Simple Dynamic Stock/Bond/Gold Portfolios

Nikhil Devanathan   Alexandros E. Tzikas   Stephen P. Boyd

September 5, 2026

Abstract

For more than four decades, the 60/40 stock/bond portfolio has served as a benchmark for delivering reasonable returns without excessive risk. More recently, a 50/30/20 stock/bond/alternative portfolio has been suggested. We use gold as the alternative and as an inflation hedge. In this paper we ask: how much improvement over these benchmark fixed-weight portfolios can be obtained using widely available public data and standard methods from quantitative finance?

We restrict ourselves to long-only dynamic portfolios of stocks, bonds, and gold, plus cash, rebalancing monthly, using only publicly available data. We evaluate portfolios on the conventional metrics: return, volatility, Sharpe ratio (computed in excess of the federal funds rate), drawdown, and turnover, in addition to consistency of performance over time, judged by the consistency of the realized annual volatility.

Over the 20–year period 2006–2026, using a conservative estimate of trading costs, we show that all risk-adjusted and drawdown metrics are improved using simple volatility control, where we dynamically mix the fixed-weight portfolios with cash so as to target a fixed volatility. This method relies on a simple estimate of portfolio volatility derived from past returns. We also demonstrate that more sophisticated portfolios based on convex optimization—similar to those used in quantitative hedge funds—yield further substantial improvement in return and risk-adjusted return. We consider two such portfolios, one that uses a simple estimate of future returns based on past returns, and one that forecasts future returns based on past returns and just a handful of widely available public economic data. These portfolios also outperform a suite of standard risk-based allocation methods, such as risk parity and minimum variance, evaluated on the same assets and data.

Our paper is accompanied by open-source software that implements the methods and replicates all results.

1 Introduction

A risk-averse investor seeking steady growth of capital is conventionally advised to hold a fixed-weight 60/40 portfolio of broad stock and bond indices, rebalanced annually [7, 9]. As the correlation between stock and bond returns has risen in recent years [15], variants that add a diversifying third asset class have gained traction, such as a 50/30/20 stock/bond/alternative portfolio [27]. Motivated by this proposal and by proposals to use gold as an inflation hedge [63], we use a 50/30/20 stock/bond/gold portfolio as a gold-augmented benchmark.

We ask a simple question: how much better can we do using standard tools of quantitative finance? We work strictly within the constraints that make this simple setting attractive to an individual investor: only the three broad assets (stocks, bonds, and gold), monthly rebalancing, long only, no leverage, no derivatives, and no proprietary data. We restrict ourselves to methods that are simple and completely transparent, can be implemented in just a few lines of code, and rely only on publicly available data.

Given these constraints, and the very small set of assets we invest in, we do not expect hedge-fund-level results. Instead we judge our methods against the classical 60/40 portfolio and our 50/30/20 gold-augmented benchmark. We find that we can obtain a substantial lift in performance compared to these benchmarks.

We state our contribution plainly. It is not methodological: volatility targeting, momentum, regression-based return forecasting, and mean-variance optimization with a risk constraint are all standard and well established. What is new is empirical: what these standard components deliver under the constraints an individual investor actually faces—three liquid assets plus cash, long only, no leverage, monthly rebalancing, only public data, and net of trading costs—measured against the 60/40 and our 50/30/20 benchmark. We evaluate them not only on the usual metrics, but also on an inflation-adjusted and a post-tax basis. Finally, we give a transparent and fully reproducible implementation, with open-source code, that a self-directed investor could run themselves.

We show this using the exchange-traded funds (ETFs) SPY (broad U.S. equity), AGG (U.S. investment-grade bond market), and GLD (gold), plus cash, with simulations over the 20–year period 2006–2026, net of a conservative estimate of trading costs. We compare three families of portfolios.

- *Fixed-weight benchmarks.* The conventional 60/40 and our 50/30/20 baseline portfolios holding SPY/AGG/GLD, rebalanced annually.

- *Volatility-controlled portfolios.* The same fixed-weight portfolios, diluted with cash each month so that their estimated (ex-ante) volatility equals a target, such as 7% annualized [22].

- *Optimization-based portfolios.* Monthly rebalanced portfolios found using mean-variance or, more generally, Markowitz-type methods, which use convex optimization to maximize a forecast of forward return subject to a limit on ex-ante volatility [11]. We consider two return forecasts: a simple trailing average of past returns, and a regression forecast based on asset and economic features. We refer to the two associated portfolios as simple Markowitz and Markowitz, respectively.

We judge these portfolios using conventional metrics such as return, risk, average and maximum drawdown. Our principal objective is Sharpe ratio, computed as the ratio of excess return (above the risk free federal funds rate) to the volatility. We also consider the steadiness or consistency of performance, judged by the variation in annual realized risk across different economic regimes. We find that volatility control improves these metrics over the fixed-weight benchmarks, with a substantial increase in Sharpe ratio along with more consistent performance. The Markowitz portfolios improve the metrics further, lifting the Sharpe ratio to 1.08, nearly twice the Sharpe ratio of the 60/40 portfolio (0.56). We find that these improvements persist when they are judged using inflation-adjusted or post-tax metrics.

Since both volatility control and our Markowitz portfolios size positions using an estimate of risk, it is fair to ask whether their advantage is just a generic benefit of risk control. To answer this we also compare against seven standard risk-based allocation methods—among them risk parity, equal risk contribution, minimum variance, and Black–Litterman—using the same assets, data, costs, and constraints. None of them reaches the performance of our Markowitz portfolio: at a common volatility target their Sharpe ratios lie between 0.60 and 0.85, against 1.08. We give this comparison in appendix F.

Open-source software that implements all methods and reproduces every result accompanies the paper and can be found at

https://github.com/cvxgrp/simple-portfolio-code.

1.1 Related work

The fixed-weight stock/bond portfolio is the benchmark against which we measure performance. Its appeal rests partly on evidence that a portfolio's long-term allocation across broad asset classes—for example, its target weights in stocks and bonds—explains most of the variation in that portfolio's returns over time, whereas security selection and short-term market-timing decisions explain substantially less [13, 14]. At a high level, equities supply the return and bonds hedge equity drawdowns in a portfolio. This hedge is good when the stock/bond correlation is negative or positive and small. However, this correlation is regime-dependent: negative when growth shocks dominate but positive when inflation shocks occur [38, 17, 15]. Nevertheless, evidence suggests that there is an overall diversification benefit to holding bonds [41]. As episodes of positive stock–bond correlation have become more prominent in recent years, especially after 2010, the diversification benefit of the 60/40 portfolio has shrunk [60], motivating the inclusion of a third, separately driven asset. Gold was proposed as it can both hedge equities and act as a safe haven in market stress [35, 5, 6]. Unlike the private alternatives typically used for the same purpose, gold is directly accessible to a self-directed investor. There are, however, counterarguments to including gold in a portfolio, as researchers have questioned its reliability as an inflation hedge or source of stable real value [24]. We demonstrate that, despite these concerns, a portfolio including gold outperforms other portfolios.

The combination of equities, fixed income, and gold also has practical precedents in investable multi-asset funds. The permanent portfolio mutual fund (PRPFX) combines stocks, government securities, and precious metals [52], while the RPAR ETF allocates across equities, nominal and inflation-linked government bonds, and commodities including gold [56]. Other multi-asset ETFs use either broad strategic allocations or momentum-based rules to vary exposure across stocks, bonds, and real assets such as commodities, real estate, and infrastructure. Such dynamic allocations are motivated by evidence that dynamic asset allocations can improve static ones [41].

Scaling exposure to a target volatility is a long-standing way to improve risk-adjusted return and attenuate the left tail of the return distribution, by de-risking when volatility is high [48, 32, 36, 25, 28, 40]. The method also has industry precedents. Risk control indices scale exposure using realized volatility and a specified volatility target [61]. This is effective because periods of high volatility often coincide with market stress and adverse returns. The evidence is not uniformly favorable. Across a large set of equity strategies, volatility management does not systematically improve out-of-sample performance [18]. Volatility control belongs to a broader family of risk-based allocation rules—such as risk parity and equal risk contribution—that size positions by risk rather than by forecast return [54, 43, 1]. Risk parity and equal risk contribution portfolios choose relative asset weights so that portfolio risk is distributed more evenly across assets. The same family includes inverse volatility weighting, sometimes called naïve risk parity [42], risk budgeting, which splits risk in prescribed (not necessarily equal) proportions [16, 55], the global minimum variance portfolio [33, 20], the maximum diversification portfolio [19], and the equal weight, or $1/N$, rule [21]. Volatility control addresses a different dimension of the allocation problem. Rather than prescribing how risk should be divided among assets, it controls the total amount of risk taken by scaling a given portfolio toward a specified volatility level. We adopt volatility targeting but keep the long-only, unlevered constraints of our setting, so the strategy de-risks into cash in volatile regimes, but does not lever up in calm ones. In appendix F we implement seven of these rules in our setting, and find that none of them attains the risk-adjusted performance of our Markowitz portfolio.

