Retirement Spending Commitment as Portfolio Selection: An Efficient Frontier of Spending and Shortfall Risk

Martin-D. Lacasse* John Leonard†

April 2026

Abstract

Problem. Probability of success (PoS) is the dominant risk metric in retirement income planning, yet it is a fundamentally incomplete measure: it records only whether a plan fails, not by how much. A plan with a PoS of 70% could involve catastrophic portfolio exhaustion or merely a modest spending reduction in a few tail scenarios — two very different risk profiles that PoS cannot distinguish. A framework that quantifies both the frequency and the magnitude of potential shortfalls is needed.

Methods. We introduce a two-stage optimization approach. In the scenario stage, a full mixed-integer linear program (MILP) — implemented in the open-source Owl Retirement Planner — is solved for each member of an ensemble of historical or Monte Carlo scenarios, yielding a *scenario basis*: the maximum real first-year spending the plan can sustain under that scenario, given a comprehensive model of US federal tax law, Social Security benefits, required minimum distributions, and Medicare costs. In the commitment stage, a single linear program finds the committed spending level g^* that maximizes spending while penalizing

*Retired scientist, ExxonMobil Research & Engineering Company and Johns Hopkins University. `martin.d.lacasse@gmail.com`. The authors declare no conflicts of interest.

†Professor Emeritus of Computer Science, Virginia Commonwealth University.

mean shortfall, sweeping a risk-aversion parameter λ to trace a Pareto frontier of committed spending versus mean shortfall.

Findings. The resulting *spending/shortfall efficient frontier* is formally analogous to the Markowitz mean-variance frontier, with committed spending replacing expected return and mean shortfall replacing variance. We prove the frontier is monotone in λ and show that its extreme points nest the safe withdrawal rate (maximum risk aversion) and the unconstrained best-case spending (zero risk aversion). The commitment-stage linear program is structurally equivalent to a conditional value-at-risk (CVaR) minimization problem, endowing the shortfall measure with well-established coherence properties.

Practical Applications. The framework enables advisors and retirees to move beyond a binary success/failure threshold toward a richer characterization of retirement risk. We illustrate with historical (1928–2025) and Monte Carlo case studies and show that the efficient frontier frequently reveals that a lower PoS target is well-justified once the magnitude of potential shortfalls is understood. The Owl platform implementing these computations is freely available as open-source software.

Keywords: retirement income planning, efficient frontier, probability of success, shortfall risk, mixed-integer linear programming

JEL Codes: G11 (Portfolio Choice; Investment Decisions), G22 (Insurance; Actuarial Studies), C61 (Optimization Techniques; Programming Models)

1 Introduction

How much can a retiree safely spend each year? This deceptively simple question is central to retirement planning, yet no universally satisfying answer exists. The difficulty lies in uncertainty: future investment returns, inflation, tax policy, longevity, and health costs are all unknown at the moment of retirement.

The traditional response to this uncertainty is the *safe withdrawal rate* (SWR), most famously formulated by Bengen (1994) and subsequently popularized as the "4% rule." The SWR is the highest initial real withdrawal rate that, applied to a fixed 60/40 portfolio, would not have exhausted funds in any historical 30-year window since 1926. The Trinity Study (Cooley et al., 1998) extended this analysis across five allocation mixes and multiple time horizons. The SWR concept has the virtue of simplicity, but it is fundamentally a worst-case bound: by guaranteeing survival in the worst observed historical sequence (1966 in the US context), it may cause a retiree to spend far less than their plan can comfortably sustain in the median scenario.

The natural alternative is Monte Carlo simulation, which generates a distribution of outcomes and reports a *probability of success* (PoS) — the fraction of simulated scenarios in which the portfolio does not run out (Finke et al., 2013). Milevsky and Robinson (2005) showed that, under a lognormal return model, an analytical formula for the probability of ruin can be derived from the reciprocal Gamma distribution, bypassing simulation entirely and revealing the structural dependence of ruin probability on the Sharpe ratio relative to the withdrawal rate. While more flexible than the SWR, PoS has a critical weakness that is increasingly recognized in the literature (Pfau, 2012a; Artzner et al., 1999): it treats all failures equally. A plan that misses its spending target by \$1 in the final month of a 30-year horizon counts as a failure alongside one that exhausts funds 15 years early. Conversely, a plan with PoS $= 95\%$ may experience catastrophic shortfalls in its 5% of failing scenarios — information that PoS discards entirely.

We propose a framework that addresses this limitation while remaining practically actionable. Its core idea is borrowed from financial economics: just as Markowitz (1952) showed that portfolio selection is not a single number (expected return) but a two-dimensional trade-off (return vs. variance), retirement spending commitment is not a single number (a safe rate) but a two-dimensional trade-off between committed spending and the mean shortfall incurred when that commitment is too ambitious for some scenarios.

