Fit to normal distribution with $E[G] = 1$, $std[G] = 0.1$[a]

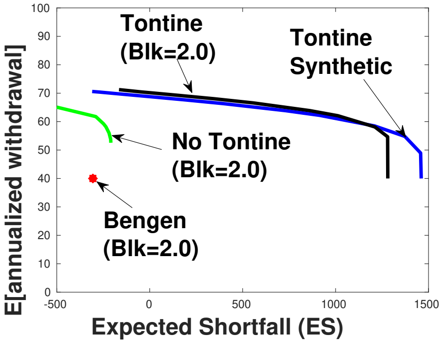
[a] Fullmer and Sabin (2019) CFA white paper

$$\text{Actual Mortality Credit} = \text{Theoretical Credit} \times \overbrace{G}^{\text{Group Gain}}$$

Random G Statistics

- Different ages, genders, investment strategies
- Perpetual pool, random initial wealth

Efficient Frontier: Historical Market (bootstrap resampling)



Synthetic: Control computed using parametric model

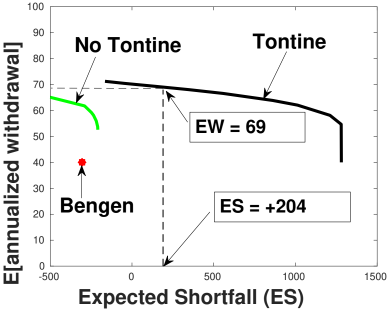
Blk = 2.0: Control tested on block bootstrap resampled historical data
↪ Blocksize = 2yrs
↪ Out-of-sample test

Bengen: 4% rule

No Tontine: Optimal strategy, no Tontine overlay

- X-axis: expected shortfall (larger better)
- Y-axis: Expected annualized withdrawal (larger better)

How bad is the Bengen 4% rule? (Bootstrap simulations)



Bengen:

↪ EW = +40K/year[a]

↪ Expected shortfall = -303K !

Tontine:

↪ EW = +69K/year

↪ Expected shortfall = +204K

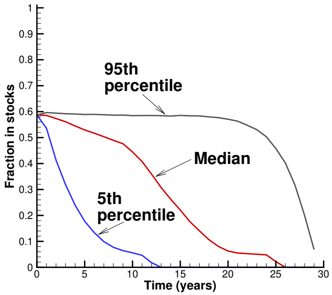
[a] EW = Expected annualized Withdrawals

- Tontine: $(EW, ES) = (69, +204)$
  → Never withdraws less then Bengen
  → Expected withdrawals 6.9%/year[23]
  → ES = +204K at age 95
    - Mortality credit for 95 year old: $\simeq$ 33% per year!

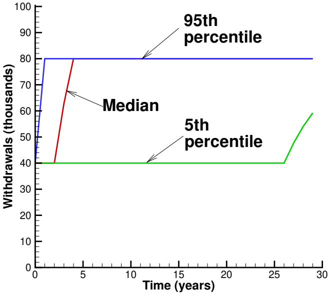
[23] 6.9% of initial capital, adjusted for inflation.

Point on Frontier: (EW,ES) = (69K/year, +204)

Percentiles: fraction in equities



Percentiles: withdrawals

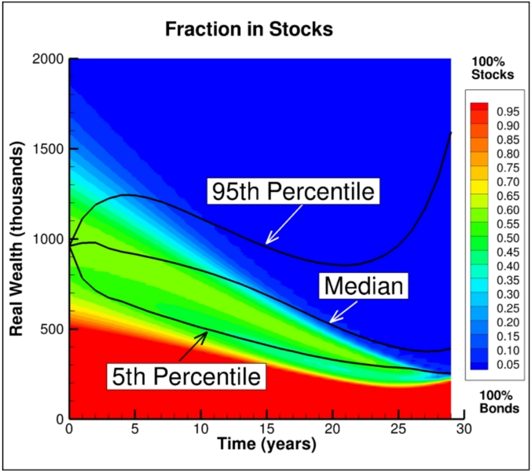


→ $ES \simeq +204$

→ 5th percentile wealth at $t = 30 \simeq +300$K

→ Average withdrawal $\simeq$69K/year

Asset Allocation Heat Map



Point on Frontier: (EW,ES) = (69K/year, +204K)

Blue: 100% bonds
Red: 100% stocks

Optimal fraction in stocks

Function of observed wealth, time

- Over 30 years
  - → Fraction in stocks $\leq$ 0.60 with 95% probability

Conclusions

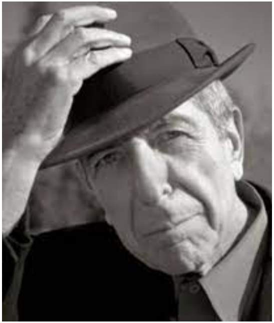
- Tontine overlay: peer to peer longevity risk sharing
  - Investment/withdrawal strategy entirely under retiree's control
- Significantly larger expected withdrawals compared to industry standard (Bengen)
  - ⇒ Significantly smaller probability of running out of cash
- Bootstrap resampling
  - ⇒ controls are robust
- Tontine provider has no risk
  - Fees can be very low
- But there is no free lunch

  "If you want more money when you are alive, you have to give up some when you are dead." *(Moshe Milevsky)*

A thought about life

To paraphrase Leonard Cohen[24]

> *"The problem with turning 70, is that I can no longer think of myself as a young man."*



- When I first heard this, I thought this was ridiculous
  - → Of course, a 70-year old is a very old man
- Now, I understand exactly what Leonard was saying

[24] Famous Canadian songwriter/singer, perhaps best known for *"Hallelujah."*