Our optimization-based portfolios combine a future return estimate with a returns covariance estimate in the mean–variance framework of [44], whose canonical scalar summary is the Sharpe ratio [58, 59], and follows the practical-implementation perspective of [8, 10]. Empirical studies suggest that fixed-weight portfolios can be difficult to beat with optimization-based portfolios, because future return forecasting is challenging [21]. In addition, Markowitz-type portfolios have received criticism for their sensitivity to the return forecast and covariance estimate [46]. A well-known remedy is the Black–Litterman model [8, 34], which shrinks a return forecast toward the returns implied by an equilibrium portfolio; in appendix F we compare against it, giving it our own return forecast as its views. We demonstrate that, with carefully constructed inputs and constraints, the Markowitz portfolios outperform the fixed-weight benchmarks. The return forecast used in the simple Markowitz portfolio is a simple trailing average of past returns, which is a common momentum signal [49, 37]. Momentum signals have performed well for many decades [37], and scaling momentum by volatility can reduce its crash risk [4]. The return forecast used in our Markowitz portfolio is a rolling regression of forward returns on asset-specific and macro-financial features. Ridge regularization stabilizes the fit when predictors are correlated or numerous relative to the rolling sample. Return prediction is difficult, and many macro-financial predictors perform poorly out of sample [62]. We therefore keep the model deliberately simple. Documented return predictors tend to decay once published [45], and complicated models are prone to over-fitting [31, 2, 3, 57]. More elaborate dynamic portfolio methods account directly for signal persistence and trading cost [29]. Our method instead uses a simpler one-period formulation.

Our optimization-based approach of first forecasting returns and then using the forecasts in order to decide on allocations shares similarities with [39, 50, 41]. Kim et al. [39] construct monthly U.S. growth and inflation indicators that define economic regimes, estimate asset returns for each regime, and shift exposures toward attractive assets based on the current regime. Mueller-Glissmann et al. [50] use macro and market indicators to identify regimes and then apply three overlays: moves between 60/40 and cash, rotations between equities and bonds, and additions of commodities/gold. Laipply et al. [41] first estimate the correlation regime between stocks and bonds using a hidden Markov model, and then make allocations using mean-variance optimization based on the historic performance of the assets in the regime.

1.2 Outline

In §2 we set our notation and describe the two fixed-weight benchmarks, their volatility-controlled versions, and our two optimization-based portfolios. In §3 we present the performance metrics for the six portfolios, along with the performance of each of the assets (SPY, AGG, and GLD). In §4 we evaluate the portfolios taking into account inflation, and also taking into account a simple model of U.S. federal taxes. We conclude in §5. In appendix A we give the details of the accounting we use, and in appendix B we give the details of the return forecast used in our Markowitz portfolio. In appendix C we analyze the statistical significance of our results. In appendix D we explore the effect of lagging the data we use in our Markowitz portfolio. We address several questions related to our choice of the (few) hyper-parameters in our Markowitz portfolio in appendix E. Finally, in appendix F we benchmark our Markowitz portfolio against seven standard risk-based and Bayesian allocation methods, all of which it outperforms, and in appendix G we report the sensitivity of our results to the assumed trading cost and cash rate.

2 Portfolios

We first discuss our notation and then introduce the six portfolios.

2.1 Notation

We denote trading days (of which there are around 252 per year) by $t = 0, 1, 2, \ldots$. We consider portfolios that hold three assets (the ETFs SPY, AGG, and GLD), plus cash. The (positive) value of the portfolio on trading day t is denoted V_t, with initial value $V_0 = 1$. The weights of the noncash assets on trading day t are denoted w_t^{spy}, w_t^{agg}, and w_t^{gld}. These represent the fraction of the total portfolio value held in each noncash asset. They are nonnegative and satisfy

$$w_t^{\text{spy}} + w_t^{\text{agg}} + w_t^{\text{gld}} \leq 1, \quad t = 0, 1, 2, \ldots.$$

The cash weight, the fraction of the portfolio value held in cash, is

$$c_t = 1 - w_t^{\text{spy}} - w_t^{\text{agg}} - w_t^{\text{gld}}, \quad t = 0, 1, 2, \ldots.$$

We denote the portfolio as a weight vector $w_t = (w_t^{\text{spy}},\ w_t^{\text{agg}},\ w_t^{\text{gld}})$. Using vector notation we have, for example, $(w_t)_2 = w_t^{\text{agg}}$. We can express c_t as $1 - \mathbf{1}^T w_t$, where $\mathbf{1} = (1, 1, 1)$.

The relative asset weights are given by

$$\eta_t^{\text{spy}} = \frac{w_t^{\text{spy}}}{\mathbf{1}^T w_t}, \qquad \eta_t^{\text{agg}} = \frac{w_t^{\text{agg}}}{\mathbf{1}^T w_t}, \qquad \eta_t^{\text{gld}} = \frac{w_t^{\text{gld}}}{\mathbf{1}^T w_t},$$

whenever $w_t \neq 0$, *i.e.*, the portfolio is not fully invested in cash.

After each day, the portfolio value and asset weights w_t update according to the returns of the assets and the prevailing risk-free interest rate on that day. On a rebalance day, we trade to get w_t to a specified target weight vector $\hat{w}_t$, taking trading cost into account. §2.2–2.4 describe the different methods to compute the target weights. Details of the daily value and weight update are given in appendix A.

2.2 Fixed-weight benchmarks

We consider two fixed-weight benchmarks: the traditional 60/40 stock/bond portfolio, with target weight vector $(0.6, 0.4, 0.0)$, and the 50/30/20 stock/bond/gold portfolio, with target weight vector $(0.5, 0.3, 0.2)$. Following custom, these benchmark portfolios are rebalanced to their targets annually, on the last trading day of each calendar year. Between rebalances the weights drift with the realized returns of the assets. These fixed-weight benchmarks do not hold cash, and their relative weights after each rebalance are the target portfolio weights.

2.3 Volatility-controlled portfolios

From each fixed-weight benchmark we construct a volatility-controlled portfolio, which we rebalance monthly, on the last trading day of each month. The volatility-controlled portfolio dilutes the fixed-weight benchmark with cash to reduce the volatility to a given target value σ^{tgt}, without changing the relative asset weights compared to the fixed-weight benchmark. When t is a rebalance day, the weights are set to

$$\hat{w}_t = \begin{cases} (\sigma^{\text{tgt}}/\sigma_t) w^{\text{tgt}}, & \sigma_t > \sigma^{\text{tgt}} \\ w^{\text{tgt}} & \sigma_t \leq \sigma^{\text{tgt}}. \end{cases} \tag{1}$$

Here w^{tgt} is the fixed-weight benchmark portfolio (*i.e.*, either $(0.6, 0.4, 0.0)$ or $(0.5, 0.3, 0.2)$), and σ_t is its estimated volatility. The corresponding cash weight is $c_t = 0$ when $\sigma_t \leq \sigma^{\text{tgt}}$, and $c_t = 1 - \sigma^{\text{tgt}}/\sigma_t$ when $\sigma_t > \sigma^{\text{tgt}}$. It is obvious from (1) that a volatility-controlled benchmark preserves the relative weights at each rebalance.

We estimate the fixed-weight benchmark portfolio volatility σ_t as the empirical standard deviation of the benchmark's daily returns over a trailing window of 11 trading days, annualized by the factor $\sqrt{252}$. A lookback period of 11 days is short by the standards of volatility estimation, and gives a noisy estimate of σ_t; but it is more reactive to changes in volatility, and leads to better portfolio performance (after a modest search over various values). Other volatility estimates, such as the Parkinson volatility estimator [51] that requires low and high portfolio values, exist but we find that the simple trailing standard deviation works well in our setting. These volatility-controlled portfolios are dynamic portfolios, *i.e.*, they change over time, because σ_t varies with the realized returns of the assets.

2.4 Optimization-based portfolios

Optimization objective. Our optimization-based portfolios depend on an estimate of the future asset returns, given by the vector α_t, and an estimate of the asset return covariance matrix, denoted Σ_t, when t corresponds to a rebalance day. We express α_t as an estimate of next month's asset returns. Let r_t^{rf} be the contemporaneous daily federal funds rate compounded over 21 trading days, and let w_t denote the drifted risky-asset weights before the rebalance. For candidate risky-asset weights w, the corresponding cash weight is $1-\mathbf{1}^T w$, so the return on the cash is $r_t^{\mathrm{rf}}(1-\mathbf{1}^T w)$. The estimated total gross return is then

$$\alpha_t^T w + r_t^{\mathrm{rf}}(1-\mathbf{1}^T w).$$

We also take trading cost into account. We assume that the assets have the same bid-ask spread s, which we take as $s = 0.0005$ (5 basis points). With proposed weight w, the estimated trading cost for trading from w_t to w is then $(s/2)\|w - w_t\|_1$, where $\|a\|_1 = |a_1| + |a_2| + |a_3|$ is the ℓ_1 norm of a 3-vector a.

Our estimate of the portfolio return over the next month, net of the trading cost, is

$$\alpha_t^T w + r_t^{\mathrm{rf}}(1-\mathbf{1}^T w) - \frac{s}{2}\|w - w_t\|_1.$$

We will maximize this estimated next-month return net of trading cost.