The framework has two stages. In the *scenario stage*, a full MILP retirement optimizer is solved for each of N_s historical or Monte Carlo scenarios, yielding a *scenario basis* g_s — the maximum real first-year spending the plan can sustain under scenario s, given all constraints of federal tax law, required minimum distributions (RMDs), Social Security benefit rules, and the retiree's chosen spending profile. In the *commitment stage*, a single small linear program selects the committed basis g^* that maximizes spending minus a penalty proportional to the expected shortfall $\sigma_s = \max(0, g^* - g_s)$ across scenarios. Sweeping the penalty weight λ from zero to infinity traces a Pareto frontier of committed spending versus mean shortfall — the *spending/shortfall efficient frontier*.

The remainder of the paper is organized as follows. Section 2 reviews the relevant literature. Section 3 briefly describes the MILP retirement engine. Section 4 reviews Markowitz's mean-variance framework. Section 5 presents the spending/shortfall efficient frontier and establishes its key properties. Section 7 illustrates the approach with representative examples. Section 8 discusses implications and limitations. Section 9 concludes.

2 Prior Work

2.1 Safe Withdrawal Rates and Probability of Success

The safe withdrawal rate literature originates with Bengen (1994), who analyzed 30-year rolling windows of US historical equity and bond returns from 1926 to 1992 and found that an initial annual withdrawal of 4.15% of the portfolio, adjusted for inflation thereafter, never exhausted a 50/50 stock-bond portfolio over any 30-year historical window. The Trinity Study (Cooley et al., 1998) broadened this analysis to five allocations and multiple horizons, reporting survival probabilities as a function of withdrawal rate. Both studies are backward-looking historical stress tests with no optimization component.

The subsequent literature examined the robustness of the 4% rule under varying market conditions (Pfau, 2012b; Kitces, 2008), spending flexibility (Pfau, 2012a), and rising-equity glide paths (Pfau and Kitces, 2014). Tail-risk objectives such as conditional value-at-risk (CVaR) penalize average loss beyond a threshold (Rockafellar and Uryasev, 2000), and coherent risk measures provide axiomatic foundations for such summaries (Artzner et al., 1999). The present paper formalizes shortfall magnitude directly in a Markowitz-style optimization framework.

2.2 Linear Programming in Retirement Planning

The use of linear programming (LP) to optimize retirement income distributions was pioneered by Welch (2015) and further developed in a sequence of papers (Welch, 2016, 2017, 2014) under the umbrella of the Interactive Optimal Retirement Planner (i-ORP). The ORP model maximizes after-tax cash available for spending subject to account balance dynamics, RMD constraints, and a simplified federal tax model. Coopersmith and Sumutka (2011, 2012) independently formulated LP models for tax-efficient withdrawal ordering across traditional, Roth, and taxable accounts. DiLellio and Ostrov (2019) extended this three-account LP framework to handle nonlinearities in tax brackets more carefully.

All of the above are *deterministic* LP models: they optimize over a single fixed scenario. i-ORP also provides a Monte Carlo mode in which many independent LP solves are performed and the resulting spending distribution is reported as a histogram (Welch, 2014). However, this architecture optimizes each scenario independently; there is no single committed first-year spending decision that is common across scenarios. The approach proposed here introduces precisely this commitment structure via a two-stage stochastic LP.

Owl (Optimal Wealth Lab) (Lacasse, 2026b) extends the LP retirement optimization framework to a full mixed-integer linear program (MILP), incorporating exact LP formulations of Social Security benefit taxability, long-term capital gains tax brackets, net investment income tax, Medicare IRMAA surcharges, ACA premium tax credits, and optimization of Social Security claiming ages. The present paper uses Owl's MILP engine to compute the scenario bases $\{g_s\}$ that feed the spending/shortfall LP.

2.3 Stochastic Programming and Stochastic Control for Retirement

Multi-stage stochastic programming has been applied to retirement planning by Konicz and Mulvey (2013); Konicz et al. (2015); Konicz and Mulvey (2015), who formulate dynamic asset allocation and contribution/withdrawal policies as scenario-tree stochastic programs. Similarly, Simsek et al. (2018) solve a multi-stage stochastic program for longevity risk management. These formulations are richer than the present approach — they permit dynamic recourse at every year — but also substantially more complex. The present paper's commitment structure (a single first-stage decision g^*, with per-scenario shortfall as the only recourse) is intentionally simpler and more interpretable for practitioners.

Milevsky and Huang (2011) derive optimal withdrawal strategies under CRRA utility with longevity risk aversion, showing that the optimal spending rate is a function of a risk-aversion coefficient that trades off current consumption against the risk of portfolio depletion. Their formulation is conceptually related to the present paper's λ parameter: both quantify willingness to accept lower spending in exchange for reduced shortfall risk. The key distinction is that Milevsky and Huang work in continuous time with CRRA utility and derive closed-form spending policies, while the present paper works in discrete time with a linear shortfall penalty and produces a plannable committed spending level rather than a time-varying spending rate.