Optimization constraints. We start with the obvious constraints

$$w^{\mathrm{spy}} \geq 0, \quad w^{\mathrm{agg}} \geq 0, \quad w^{\mathrm{gld}} \geq 0, \qquad w^{\mathrm{spy}} + w^{\mathrm{agg}} + w^{\mathrm{gld}} \leq 1,$$

which state that our asset and cash weights must be nonnegative. These are simple linear constraints on the variable w.

We also impose a constraint on portfolio risk. The estimated portfolio volatility is $(w^T\Sigma_t w)^{1/2}$. We limit this to be below a given target volatility σ^{tgt}, by imposing the constraint

$$(w^T\Sigma_t w)^{1/2} \leq \sigma^{\mathrm{tgt}}.$$

This is a convex constraint on the variable w.

We also limit the relative risky-asset weight deviation from $w^{\mathrm{tgt}} = (0.5, 0.3, 0.2)$. Writing $\eta = w/(\mathbf{1}^T w)$ when $\mathbf{1}^T w > 0$, we impose

$$\|\eta - w^{\mathrm{tgt}}\|_1 = |\eta^{\mathrm{spy}} - 0.5| + |\eta^{\mathrm{agg}} - 0.3| + |\eta^{\mathrm{gld}} - 0.2| \leq 1,$$

or, equivalently,

$$\|w - (\mathbf{1}^T w)w^{\mathrm{tgt}}\|_1 \leq \mathbf{1}^T w. \tag{2}$$

This is a convex constraint on w.

Optimization problem. On each rebalance day t, we find our target portfolio weights $\hat{w}_t$ by solving the optimization problem

$$\begin{array}{ll} \text{maximize} & \alpha_t^T w + r_t^{\text{rf}}(1 - \mathbf{1}^T w) - (s/2)\|w - w_t\|_1 \\ \text{subject to} & w^{\text{spy}} \geq 0, \;\; w^{\text{agg}} \geq 0, \;\; w^{\text{gld}} \geq 0 \\ & w^{\text{spy}} + w^{\text{agg}} + w^{\text{gld}} \leq 1, \\ & \left\| w - (\mathbf{1}^T w) w^{\text{tgt}} \right\|_1 \leq \mathbf{1}^T w, \\ & (w^T \Sigma_t w)^{1/2} \leq \sigma^{\text{tgt}}, \end{array} \tag{3}$$

with variable $w \in \mathbf{R}^3$. Here we choose the weights to maximize the estimated next-month return net of the estimated cost of the rebalance while limiting the estimated risk to not exceed the target value and keeping the relative risky-asset allocation within the stated ℓ_1 distance of the 50/30/20 portfolio. This is a small convex optimization problem [12] and a variant of the classical Markowitz portfolio optimization problem [44]. We model the problem in CVXPY [23] and solve it with the Clarabel solver [30]. This very small problem can be solved in well under one millisecond.

Optimization problem data. The portfolio problem (3) requires the data α_t, r_t^{rf}, s, w_t, Σ_t, and σ^{tgt}. The target volatility σ^{tgt} is specified. We take the bid-ask spread to be $s = 0.0005$. We take the estimated covariance Σ_t to be the empirical covariance matrix of the three assets' daily returns over a trailing window of 11 trading days, annualized. As with our estimates of the volatilities of the fixed-weight benchmarks, this gives a noisy but reactive estimate, which improves portfolio performance. We use two different methods to estimate α_t, one very simple and one somewhat less simple.

Simple Markowitz. For the simple Markowitz portfolio, we take α_t to be the exponentially weighted moving average of each asset's past daily returns, with a half-life of 252 trading days (one year), multiplied by 21 to put it on the next-month scale. This is a simple momentum signal: it favors assets that have performed well over the trailing year, and it uses only the return history.

Markowitz. For our Markowitz portfolio, α_t is a return forecast based on previous asset returns, and in addition some quantities that are publicly available on day t. These fall into three groups.

- *Asset-specific features*, such as the trailing returns, trading volumes, and volatilities.
- *Macro-financial features*, such as yields, inflation rates, and the VIX measure of expected U.S. equity-market volatility.
- *Fama-French factors* [26], such as the market excess return.

The forecast is produced by a rolling regression that links these quantities to the assets' future returns; its construction is causal (at time t, the forecast only depends on data available at time t) and is described in appendix B. Our estimation method is simple, transparent, and interpretable, unlike more complex methods such as random forests or neural networks.

2.5 Data and simulation

We evaluate the six portfolios over the 20 years from Jan. 1st 2006 to Jan. 1st 2026. Daily adjusted close prices for SPY, AGG, and GLD were obtained from Yahoo Finance; the federal funds rate (FRED series `DFF`) and core CPI (`CPILFESL`) are downloaded from FRED. Our simulation tracks daily portfolio value, and includes a trading cost of 5 basis points (a 2.5 basis point half-spread) on asset trading, charged against the portfolio value at each rebalance. The optimization-based portfolios anticipate this same cost in their objective. In appendix G we report the sensitivity of our results to the assumed trading cost and to the rate at which cash accrues.

2.6 Metrics

We report six metrics, all computed from the portfolio value time series. *Annualized return* is the compound annual growth rate (CAGR), *i.e.*, the constant annual rate of growth that takes the portfolio from its initial to its final value over the simulation. *Annualized volatility* is the standard deviation of the daily returns, annualized by the factor $\sqrt{252}$. The *Sharpe ratio* is the compound annual growth rate of the portfolio value relative to the compounded federal funds rate, divided by the annualized volatility. This is a geometric, CAGR-based measure, and not the conventional Sharpe ratio, which uses the arithmetic mean of the periodic excess returns. The two agree closely here: over the full sample the conventional Sharpe ratios are 1.08 for the Markowitz portfolio (against the 1.08 we report), 0.92 (0.91) for the simple Markowitz portfolio, 0.83 (0.82) for the 50/30/20 VC portfolio, 0.73 (0.71) for the 60/40 VC portfolio, 0.73 (0.70) for 50/30/20, and 0.60 (0.56) for 60/40. The conventional measure is slightly more favorable to the higher-volatility benchmarks, so the margins we report are, if anything, mildly conservative. *Maximum drawdown* (max DD) is the largest percentage decline of the portfolio value from its previous running peak, *i.e.*, the worst peak-to-trough loss. *Mean drawdown* (mean DD) is the average percentage decline from the running peak over time. *Turnover* is half the total absolute change in the non-cash weights, so that selling one asset to buy another counts the exchanged value once, averaged over trading days and annualized.

3 Results

In all results reported below we use a common annualized volatility target of $\sigma^{\mathrm{tgt}} = 7\%$ for both volatility-controlled portfolios and the optimization-based portfolios, so that the portfolios are compared at the same level of targeted risk. (The realized risks are, however, a bit different.)

3.1 Metrics

Cumulative return. Figure 1 shows the cumulative return of the six portfolios, on a logarithmic vertical scale, so a straight line corresponds to a constant rate of return. The two volatility-controlled portfolios track visibly smoother paths than their fixed-weight counterparts, particularly through the 2008 and 2020 drawdowns, and the optimization-based

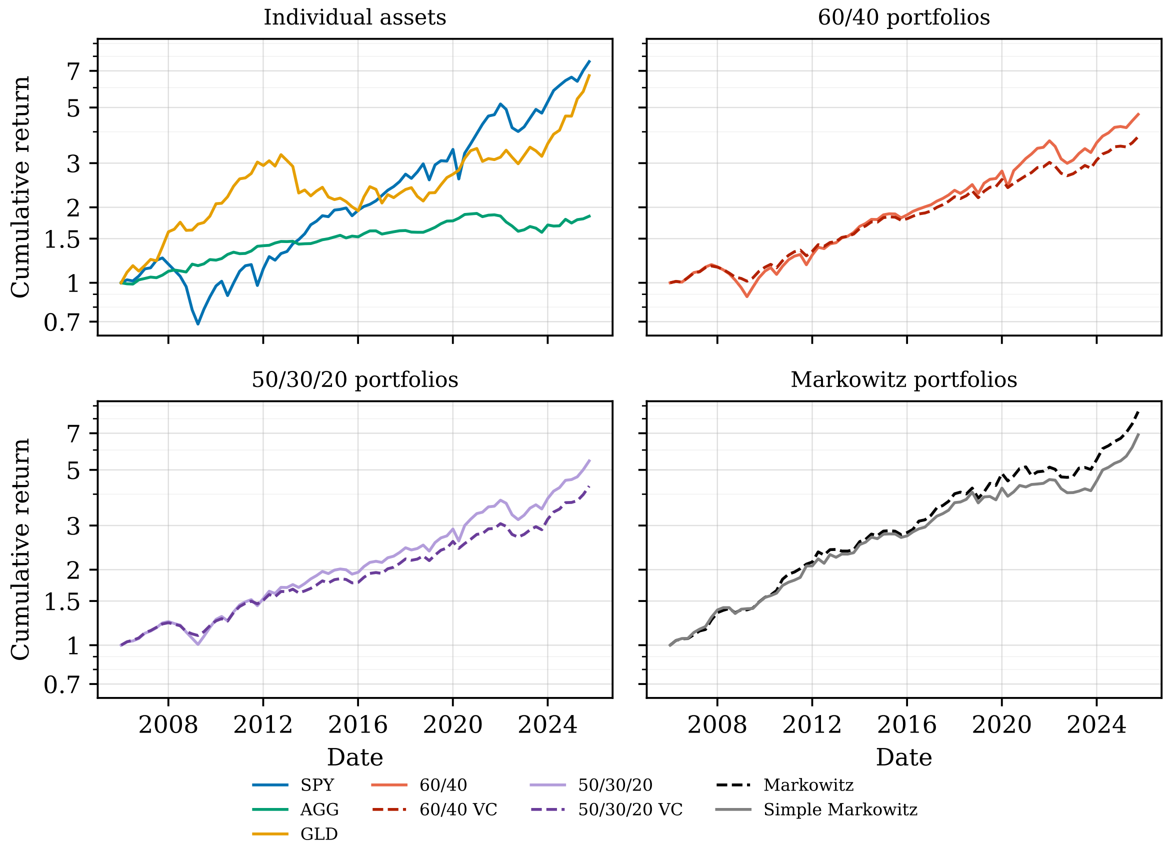


Figure 1: Cumulative return of the three assets (top left) and the six portfolios. VC corresponds to volatility control.

portfolios dominate all four benchmarks over the full period. Both optimization-based portfolios have a positive return in 2008, of 2.6% for the Markowitz portfolio and 1.2% for the simple Markowitz portfolio.

Metrics. Table 1 reports the metrics for the six portfolios. We can see that volatility control results in realized volatility near the target value and reduces maximum and mean drawdown, at the cost of lower return but higher Sharpe ratio, despite higher turnover. The optimization-based portfolios improve the risk-adjusted returns further, with higher Sharpe ratios and drawdowns well below the fixed-weight benchmarks. The Markowitz portfolio achieves a Sharpe ratio of 1.08, a substantial step up from both the fixed-weight benchmarks and the volatility-controlled portfolios. Accounting for realized risk, it delivers nearly twice the excess return of the 60/40 benchmark, which has Sharpe ratio 0.56. The simple Markowitz portfolio lands between the volatility-controlled benchmarks and the Markowitz portfolio, with a Sharpe ratio of 0.91. The volatility-controlled and optimization-based portfolios have substantially higher turnover compared to the annually rebalanced fixed-weight benchmarks; we remind the reader that all of the results we report take into account trading costs. The same metrics for seven standard risk-based allocation methods, run on the same assets and data, are given in appendix F; all of them fall well short of the Markowitz

Portfolio	Return	Volatility	Sharpe	Max DD	Mean DD	Turnover
Markowitz	11.6%	9.0%	1.08	18.1%	3.0%	284.4%
Simple Markowitz	10.4%	9.3%	0.91	15.8%	3.1%	278.9%
50/30/20 VC	7.8%	7.3%	0.82	15.9%	2.4%	79.0%
60/40 VC	7.1%	7.4%	0.71	16.9%	2.6%	85.0%
50/30/20	9.1%	10.3%	0.70	27.1%	3.0%	4.0%
60/40	8.1%	11.3%	0.56	33.7%	3.8%	3.4%
GLD	10.6%	17.9%	0.49	45.6%	17.0%	0.0%
AGG	3.1%	5.3%	0.26	18.4%	2.9%	0.0%
SPY	10.8%	19.4%	0.46	55.2%	7.5%	0.0%
Cash	1.7%	0.0%	0.00	0.0%	0.0%	0.0%

Table 1: Performance metrics for the six portfolios (top) and the individual assets and cash (bottom).

portfolio.

A detailed analysis of the statistical significance of these results is given in appendix C.

Consistency. Table 1 gives the aggregate performance of the portfolios over many years. One important additional metric not captured in these aggregate values is the consistency of performance over time, through different market regimes. We illustrate this by examining the annual realized volatilities of the six portfolios, shown in figure 2. We see that volatility control leads to much more consistent realized volatility, as judged by the variability of realized calendar year volatility; with the volatility-controlled and Markowitz portfolios all comparably consistent. The two fixed-weight benchmarks experience spikes in volatility, ranging from around 4% to above 20%. The volatility-controlled and Markowitz portfolios bring this range down, from around 4% to 12%.

It is well known that return-forecasting strategies can exhibit decreasing performance over long periods of time, presumably as other traders learn the (small) arbitrage opportunities being exploited [45]. Our Markowitz portfolio shows no clear sign of this phenomenon over our simulation period. Its annualized returns in the four consecutive five-year periods are 14.0%, 8.0%, 12.5%, and 11.8%, at annualized volatilities of 8.4%, 9.2%, 9.6%, and 8.6%, for Sharpe ratios of 1.34, 0.86, 1.16, and 0.97; the Sharpe ratios of all six portfolios by subperiod are given in table 2.

3.2 Weights over time

Figure 3 shows the weights for the six portfolios over time. The fixed-weight benchmarks (panels 3a and 3b) drift a bit between annual rebalances; the volatility-controlled variants (panels 3c and 3d) visibly scale down during high-volatility market regimes, holding more in cash; and the Markowitz portfolio (panel 3f) shifts weight dynamically across all three assets each month, as does the simple Markowitz portfolio (panel 3e).

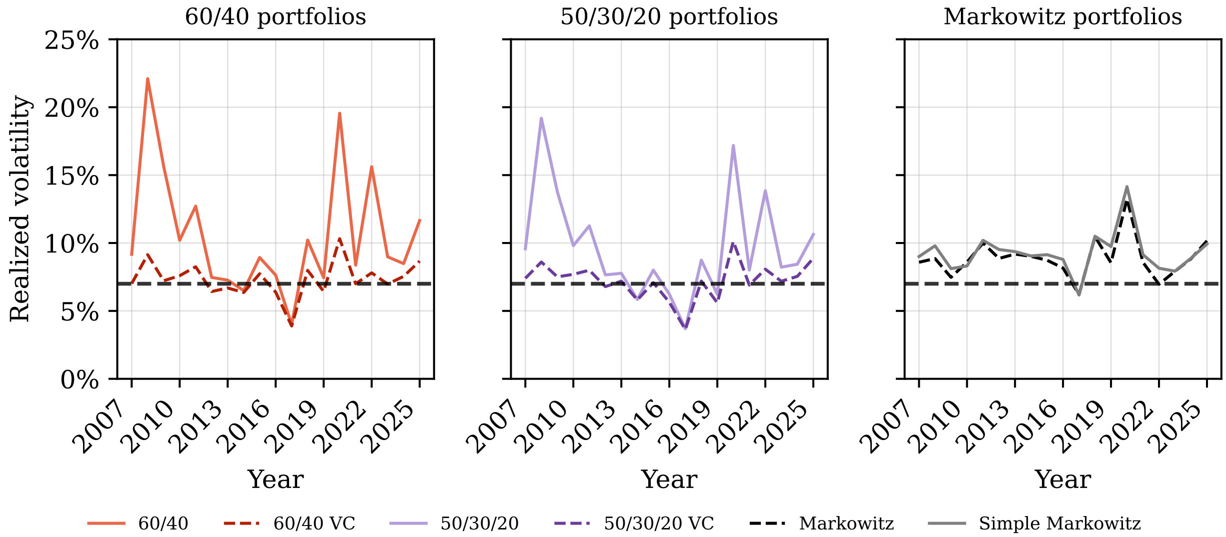
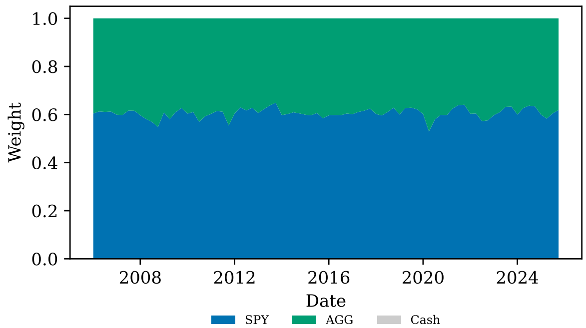


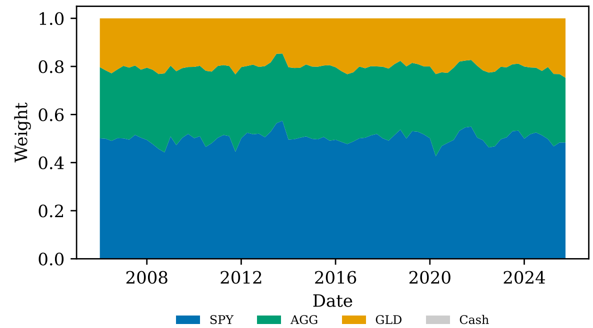
Figure 2: Calendar-year realized volatilities.