The closest methodological cousin to the efficient frontier proposed here is the work of Forsyth (2020); Forsyth et al. (2021), who pose DC plan decumulation as a stochastic optimal control problem and trace an (expected total withdrawal, mean shortfall of terminal wealth) frontier by varying the relative weight of the two objectives in the Hamilton-Jacobi-Bellman objective function. Their approach is more general in that it permits dynamic spending and asset allocation controls recalibrated annually; it differs in that it uses continuous-time stochastic control rather than LP, and the risk measure is terminal wealth shortfall rather than per-year spending shortfall relative to a committed target. The present paper achieves a similar frontier concept through a much simpler two-stage LP structure, embedded within a fully tax-aware MILP engine.

2.4 Efficient Frontiers in Retirement Income

Pfau (2013) constructed a retirement income efficient frontier by varying the allocation among stocks, bonds, and annuities for a fixed spending target, plotting the Pareto curve of (probability of meeting the spending goal, expected residual legacy). This is the closest practitioner-oriented analogue to the present framework. The key distinction is that Pfau's frontier sweeps over *product mix* for a fixed spending target using Monte Carlo grid search, while the present paper sweeps over *committed spending levels* for a fixed portfolio structure using a formal LP. Pfau's framework is not an optimization; the present framework is.

2.5 CVaR and Expected Shortfall in Optimization

Rockafellar and Uryasev (2000) showed that minimizing the conditional value-at-risk (CVaR) — the expected loss in the worst α fraction of scenarios — is equivalent to a linear program. Specifically, they demonstrated that CVaR at confidence level α can be minimized by introducing a scalar auxiliary variable and N_s nonnegative shortfall variables, one per scenario. The commitment-stage LP proposed in this paper is structurally identical: the shortfall variables $\sigma_s = \max(0, g^* - g_s)$ are the per-scenario loss variables of Rockafellar-Uryasev, and the penalty $(\lambda/N_s)\sum_s \sigma_s$ is the (unscaled) CVaR of the spending shortfall distribution. Rockafellar-Uryasev therefore provides the mathematical foundation confirming that the proposed LP is tractable and that the risk measure is coherent in the sense of Artzner et al. (1999).

3 The MILP Retirement Optimizer

The scenario stage requires solving a retirement optimization problem for each of the N_s scenarios. We use Owl's MILP engine for this purpose; this section provides a brief self-contained description. A complete mathematical treatment is given in Lacasse (2026a).

Plan structure. A retirement plan covers a horizon of N_n years for a household of N_i individuals (typically one or two). The plan begins in calendar year y_0 (the "plan year"), typically the current year. Four account types are modeled: taxable accounts ($j = 0$), traditional tax-deferred accounts ($j = 1$), Roth accounts ($j = 2$), and health savings accounts ($j = 3$).

Decision variables. The MILP selects, for each year $n = 0, \ldots, N_n - 1$:

- b_{ijn}: withdrawals from account type j for individual i in year n;
- x_{in}: Roth conversion amounts for individual i;
- g_n: net spending in year n (inflation-adjusted).

Additional continuous variables model tax bracket occupancy, long-term capital gains allocation across the 0%/15%/20% brackets, Medicare IRMAA tier selection, and Social Security benefit levels. Binary variables select the applicable tax bracket, IRMAA tier, and (when optimizing claiming ages) Social Security filing month.

Constraints. The model enforces: (i) annual cash flow balance; (ii) account balance dynamics with portfolio returns; (iii) federal income tax computation including ordinary income brackets, long-term capital gains brackets, net investment income tax, and Social Security benefit taxability (0%, 50%, or 85%); (iv) Medicare IRMAA surcharges as a function of MAGI with a two-year lag; (v) RMDs as hard lower bounds on traditional account withdrawals; (vi) ACA premium tax credits when applicable; (vii) a user-specified normalized spending profile ξ_n such that $g_n = g^* \cdot \xi_n$ for a scalar basis g^*.

Objective. The primary objective considered in this paper is to maximize the first-year real spending basis $g^* = g_0/\xi_0$. The scenario basis g_s is the unconstrained optimal value of g^* that Owl can achieve under the rate series of scenario s.

Rates. Each scenario s corresponds to a specific time series of nominal returns for equities and bonds and an inflation series. In historical mode, scenario s uses the actual US return data starting in year y_s (from the Owl rates database, 1928–2025, sourced from Damodaran (2025)). In Monte Carlo mode, Owl supports several stochastic return models; the examples in this paper use the *histolognormal* model, which fits a multivariate lognormal distribution to the full historical record by estimating the log-space mean vector and covariance matrix from historical log-returns and sampling i.i.d. draws. This is the standard maximum-likelihood fit of a correlated lognormal to historical data, the same class of model used as a baseline in Finke et al. (2013) and reviewed in Collins et al. (2015). The lognormal is preferable to the multivariate normal for asset returns because portfolio values are products of gross-return factors, making log-returns the natural additive quantity; the VAR framework of Campbell and Viceira (2002) represents the next level of sophistication, incorporating mean reversion and time-varying risk premiums, but is not required for the illustrative purposes of this paper.