Portfolio	2006–2011	2011–2016	2016–2021	2021–2026	2006–2026
Markowitz	1.34	0.86	1.16	0.97	1.08
Simple Markowitz	1.11	0.94	0.82	0.81	0.91
50/30/20 VC	0.63	0.65	1.17	0.84	0.82
60/40 VC	0.35	0.99	0.96	0.57	0.71
50/30/20	0.40	0.75	1.05	0.75	0.70
60/40	0.13	0.99	0.87	0.48	0.56

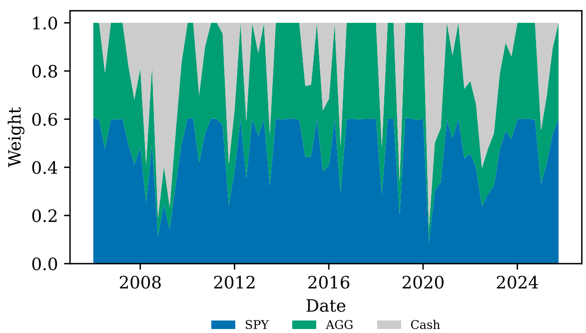
Table 2: Sharpe ratios of the six portfolios by five-year subperiods, and over the full period.



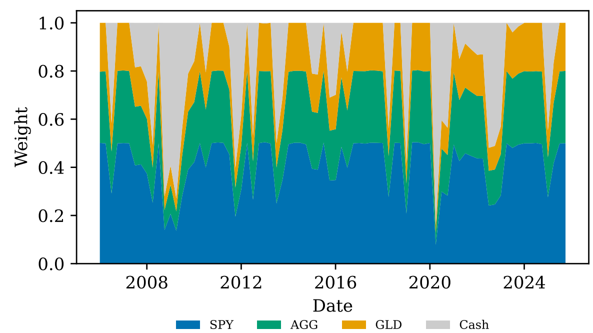
(a) 60/40



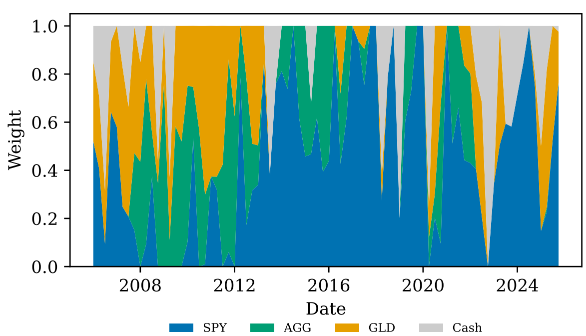
(b) 50/30/20



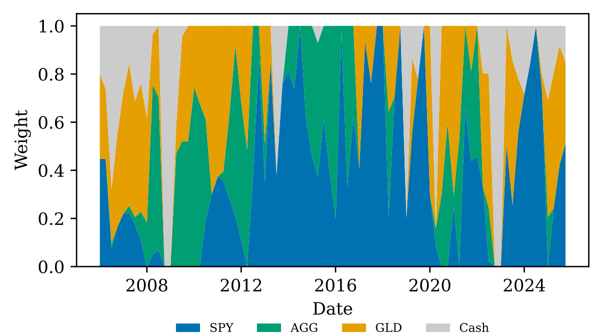
(c) 60/40 VC



(d) 50/30/20 VC



(e) Simple Markowitz



(f) Markowitz

Figure 3: Portfolio weights over time.

Portfolio	SPY	AGG	GLD	Cash
Markowitz	41.5%	20.1%	23.3%	15.1%
Simple Markowitz	46.5%	19.1%	21.2%	13.2%
50/30/20 VC	42.0%	25.1%	16.8%	16.2%
60/40 VC	48.9%	32.4%	0.0%	18.7%
50/30/20	50.3%	29.2%	20.6%	0.0%
60/40	60.8%	39.2%	0.0%	0.0%

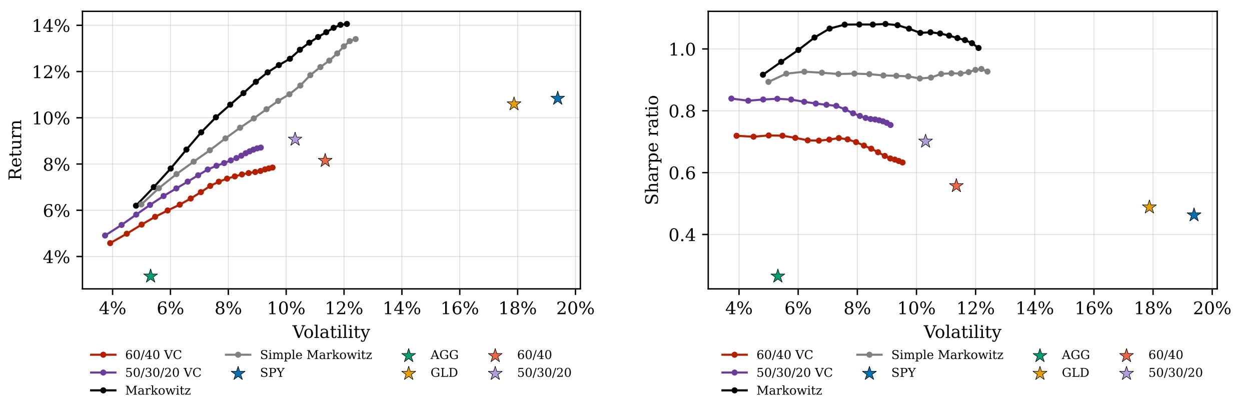
Table 3: Average weights over time for the six portfolios, averaged over all trading days in the simulation.

3.3 Average weights over time

Table 3 gives the average weights over time for each portfolio. Interestingly, the Markowitz portfolio, which has the best performance, holds about 15% of its value in cash on average, similar to the volatility-controlled benchmarks. A fixed-weight portfolio with these average weights, however, does poorly: rebalanced annually, it realizes a return of 8.5% at a volatility of 9.1%, for a Sharpe ratio of 0.73—at essentially the same volatility as the Markowitz portfolio, but well below its Sharpe ratio of 1.08. The benefit of the Markowitz portfolio is therefore not simply in holding more cash, but in varying the cash holdings (as well as the mix of stocks, bonds, and gold) strategically.

3.4 Risk-return trade-off

In results reported above we target an annualized volatility of 7%. Here we study the risk-return trade-off by varying the volatility target in the volatility-controlled and optimization-based portfolios, over the range 3% to 12%, in 0.5% increments. Figure 4a shows realized return versus realized volatility for these three portfolios, and figure 4b shows Sharpe ratio versus realized volatility. We can see that the Markowitz portfolio dominates the volatility-controlled benchmarks across the entire range of volatility targets, both in realized return at a given level of risk, and in Sharpe ratio. The raw assets and the fixed-weight benchmarks are shown as single points; all of them lie well below the Markowitz frontier in both panels.



(a) Realized return versus realized volatility.

(b) Sharpe ratio versus realized volatility.

Figure 4: Performance of portfolios as the risk target is varied from 3% to 12% in 0.5% increments. Raw assets and benchmarks without volatility control are shown as single stars.

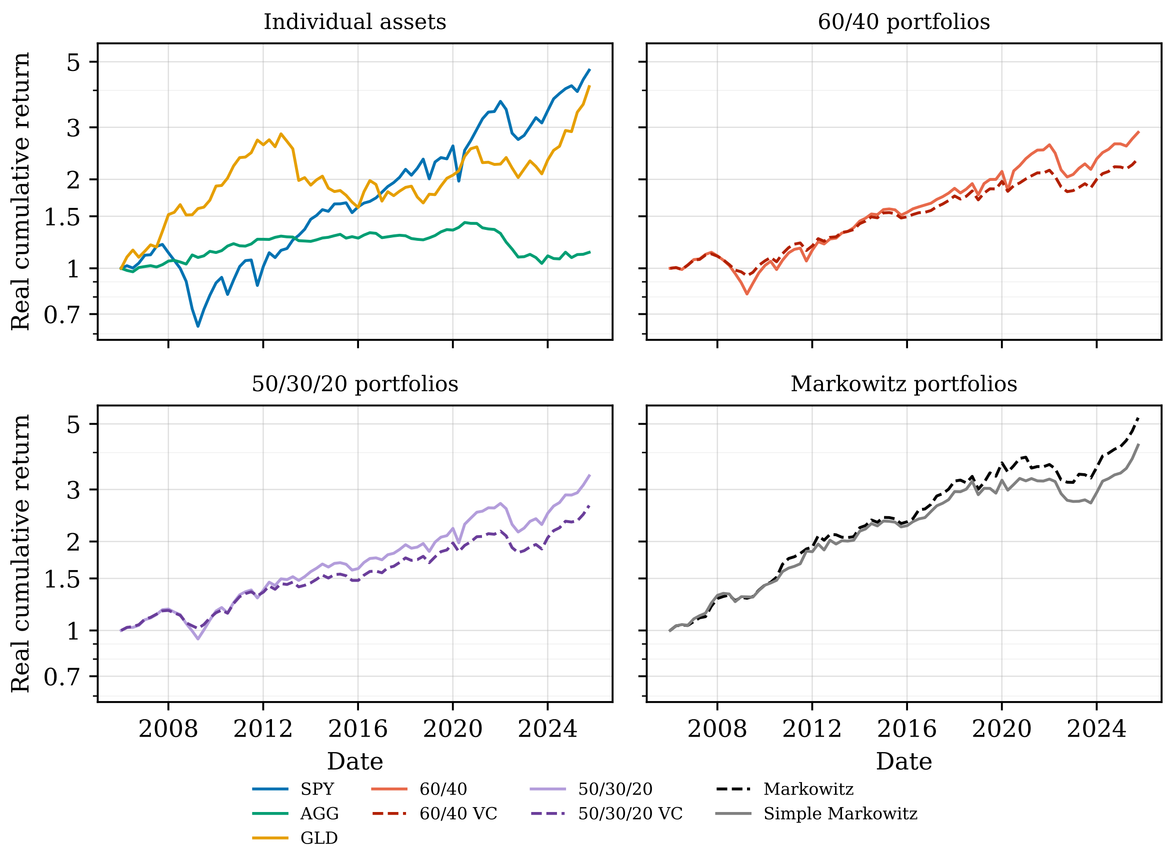


Figure 5: CPI-adjusted (real) cumulative return of the three assets (top left) and the six portfolios.