4 The Markowitz Mean-Variance Efficient Frontier

Before presenting the spending/shortfall frontier, we briefly review the Markowitz framework that motivates it.

The mean-variance problem. Consider a universe of M risky assets with expected return vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$. An investor chooses portfolio weights $\boldsymbol{w}$ to solve, for each level of risk aversion $\lambda \geq 0$:

$$\max_{\boldsymbol{w}} \quad \boldsymbol{\mu}^\top \boldsymbol{w} - \lambda\, \boldsymbol{w}^\top \boldsymbol{\Sigma} \boldsymbol{w} \quad \text{subject to} \quad \boldsymbol{w} \geq \boldsymbol{0},\ \boldsymbol{1}^\top \boldsymbol{w} = 1. \tag{1}$$

As λ varies from 0 (risk-neutral; maximum expected return regardless of variance) to ∞ (minimum variance regardless of return), the solution traces the *efficient frontier*: the set of portfolios that achieve the maximum expected return for each level of variance. The investor selects the point on this curve matching their personal risk tolerance.

The role of λ. The parameter λ does not need to be estimated; it is a free parameter that *parameterizes* the frontier. Any specific $(\text{return}, \text{variance})$ pair on the frontier corresponds to exactly one λ. In practice, the investor chooses a point on the frontier based on how much variance they are willing to accept per unit of expected return.

5 The Spending/Shortfall Efficient Frontier

5.1 Scenario Bases

For each of N_s scenarios, let the MILP retirement optimizer (Section 3) be solved under the rate series of scenario s. Denote the resulting optimal first-year real spending basis by g_s (in today's dollars per year). The collection $\{g_s\}_{s=1}^{N_s}$ represents the range of spending outcomes the plan could sustain across the scenario ensemble, with no commitment imposed.

5.2 The Committed Spending LP

Suppose the retiree commits to a single first-year real spending basis g^*, to be applied regardless of which scenario is realized. In scenario s, if $g^* > g_s$, the plan experiences a *shortfall*:

$$\sigma_s = \max(0,\ g^* - g_s), \quad s = 1, \ldots, N_s. \tag{2}$$

The shortfall σ_s is the annual real spending reduction the plan would have to absorb in scenario s if forced to commit to g^* despite that scenario only being able to sustain $g_s < g^*$.

Paralleling the Markowitz objective (1), we define the committed spending problem as: choose g^* to maximize spending minus a penalty proportional to expected shortfall:

$$\max_{g^*, \boldsymbol{\sigma}} \quad g^* - \frac{\lambda}{N_s} \sum_{s=1}^{N_s} \sigma_s \tag{3}$$

subject to:

$$\sigma_s \geq g^* - g_s, \quad s = 1, \ldots, N_s, \tag{4}$$

$$\sigma_s \geq 0, \quad s = 1, \ldots, N_s, \tag{5}$$

$$0 \leq g^* \leq \max_s g_s. \tag{6}$$

This is a linear program in $1 + N_s$ variables and $2N_s + 2$ constraints. For a given λ, it is solved in milliseconds for any practical N_s (typically 50–200).

Connection to CVaR. As shown by Rockafellar and Uryasev (2000), the penalty $(\lambda/N_s)\sum_s \sigma_s$ in (3)–(5) is exactly the (scaled) CVaR of the shortfall distribution when g^* plays the role of the value-at-risk threshold. The LP (3)–(6) therefore has a clear risk-theoretic interpretation: it maximizes committed spending while minimizing the mean shortfall in the worst-case scenarios. Because CVaR is a coherent risk measure (Artzner et al., 1999), the resulting frontier inherits desirable theoretical properties.

5.3 The Efficient Frontier and Its Properties

Solving the LP (3)–(6) for a range of λ values traces a curve in the $(g^*, \bar{\sigma})$ plane, where $\bar{\sigma} = (1/N_s)\sum_s \sigma_s$ is the mean shortfall — the unconditional average over all N_s scenarios, assigning zero to scenarios without shortfall. (The conditional shortfall per failing scenario is $\bar{\sigma}/(1-\rho)$, where ρ is the fraction of scenarios with $\sigma_s = 0$.) We call this the *spending/shortfall efficient frontier*. The following proposition confirms its Pareto structure.

Proposition 1 (Frontier Monotonicity)**.** *As λ increases, the optimal committed spending $g^*(\lambda)$ is non-increasing and the mean shortfall $\bar{\sigma}(\lambda)$ is non-increasing.*

Proof. Let $\lambda_1 < \lambda_2$ with corresponding optimal solutions $(g_1^*, \boldsymbol{\sigma}_1)$ and $(g_2^*, \boldsymbol{\sigma}_2)$. By optimality:

$$g_1^* - \lambda_1 \bar{\sigma}_1 \geq g_2^* - \lambda_1 \bar{\sigma}_2, \tag{7}$$

$$g_2^* - \lambda_2 \bar{\sigma}_2 \geq g_1^* - \lambda_2 \bar{\sigma}_1. \tag{8}$$

Adding (7) and (8) gives $0 \leq (\lambda_2 - \lambda_1)(\bar{\sigma}_1 - \bar{\sigma}_2)$. Since $\lambda_2 > \lambda_1$ we conclude $\bar{\sigma}_1 \geq \bar{\sigma}_2$. Substituting back into (7) gives $g_1^* \geq g_2^*$. □

The proposition confirms that the frontier is a Pareto curve: higher risk aversion (λ) yields lower committed spending and lower mean shortfall. The two extreme points have intuitive interpretations:

- At $\lambda = 0$ (risk-neutral): $g^* = \max_s g_s$. The committed spending equals the highest scenario basis, ignoring all shortfall risk.

- As $\lambda \to \infty$ (maximin): $g^* \to \min_s g_s$, with zero shortfall in all scenarios. This is the *safe withdrawal rate* — the largest commitment that survives even the worst scenario.

The analogy with the Markowitz frontier is exact in structure, with the correspondences summarized in Table 1. The $\lambda \to \infty$ extreme deserves particular note: it is the spending analogue of holding all wealth in the minimum-variance portfolio. Just as Markowitz (1952) demonstrated that rational investors need not minimize variance to its theoretical extreme — the efficient frontier offers superior risk-adjusted outcomes for any given level of risk tolerance — the spending/shortfall frontier demonstrates that rational retirees need not adopt the SWR: moving slightly away from the zero-shortfall extreme typically yields a substantially higher committed spending level for only a modest increase in mean shortfall.

Table 1: Analogy between the Markowitz mean-variance frontier and the spending/shortfall frontier.

Concept	**Markowitz (MPT)**	**This paper**
Decision variable	Portfolio weights $\boldsymbol{w}$	Committed spending g^*
Reward to maximize	Expected return $\boldsymbol{\mu}^\top \boldsymbol{w}$	Committed spending g^*
Risk to minimize	Variance $\boldsymbol{w}^\top \boldsymbol{\Sigma} \boldsymbol{w}$	Mean shortfall $\bar{\sigma}$
Risk aversion	λ (penalty on variance)	λ (penalty on shortfall)
Efficient frontier	Return vs. standard deviation	Spending vs. mean shortfall
Conservative extreme	Minimum-variance portfolio	Safe withdrawal rate ($\min_s g_s$)
Aggressive extreme	Maximum-return portfolio	$g^* = \max_s g_s$

5.4 Selecting the Committed Spending Level

The efficient frontier presents a Pareto curve; the retiree must select a point. Rather than specifying λ directly, we parameterize the choice by a *target success rate* $\rho \in [0,1]$: the desired fraction of scenarios in which g^* does not exceed the scenario basis (i.e., $\sigma_s = 0$). For a target ρ, we select the least conservative λ^* such that the shortfall probability $P(\sigma_s > 0) \leq 1 - \rho$. The result $g^*(\lambda^*)$

is the maximum committed spending level meeting the target success rate. Since the frontier is precomputed over a sweep of λ values, this selection is a simple lookup requiring no additional LP solves.

Visualization. We display the frontier in two complementary panels:

1. **Success rate curve**: g^* vs. shortfall probability $P(\sigma_s > 0)$. The target ρ selects a point whose g^* coordinate gives the committed spending.

2. **Efficient frontier**: g^* vs. mean shortfall $\bar{\sigma}$. This is the classical Pareto curve, analogous to the mean-variance frontier.

Remark 1. *The scenario bases $\{g_s\}$ are a* sufficient statistic *for the commitment stage: all information from the N_s MILP solutions relevant to the spending commitment decision is captured in one number per scenario. This separation of the computationally expensive scenario stage from the trivially fast commitment stage is a key practical advantage of the two-stage formulation.*

6 Longevity as a Scenario Dimension

6.1 SSA Period Life Tables

The plan horizon N_n used in the MILP optimizer is typically set to the retiree's life expectancy at plan inception. This is a deterministic approximation that ignores the stochastic nature of individual lifespans. We now show how longevity uncertainty can be incorporated into the efficient frontier framework without increasing the number of scenario solves.

Let $q_{x,k}$ denote the annual death probability for an individual of sex $k \in \{\mathrm{M}, \mathrm{F}\}$ who has reached exact age x, taken from the Social Security Administration period life table (Social Security Administration, 2025). The survival function, conditional on being alive at age x_0, is

$$S_k(x; x_0) = \prod_{j=x_0}^{x-1} (1 - q_{j,k}), \quad x = x_0, x_0 + 1, \ldots, 119, \tag{9}$$

and the corresponding probability mass function (PMF) for the discrete lifespan T_k is

$$P(T_k = x; x_0) = S_k(x; x_0) \cdot q_{x,k}, \quad x \in \{x_0, \dots, 119\}, \tag{10}$$

normalized to sum to 1. The age 119 is assigned $q_{119,k} = 1$ by convention (certain death by age 120). Equation (10) defines the conditional lifespan distribution used to generate stochastic plan horizons in the scenario stage.