4 Inflation and taxes

In this section we examine the post-inflation and post-tax performance of the six portfolios.

4.1 Inflation-adjusted performance

We deflate the portfolio value by core CPI (FRED series `CPILFESL`) and recompute the cumulative return and Sharpe ratio, the latter now measuring return in excess of realized inflation rather than the federal funds rate.

Figure 5 shows the CPI-adjusted (real) cumulative return of the three assets and the six portfolios—the inflation-adjusted analog of figure 1—and table 4 reports the realized annualized return and realized Sharpe ratio. Inflation lowers every return, but the ranking is essentially unchanged. The Markowitz portfolio retains the highest real Sharpe ratio, 0.99; its real return of 8.8% exceeds that of raw equity (SPY, 8.1%) and gold (GLD, 7.9%) while realizing less than half their volatility. The advantage of the Markowitz portfolio over the fixed-weight and volatility-controlled benchmarks thus persists after adjusting for inflation.

Portfolio	Return	Sharpe
Markowitz	8.8%	0.99
Simple Markowitz	7.7%	0.82
50/30/20 VC	5.1%	0.70
60/40 VC	4.4%	0.60
50/30/20	6.4%	0.62
60/40	5.5%	0.49
GLD	7.9%	0.44
AGG	0.6%	0.12
SPY	8.1%	0.42

Table 4: CPI-adjusted annualized return and Sharpe ratio, both in excess of realized inflation, for the six portfolios and the three assets.

4.2 Tax adjusted performance

The results given so far are pre-tax. In a taxable account, rebalancing realizes capital gains (or losses), and since our dynamic portfolios trade far more than the annually rebalanced benchmarks, we ask how much of their advantage survives tax. We estimate post-tax performance using a simple accounting model described below.

Tax brackets. We consider four brackets, each with a pair of marginal rates $(\tau^{\mathrm{lt}}, \tau^{\mathrm{st}})$: the long-term rate τ^{lt} on gains from positions held more than one year, and the short-term (ordinary-income) rate τ^{st} on positions held one year or less. They correspond to representative U.S. federal single-filer rates:

- **B1**, (0%, 0%): a tax-advantaged or tax-exempt account (*e.g.*, an IRA or 401(k)), or the lowest bracket. This corresponds to ignoring taxes, *i.e.*, the pre-tax results.
- **B2**, (15%, 22%): a middle-income investor.
- **B3**, (15%, 35%): a high-income investor.
- **B4**, (20%, 37%): the top bracket.

Two adjustments apply on top of these headline rates. The net investment income tax (NIIT) adds a 3.8% surtax on investment income above a modified adjusted gross income of $200,000 for a single filer, so it applies in B3 and B4 but not in B2; the effective pairs are therefore (18.8%, 38.8%) and (23.8%, 40.8%). Also, GLD is a grantor trust holding physical metal, so a long-term gain on its shares is a collectibles gain, taxed at ordinary rates capped at 28% rather than at the rate that applies to stock. Its long-term rate is thus 22% in B2 and 31.8% in both B3 and B4.

Accounting. We track individual tax lots. Each purchase opens a lot recording its acquisition date and cost basis, and each sale realizes a gain or loss (proceeds minus basis), classified as long-term or short-term by whether the lot was held more than one year. Whenever a transaction realizes a gain or loss at a rebalance, we deduct the tax liability immediately from the cash account. (In a more realistic simulation the taxes for the preceding calendar year would be deducted from the cash account each April.) Our results are therefore conservative: settling at once rather than the following April gives up the use of the money in the interim, and the more a portfolio trades the more of that deferral it forgoes. Interest earned on cash (at the federal funds rate) is taxed as ordinary income, *i.e.*, at the short-term rate rather than the long-term rate.

Our simulations run on adjusted close prices, whose returns reinvest every distribution, so a model that taxed only realized capital gains would silently defer dividend and coupon income into capital appreciation. We therefore separate the income component of each asset's daily total return and tax it in the period received: SPY's distributions as qualified dividends, at the long-term rate, and AGG's as interest, at the ordinary rate. (GLD makes no distributions.) The basis of the holding is stepped up by the amount taxed, so the same dollars are not taxed a second time when the position is sold. Over 2006–2025 this income averaged 1.9% a year for SPY and 3.1% for AGG.

We also apply the wash-sale rule. A realized loss is deductible only if no purchase of the same asset falls within 30 calendar days either side of the sale; otherwise the loss is disallowed and added to the basis of the replacement shares. Since consecutive monthly rebalances are 28 to 31 days apart, a large share of our realized losses fall inside that window. We ignore state and local taxes.

When we sell assets, we use a simple greedy rule to determine which lots to sell. We choose the lots so as to minimize the immediate tax liability. This depends on the basis of the lots, their short-term or long-term status, and the two marginal tax rates. We start from cash, and at the end of the simulation horizon we liquidate the entire portfolio and tax the remaining unrealized gains, so even a zero-turnover holding (buy-and-hold SPY) is taxed, since its accumulated gain must eventually be realized.

A tax-aware investor could do better, by harvesting losses and deferring gains more aggressively [47]; our simple analyses here estimate the post-tax performance of a naïve investor who simply runs each strategy and pays the resulting taxes. Also, the tax analysis assumes that the investor is carrying out the rebalancing. If the portfolios were instead offered as ETFs, the investor would only pay taxes when shares of it are sold. This tax advantage to the investor would presumably exceed ETF fees.

Results. Tables 5 and 6 report the post-tax annualized return and Sharpe ratio (in excess of the federal funds rate) under the four brackets. For the Sharpe ratio we hold volatility fixed at its pre-tax (B1) value, since the tax drag is a largely deterministic reduction in return, not investment risk; the B1 column thus reproduces table 1 exactly.

Three patterns stand out. First, the tax drag grows with turnover: the annually rebalanced benchmarks and raw assets realize gains slowly and mostly at long-term rates, so SPY (zero turnover) loses about 1.6 points of return at the top bracket—most of it from the terminal liquidation—whereas the volatility-controlled and Markowitz portfolios (turnover

Portfolio ($\tau^{\rm lt}/\tau^{\rm st}$)	B1 (0/0)	B2 (15/22)	B3 (15/35)	B4 (20/37)
Markowitz	11.6%	9.2%	7.8%	7.5%
Simple Markowitz	10.4%	8.4%	7.3%	7.0%
50/30/20 VC	7.8%	6.5%	5.9%	5.6%
60/40 VC	7.1%	5.9%	5.5%	5.2%
50/30/20	9.1%	7.9%	7.4%	7.2%
60/40	8.1%	7.1%	6.7%	6.4%
GLD	10.6%	9.4%	8.8%	8.8%
AGG	3.1%	2.5%	1.9%	1.9%
SPY	10.8%	9.9%	9.6%	9.2%

Table 5: Post-tax annualized return for the six portfolios and the three assets under the four tax brackets B1–B4, with ($\tau^{\rm lt}/\tau^{\rm st}$) the long-term and short-term marginal rates in percent.

Portfolio ($\tau^{\rm lt}/\tau^{\rm st}$)	B1 (0/0)	B2 (15/22)	B3 (15/35)	B4 (20/37)
Markowitz	1.08	0.83	0.67	0.64
Simple Markowitz	0.91	0.71	0.59	0.56
50/30/20 VC	0.82	0.64	0.56	0.53
60/40 VC	0.71	0.56	0.50	0.46
50/30/20	0.70	0.59	0.55	0.52
60/40	0.56	0.47	0.43	0.41
GLD	0.49	0.42	0.39	0.39
AGG	0.26	0.14	0.04	0.03
SPY	0.46	0.41	0.40	0.38

Table 6: Post-tax Sharpe ratio (in excess of the federal funds rate) for the six portfolios and the three assets under the four tax brackets B1–B4. The volatility in the denominator is held fixed at its pre-tax (B1) value.

79%–284%) give up much more. Second, income matters as much as turnover for the bond-heavy positions. AGG distributes about 3.1% a year as interest, taxed at the ordinary rate, so its post-tax return falls from 3.1% to 1.9% and its Sharpe ratio from 0.26 to 0.03 at the top bracket. The zero-turnover buy-and-hold benchmark is not the tax-efficient one; the tax-efficient asset is the one that pays no income, which here is gold. Third, the Markowitz portfolio's advantage narrows under taxation but remains clear: its Sharpe ratio falls from 1.08 to 0.83, 0.67, and 0.64 across B2–B4, yet it stays the highest in every bracket—at the top bracket it reaches 0.64, above the 0.56 of the next-best portfolio. Its edge thus survives taxation, though a high-bracket investor would weigh its higher turnover more heavily.