6.2 Longevity-Augmented Scenario Bases

In the fixed-horizon formulation of Section 5, every scenario s shares the same planning horizon N_n. To incorporate longevity risk, we augment each scenario with an independently drawn lifespan.

Single retiree. For a single individual of sex k and current age x_0, draw $\tau_s \sim P(T_k; x_0)$ independently for each scenario $s = 1, \dots, N_s$. Solve the MILP optimizer with horizon τ_s under the return sequence of scenario s. The resulting optimal spending basis g_s now encodes both the return risk of scenario s and the longevity risk embodied in τ_s.

Couple. For a couple with individual lifespans $\tau_{1,s} \sim P(T_{k_1}; x_{0,1})$ and $\tau_{2,s} \sim P(T_{k_2}; x_{0,2})$ drawn independently, the plan horizon extends until the second death. The last-survivor horizon is

$$\tau_s = \max(\tau_{1,s}, \tau_{2,s}). \tag{11}$$

The MILP optimizer is solved with horizon τ_s and the full couple model, including the survivor benefit provisions already embedded in the planner.

In both cases, once $\{g_s\}_{s=1}^{N_s}$ is computed, the commitment LP of Section 5 is applied unchanged. The structure of the outer LP — and hence the monotonicity proof of Proposition 1 — is independent of how the scenario bases were generated.

Remark 2. *The longevity-augmented frontier is a distributional average of fixed-horizon frontiers: each g_s is drawn from a mixture over plan horizons weighted by the lifespan PMF. Because the PMF is smooth and the MILP objective is continuous in N_n, the resulting scenario ensemble yields a smoother frontier than a single fixed-horizon run, particularly in historical mode where short scenario counts can produce a jagged curve.*

6.3 Preservation of Frontier Properties

The Proposition 1 and its proof in Section 5 depend only on the structure of the commitment LP — specifically, that the scenario bases $\{g_s\}$ are fixed scalars when the commitment LP is solved. The mechanism by which g_s was produced — fixed horizon, stochastic returns, stochastic horizon, or any combination — is irrelevant to the proof. The monotonicity of the spending/shortfall frontier therefore holds without modification in the longevity-augmented setting.

6.4 Computational Considerations

The longevity augmentation does not increase the scenario count: one lifespan draw per scenario is sufficient, as the return sequence and the horizon are jointly realized for each scenario s. The scenario bases $\{g_s\}$ and the drawn horizons $\{\tau_s\}$ are computed once and reused across the full λ-sweep of the commitment stage.

An important practical benefit is that short-horizon draws ($\tau_s \ll \bar{N}_n$, corresponding to early-death scenarios) produce smaller MILP instances that solve faster, partially offsetting the overhead of cloning and re-initializing the plan for each scenario. For long-lived scenarios the computational cost is unchanged.

7 Examples

7.1 Historical Scenarios

Setup.

Results.

The 1965–1966 sequence.

7.2 Monte Carlo Scenarios

Setup.

Results.

7.3 Effect of Spending Profile

8 Discussion

8.1 Probability of Success as an Incomplete Risk Measure

The spending/shortfall frontier makes visible a distinction that PoS obscures: not all shortfalls are equal. A plan with PoS = 70% could involve trivial shortfalls (average of $2,000/yr in bad scenarios) or catastrophic ones (portfolio exhaustion at age 75). The mean shortfall axis of the efficient frontier directly measures the average magnitude of the gap — a quantity invisible in a PoS-only analysis.

This distinction mirrors the shift in portfolio risk management from standard deviation to CVaR: standard deviation treats upside and downside symmetrically and gives no information about the magnitude of tail losses. The mean shortfall $\bar{\sigma}$ proposed here is the retirement income analogue of CVaR — a linear, coherent measure of downside risk that captures magnitude, not merely probability.

8.2 Dynamic Nature of Retirement Planning

A committed spending basis g^* is not a lifetime contract. In practice, most retirees re-evaluate their plan annually as new information arrives: updated portfolio values, tax law changes, revised return forecasts, unexpected expenses, or health changes. The committed spending decision is therefore a *one-period commitment* — the spending level adopted for the coming year — rather than a permanent pledge.

This annual re-optimization is analogous to model-predictive control in engineering: a receding-horizon policy in which the commitment stage is solved fresh each year with updated information. The framework proposed here is the single-period kernel of such a receding-horizon policy.

8.3 Computational Considerations

Why an optimizer is required. The efficient frontier depends critically on the scenario basis g_s — the *maximum* real spending the plan can sustain under scenario s. This quantity is only available from an optimizer. Traditional forward-projection tools simulate a fixed spending level year by year and report the year in which the portfolio is exhausted; they record only *whether* a plan fails at a given spending rate, not the maximum spending level the scenario could actually support. Without g_s, the shortfall $\sigma_s = \max(0, g^* - g_s)$ cannot be computed, and the commitment LP has no inputs. The two-stage MILP framework is therefore not merely a convenience — it is what makes the efficient frontier computationally feasible.

The scenario stage is computationally intensive: it requires N_s MILP solves, each taking a few seconds on modern hardware. For $N_s = 100$ historical scenarios and a 25-year plan horizon, the scenario stage completes in a few minutes on a laptop. The commitment stage and the full frontier sweep over approximately 60 values of λ, sampled on a logarithmic scale to obtain uniform visual density across the full curve, complete in under one second. Crucially, once the scenario stage is complete, changing the target success rate ρ triggers only the fast commitment-stage lookup — the scenario MILP solves are not repeated.

8.4 Limitations and Extensions

Fixed asset allocation. The scenario bases are computed with a fixed user-specified asset allocation. The framework does not jointly optimize spending and allocation. This is a natural extension for future work.

Single committed spending basis. The commitment variable g^* is a scalar. Dynamic spending rules could in principle be incorporated by replacing the scalar commitment with a spending policy, though this would substantially complicate the outer optimization.

Scenario representativeness. In historical mode, the frontier is only as representative as the historical data. The US post-1928 record covers a limited number of distinct market regimes, and the worst historical sequence (1966) is a single data point. Monte Carlo mode mitigates this at the cost of model assumptions in the return-generating process. Under i.i.d. return models (e.g., lognormal), the sample minimum spending converges toward zero as the number of simulations grows — not because spending can go negative, but because i.i.d. sampling eventually generates sequences of many consecutive poor years that real markets suppress through mean reversion, volatility clustering, and policy responses. Reporting the 5th percentile rather than the sample minimum provides a stable tail indicator that converges as N_s grows. The committed spending g^* itself is unaffected by this issue, since it is determined by the success-rate constraint rather than the single worst draw. Models with autocorrelation structure — block bootstrap, stationary bootstrap (Politis and Romano, 1994), or GARCH — produce more realistic extreme sequences and are preferable when tail behavior is the primary concern.

Longevity uncertainty. Longevity risk is modeled in Section 6: plan horizons are drawn from SSA period life tables, one per scenario, incorporating both single-life and last-survivor couple cases without increasing the scenario count.

9 Conclusion

We have presented a spending/shortfall efficient frontier for retirement planning that formalizes the intuition of the safe withdrawal rate and the probability-of-success metric within a rigorous optimization framework. The framework has two stages: a MILP retirement optimizer (solving for the optimal unconstrained spending basis under each of N_s scenarios) and a commitment-stage LP (finding the Pareto-optimal committed spending level given a risk-aversion parameter λ). Proposition 1 confirms that sweeping λ traces a monotone Pareto frontier formally analogous to the Markowitz mean-variance frontier, with committed spending in place of expected return and expected shortfall in place of variance.

The key practical insight is that probability of success is an incomplete risk measure: it tells a retiree only how often their plan might fall short, not by how much. The spending/shortfall frontier makes the magnitude of potential shortfalls directly visible, enabling a more nuanced spending commitment. A retiree who understands that a 70% success rate involves an average shortfall of only $2,000/yr may reasonably choose that point over a 90% success rate that requires $15,000/yr less spending — a trade-off that PoS alone cannot reveal.

The framework is implemented in the open-source Owl (Lacasse, 2026b), available at `https://github.com/mdlacasse/Owl`.

An avenue for future work is extending the efficient frontier concept to traditional forward-projection tools, which do not return an optimal spending basis but report the year in which the portfolio is exhausted. A natural proxy for the shortfall magnitude is the present value of the spending gap between the failure year and the planning horizon, discounted at the inflation rate. While cruder than the MILP-derived g_s, such an approximation would bring shortfall-aware frontier analysis to the broad universe of simulation tools already in use by practitioners, substantially widening the framework's reach without requiring a full optimization engine.

Acknowledgments

Declaration of Interest

The authors report no conflicts of interest. Owl (Optimal Wealth Lab) is open-source software released under the GNU General Public License; the authors receive no financial benefit from its use.

References

Artzner, P., F. Delbaen, J.-M. Eber, and D. Heath (1999). Coherent measures of risk. *Mathematical Finance 9*(3), 203–228.

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

Campbell, J. Y. and L. M. Viceira (2002). *Strategic Asset Allocation: Portfolio Choice for Long-Term Investors*. Oxford: Oxford University Press.

Collins, P. J., H. D. Lam, and J. Stampfli (2015). How risky is your retirement income risk model? *Financial Services Review 24*, 193–216.

Cooley, P. L., C. M. Hubbard, and D. T. Walz (1998, February). Retirement savings: Choosing a sustainable withdrawal rate. *Journal of the American Association of Individual Investors*, 16–21. Also known as the Trinity Study.

Coopersmith, L. W. and A. R. Sumutka (2011, September). Tax-efficient retirement withdrawal planning using a linear programming model. *Journal of Financial Planning*.