5 Conclusions and discussion

Return forecast. We do not believe that our return forecasting method is the best possible; we consider it merely good enough to provide a performance lift. More sophisticated return forecasting could certainly be developed, even using the same features that we use. We also imagine that further improvement would come from using more exotic (even if still public) data sources.

Timing the market? As shown in §3.2, the average weights of the Markowitz portfolio does not by itself explain its performance: a fixed-weight portfolio that statically holds those average weights yields a Sharpe ratio of only 0.73, well below the 1.08 of the dynamic portfolio at similar volatility. The extra performance comes from how the portfolio varies its exposure over time, shifting among stocks, bonds, gold, and cash from month to month in response to the return forecast and the risk model—which is what one would call market timing. Consistent with this, the portfolio de-risks into cash during turbulent periods: it is the only portfolio we consider with a positive Sharpe ratio through the 2008 crisis, and its realized volatility stays in a much narrower band than the fixed-weight benchmarks (figure 2).

Comparison with other methods. In appendix F we benchmark our portfolios against seven standard risk-based allocation methods—among them risk parity, minimum variance, maximum diversification, and Black–Litterman—run on the same assets, data, costs, and constraints. All of them land near our simple volatility-controlled benchmark, with Sharpe ratios between 0.60 and 0.85, and well below the 1.08 of the Markowitz portfolio. Sizing positions by risk is evidently not enough on its own; the lift comes from combining a return forecast with a hard limit on estimated risk.

Acknowledgments

We thank David Demers, Emmanuel J. Candès, Trevor Hastie, Mykel J. Kochenderfer, Logan Bell, Alexander Karolin, Oskar Wallberg, and Daniel Cederberg for many helpful discussions. We are especially grateful to Rishi Narang, Ronald N. Kahn, and Maximilian Schaller for their detailed comments and suggestions.

References

[1] Clifford S. Asness, Andrea Frazzini, and Lasse H. Pedersen. Leverage Aversion and Risk Parity. *Financial Analysts Journal*, 68(1):47–59, January 2012.

[2] David H. Bailey, Jonathan M. Borwein, Marcos López de Prado, and Qiji Jim Zhu. The Probability of Backtest Overfitting. *The Journal of Computational Finance*, 20(4):39–69, April 2017.

[3] David H. Bailey and Marcos López de Prado. The Deflated Sharpe Ratio: Correcting for Selection Bias, Backtest Overfitting, and Non-Normality. *The Journal of Portfolio Management*, 40(5):94–107, September 2014.

[4] Pedro Barroso and Pedro Santa-Clara. Momentum Has Its Moments. *Journal of Financial Economics*, 116(1):111–120, April 2015.

[5] Dirk G. Baur and Brian M. Lucey. Is Gold a Hedge or a Safe Haven? An Analysis of Stocks, Bonds and Gold. *Financial Review*, 45(2):217–229, April 2010.

[6] Dirk G. Baur and Thomas K. McDermott. Is Gold a Safe Haven? International Evidence. *Journal of Banking & Finance*, 34(8):1886–1898, August 2010.

[7] William J. Bernstein. *The Intelligent Asset Allocator: How to Build Your Portfolio to Maximize Returns and Minimize Risk*. McGraw-Hill Education, 2000.

[8] Fischer Black and Robert Litterman. Global Portfolio Optimization. *Financial Analysts Journal*, 48(5):28–43, September 1992.

[9] John C. Bogle. *The Little Book of Common Sense Investing: The Only Way to Guarantee Your Fair Share of Stock Market Returns*. Little Books. Big Profits. Wiley, 2017.

[10] Stephen Boyd, Steven Diamond, Kwangmoo Koh, Peter Nystrup, Enzo Busseti, Ronald N. Kahn, and Jan Speth. Multi-Period Trading via Convex Optimization. *Foundations and Trends in Optimization*, 3(1):1–76, August 2017.

[11] Stephen Boyd, Kasper Johansson, Ronald Kahn, Philipp Schiele, and Thomas Schmelzer. Markowitz Portfolio Construction at Seventy. *The Journal of Portfolio Management*, 50(8):117–160, June 2024.

[12] Stephen Boyd and Lieven Vandenberghe. *Convex Optimization*. Cambridge University Press, 2004.

[13] Gary P. Brinson, L. Randolph Hood, and Gilbert L. Beebower. Determinants of Portfolio Performance. *Financial Analysts Journal*, 42(4):39–44, July 1986.

[14] Gary P. Brinson, Brian D. Singer, and Gilbert L. Beebower. Determinants of Portfolio Performance II: An Update. *Financial Analysts Journal*, 47(3):40–48, May 1991.

[15] Alfie Brixton, Jordan Brooks, Pete Hecht, Antti Ilmanen, Thomas Maloney, and Nicholas McQuinn. A Changing Stock–Bond Correlation: Drivers and Implications. *The Journal of Portfolio Management*, 49(4):64–80, January 2023.

[16] Benjamin Bruder and Thierry Roncalli. Managing Risk Exposures Using the Risk Budgeting Approach. Working paper, Lyxor Asset Management, 2012.

[17] John Y. Campbell, Adi Sunderam, and Luis M. Viceira. Inflation Bets or Deflation Hedges? The Changing Risks of Nominal Bonds. *Critical Finance Review*, 6(2):263–301, September 2017.

[18] Scott Cederburg, Michael S. O'Doherty, Feifei Wang, and Xuemin (Sterling) Yan. On the Performance of Volatility-Managed Portfolios. *Journal of Financial Economics*, 138(1):95–117, October 2020.

[19] Yves Choueifaty and Yves Coignard. Toward Maximum Diversification. *The Journal of Portfolio Management*, 35(1):40–51, October 2008.

[20] Roger G. Clarke, Harindra de Silva, and Steven Thorley. Minimum-Variance Portfolios in the U.S. Equity Market. *The Journal of Portfolio Management*, 33(1):10–24, October 2006.

[21] Victor DeMiguel, Lorenzo Garlappi, and Raman Uppal. Optimal Versus Naive Diversification: How Inefficient Is the 1/N Portfolio Strategy? *The Review of Financial Studies*, 22(5):1915–1953, May 2009.

[22] Nikhil Devanathan, Dylan Rueter, Stephen Boyd, Emmanuel Candès, Trevor Hastie, Mykel J. Kochenderfer, Arpit Apoorv, David Soronow, and Igor Zamkovsky. Single-Asset Adaptive Leveraged Volatility Control. arXiv:2603.01298v2 [q-fin.PM], March 2026.

[23] Steven Diamond and Stephen Boyd. CVXPY: A Python-Embedded Modeling Language for Convex Optimization. *Journal of Machine Learning Research*, 17(83):1–5, 2016.

[24] Claude B. Erb and Campbell R. Harvey. The Golden Dilemma. *Financial Analysts Journal*, 69(4):10–42, July 2013.

[25] Mebane T. Faber. A Quantitative Approach to Tactical Asset Allocation. *The Journal of Wealth Management*, 9(4):69–79, January 2007.

[26] Eugene F. Fama and Kenneth R. French. Common Risk Factors in the Returns on Stocks and Bonds. *Journal of Financial Economics*, 33(1):3–56, February 1993.

[27] Laurence D. Fink. 2025 Annual Chairman's Letter to Investors. Annual letter, BlackRock, Inc., April 2025.

[28] Jeff Fleming, Chris Kirby, and Barbara Ostdiek. The Economic Value of Volatility Timing. *The Journal of Finance*, 56(1):329–352, February 2001.

[29] Nicolae Gârleanu and Lasse Heje Pedersen. Dynamic Trading with Predictable Returns and Transaction Costs. *The Journal of Finance*, 68(6):2309–2340, December 2013.

[30] Paul J. Goulart and Yuwen Chen. Clarabel: An Interior-Point Solver for Conic Programs with Quadratic Objectives. *Mathematical Programming Computation*, May 2026.

[31] Shihao Gu, Bryan Kelly, and Dacheng Xiu. Empirical Asset Pricing via Machine Learning. *The Review of Financial Studies*, 33(5):2223–2273, February 2020.

[32] Campbell R. Harvey, Edward Hoyle, Russell Korgaonkar, Sandy Rattray, Matthew Sargaison, and Otto Van Hemert. The Impact of Volatility Targeting. *The Journal of Portfolio Management*, 45(1):14–33, October 2018.

[33] Robert A. Haugen and Nardin L. Baker. The Efficient Market Inefficiency of Capitalization-Weighted Stock Portfolios. *The Journal of Portfolio Management*, 17(3):35–40, April 1991.

[34] Guangliang He and Robert Litterman. The Intuition Behind Black-Litterman Model Portfolios. Working paper, Goldman Sachs Investment Management Research, 1999.

[35] David Hillier, Paul Draper, and Robert Faff. Do Precious Metals Shine? An Investment Perspective. *Financial Analysts Journal*, 62(2):98–106, March 2006.