Coopersmith, L. W. and A. R. Sumutka (2012, April). Tax-efficient retirement withdrawal planning using a comprehensive tax model. *Journal of Financial Planning*.

Damodaran, A. (2025). Annual returns on stock, t.bills, t.bonds and inflation—united states (1928–present). NYU Stern School of Business, available at `https://pages.stern.nyu.edu/~adamodar/`. Historical annual returns dataset, updated annually.

DiLellio, J. A. and D. N. Ostrov (2019). Constructing tax-efficient withdrawal strategies for retirees with traditional 401(k)/iras, roth 401(k)/iras, and taxable accounts. *Financial Services Review 28*(2), 67–95.

Finke, M. S., W. D. Pfau, and D. M. Blanchett (2013, June). The 4 percent rule is not safe in a low-yield world. *Journal of Financial Planning 26*(6), 46–55.

Forsyth, P. A. (2020). A stochastic control approach to defined contribution plan decumulation: "the nastiest, hardest problem in finance". *ASTIN Bulletin: The Journal of the IAA*. arXiv:2008.06598.

Forsyth, P. A., K. R. Vetzal, and G. Westmacott (2021). Optimal control of the decumulation of a retirement portfolio with variable spending and dynamic asset allocation. *ASTIN Bulletin: The Journal of the IAA 51*(3), 905–938.

Kitces, M. E. (2008, May). Resolving the paradox – is the safe withdrawal rate sometimes too safe? *The Kitces Report*.

Konicz, A. K. and J. M. Mulvey (2013). Applying a stochastic financial planning system for an individual: Immediate or deferred life annuities? *Journal of Retirement 1*(1), 90–103.

Konicz, A. K. and J. M. Mulvey (2015). Optimal savings management for individuals with defined contribution pension plans. *European Journal of Operational Research 243*(1), 233–247.

Konicz, A. K., D. Pisinger, K. M. Rasmussen, and M. Steffensen (2015). A combined stochastic programming and optimal control approach to personal finance and pensions. *OR Spectrum 37*(2), 583–616.

Lacasse, M.-D. (2026a). Optimal wealth lab (Owl): Mathematical foundations of a retirement planning optimizer. Technical report, GitHub. Available at `https://github.com/mdlacasse/Owl/blob/main/papers/owl.pdf`.

Lacasse, M.-D. (2026b). Owl: Optimal wealth lab – a mixed-integer linear programming retirement financial planner. `https://github.com/mdlacasse/Owl`. Version 2026.04.08. Additional contributors are listed in the `AUTHORS` file at `https://github.com/mdlacasse/Owl/blob/main/AUTHORS`. The case file `Case_chris+pat.toml` is included in the repository for reproducibility.

Markowitz, H. (1952). Portfolio selection. *Journal of Finance 7*(1), 77–91.

Milevsky, M. A. and H. Huang (2011). Spending retirement on planet Vulcan: The impact of longevity risk aversion on optimal withdrawal rates. *Financial Analysts Journal 67*(2), 45–58.

Milevsky, M. A. and C. Robinson (2005). A sustainable spending rate without simulation. *Financial Analysts Journal 61*(6), 89–100.

Pfau, W. D. (2012a, March). Spending flexibility and safe withdrawal rates. *Journal of Financial Planning*.

Pfau, W. D. (2012b, January). Withdrawal rates, savings rates, and valuation-based asset allocation. *Journal of Financial Planning*.

Pfau, W. D. (2013, February). A broader framework for determining an efficient frontier for retirement income. *Journal of Financial Planning*. SSRN 2151259.

Pfau, W. D. and M. E. Kitces (2014). Reducing retirement risk with a rising equity glide-path. *Journal of Financial Planning*. SSRN 2324930.

Politis, D. N. and J. P. Romano (1994). The stationary bootstrap. *Journal of the American Statistical Association 89*(428), 1303–1313.

Rockafellar, R. T. and S. Uryasev (2000). Optimization of conditional value-at-risk. *Journal of Risk 2*(3), 21–41.

Simsek, K. D., M. J. Kim, W. C. Kim, and J. M. Mulvey (2018). Optimal longevity risk management in the retirement stage of the life cycle. *Journal of Retirement 5*(3), 73–92.

Social Security Administration (2025). Actuarial life table, period life table 2021. `https://www.ssa.gov/oact/STATS/table4c6.html`. Accessed April 2026.

Welch, J. S. (2014). The optimal retirement planner (ORP). Available at `https://www.i-orp.com` (offline as of 2024). LP-based retirement planner; source code not publicly released.

Welch, J. S. (2015). Mitigating the impact of personal income taxes on retirement savings distributions. *Journal of Personal Finance*.

Welch, J. S. (2016). Measuring the financial consequences of IRA to Roth IRA conversions. *Journal of Personal Finance*.

Welch, J. S. (2017, August). A 3-step procedure for computing sustainable retirement savings withdrawals. *Journal of Financial Planning*.