[36] Alexandre Hocquard, Sunny Ng, and Nicolas Papageorgiou. A Constant-Volatility Framework for Managing Tail Risk. *The Journal of Portfolio Management*, 39(2):28–40, January 2013.

[37] Brian Hurst, Yao Hua Ooi, and Lasse Heje Pedersen. A Century of Evidence on Trend-Following Investing. *The Journal of Portfolio Management*, 44(1):15–29, October 2017.

[38] Antti Ilmanen. Stock-Bond Correlations. *The Journal of Fixed Income*, 13(2):55–66, September 2003.

[39] Min Jeong Kim and Dohyoung Kwon. Dynamic Asset Allocation Strategy: An Economic Regime Approach. *Journal of Asset Management*, 24(2):136–147, 2023.

[40] Chris Kirby and Barbara Ostdiek. It's All in the Timing: Simple Active Portfolio Strategies that Outperform Naïve Diversification. *Journal of Financial and Quantitative Analysis*, 47(2):437–467, April 2012.

[41] Steve Laipply and Ananth Madhavan. Can Bonds Still Diversify Multi-Asset Portfolios? Income Versus Duration in Distinct Correlation Regimes. *Income versus Duration in Distinct Correlation Regimes (September 15, 2025)*, 2025.

[42] Raul Leote de Carvalho, Xiao Lu, and Pierre Moulin. Demystifying Equity Risk-Based Strategies: A Simple Alpha Plus Beta Description. *The Journal of Portfolio Management*, 38(3):56–70, April 2012.

[43] Sébastien Maillard, Thierry Roncalli, and Jérôme Teïletche. The Properties of Equally Weighted Risk Contribution Portfolios. *The Journal of Portfolio Management*, 36(4):60–70, July 2010.

[44] Harry Markowitz. Portfolio Selection. *The Journal of Finance*, 7(1):77–91, March 1952.

[45] R. David McLean and Jeffrey Pontiff. Does Academic Research Destroy Stock Return Predictability? *The Journal of Finance*, 71(1):5–32, January 2016.

[46] Richard O. Michaud. The Markowitz Optimization Enigma: Is "Optimized" Optimal? *Financial Analysts Journal*, 45(1):31–42, January 1989.

[47] Nicholas Moehle, Mykel J. Kochenderfer, Stephen Boyd, and Andrew Ang. Tax-Aware Portfolio Construction via Convex Optimization. *Journal of Optimization Theory and Applications*, 189(2):364–383, May 2021.

[48] Alan Moreira and Tyler Muir. Volatility-Managed Portfolios. *The Journal of Finance*, 72(4):1611–1644, May 2017.

[49] Tobias J. Moskowitz, Yao Hua Ooi, and Lasse Heje Pedersen. Time Series Momentum. *Journal of Financial Economics*, 104(2):228–250, May 2012.

[50] Christian Mueller-Glissmann and Andrea Ferrario. Dynamic Asset Allocation using Machine Learning: Seeing the Forest for the Trees. *Journal of Portfolio Management*, 50(5):132, 2024.

[51] Michael Parkinson. The Extreme Value Method for Estimating the Variance of the Rate of Return. *Journal of business*, pages 61–65, 1980.

[52] Permanent Portfolio Family of Funds. Permanent Portfolio, 2026. Accessed: 2026-09-01.

[53] Dimitris N. Politis and Joseph P. Romano. The Stationary Bootstrap. *Journal of the American Statistical Association*, 89(428):1303–1313, December 1994.

[54] Edward E. Qian. Risk Parity Portfolios: Efficient Portfolios Through True Diversification. White paper, PanAgora Asset Management, September 2005.

[55] Thierry Roncalli. *Introduction to Risk Parity and Budgeting*. Chapman and Hall/CRC Financial Mathematics Series. Chapman and Hall/CRC, 2013.

[56] RPAR Risk Parity ETF. RPAR Risk Parity ETF, 2026. Accessed: 2026-09-01.

[57] Maria Laura Santoni, Vincent Jouanne, and Matthew L Scullin. Equity Strategy Backtesting: Luck or Edge? the Minervascore as a Statistical Robustness Grade. *arXiv preprint arXiv:2608.23808*, 2026.

[58] William F. Sharpe. Mutual Fund Performance. *The Journal of Business*, 39(1):119–138, 1966.

[59] William F. Sharpe. The Sharpe Ratio. *The Journal of Portfolio Management*, 21(1):49–58, October 1994.

[60] George Smith. Is the 60/40 Portfolio Still Relevant? Exploring Alternatives. LPL Research Blog, December 2025. Additional content provided by Kent Cullinane, CFA. Accessed: 2026-06-07.

[61] S&P Dow Jones Indices. Risk Control Indices, 2026. Accessed: 2026-09-01.

[62] Ivo Welch and Amit Goyal. A Comprehensive Look at The Empirical Performance of Equity Premium Prediction. *The Review of Financial Studies*, 21(4):1455–1508, July 2008.

[63] Mehnaz Yasmin. Morgan Stanley CIO Favors 60/20/20 Portfolio Strategy with Gold Inflation Hedge. Reuters Markets, September 2025.

A Portfolio accounting

Let $r_t \in \mathbf{R}^3$ be the vector of close-to-close daily simple returns on day t, and let r_t^{rf} be the daily risk-free rate (the federal funds rate expressed per trading day). The cash weight is $c_t = 1 - (w_t^{\mathrm{spy}} + w_t^{\mathrm{agg}} + w_t^{\mathrm{gld}})$.

On a day that is *not* a rebalance day, each asset holding grows with its return and cash grows at the risk-free rate. The portfolio value V_t updates as

$$V_t = V_{t-1} \left(\sum_{i=1}^{3} w_{t-1}^{i} \, (1 + r_t^{i}) + c_{t-1} \, (1 + r_t^{\mathrm{rf}}) \right),$$

and the drifted weights become

$$(w_t)_i = \frac{V_{t-1}}{V_t} \, w_{t-1}^{i} \, (1 + r_t^{i}), \quad i = 1, 2, 3.$$

The numerator is the value of the ith asset after its return, and the denominator is the new portfolio value, so the ratio is the new weight of the ith asset.

On a rebalance day, the value and weights are first updated as above, giving the drifted weights w_t and the pre-trade value V_t. We then compute the target weights $\hat{w}_t$ by one of the methods in §2. We finally trade so that the weights equal the desired weights $\hat{w}_t$; buying or selling an asset moves the difference into or out of cash, so the post-trade cash weight is $\hat{c}_t = 1 - (\hat{w}_t^{\mathrm{spy}} + \hat{w}_t^{\mathrm{agg}} + \hat{w}_t^{\mathrm{gld}})$.

The total traded fraction of the portfolio is $\sum_{i=1}^{3} |\hat{w}_t^i - w_t^i|$, counting the three (noncash) assets only; moving value into or out of cash is not itself a trade. Trading incurs a trading cost at the half-spread $s/2$. With $s = 5$ basis points, the trading cost of a rebalance is

$$\frac{s}{2} \sum_{i=1}^{3} \left| \hat{w}_t^i - w_t^i \right| V_t.$$

This cost is deducted from the cash account, so the post-trade value is the pre-trade value minus the trading cost, and the post-trade weights are re-computed using the updated portfolio value.

We define the turnover of a rebalance as half the total traded fraction, $\frac{1}{2} \sum_{i=1}^{3} |\hat{w}_t^i - w_t^i|$, so that selling one asset to buy another counts the exchanged value once. We report annualized turnover as 252 times the average daily turnover, with zero turnover on non-rebalancing days.

Group	Number	Features
Asset trend	12	Trailing returns of ETFs over 63, 126, and 252 days, and the 252-day trailing return divided by its trailing volatility.
Asset volatility	3	Log short (21 day) minus log long-run (252 day) volatility of ETFs.
Asset liquidity	9	Average log trading volume of ETFs over 21, 63, and 126 days.
Rates	7	Treasury yields at 3 months, 2, 5, 10, and 30 years, the 10-year real yield, and the 10-year breakeven inflation rate.
Market state	3	Smoothed (63 day average) log VIX, smoothed (512 day average) broad dollar index, and the 2-year minus 3-month yield-curve slope.
Fama–French	8	Smoothed (63 day average) returns of market, size, value, profitability, investment, momentum, short-term reversal, and long-term reversal factors.

Table 7: The feature vector z_t used for the return forecast.

B Return forecasting

In this section we describe the return forecast α_t used by the Markowitz portfolio, which is based on simple regression.

B.1 Target and features

Target and features. The forecast target is $r_t^{\text{forw}} \in \mathbf{R}^3$, the 100-trading day forward return of the three assets, annualized by scaling by 252/100. It is computed from the same adjusted close prices used in the simulations (§2.5). The feature vector $z_t \in \mathbf{R}^p$ is described in table 7. The asset-level features—trailing returns, volatilities, and trading volumes—are computed from the ETF price and volume data; the rates and market-state features are downloaded from FRED; and the factor features are downloaded from the Kenneth French data library. For forecasting on day t, all features are computed using only data available on day t, on a rolling basis. The total number of features is $p = 42$.

B.2 Forecasting method

On each day t, we fit a multivariate ridge regression from the feature vector z_τ to the forward-return vector r_τ^{forw}. The training set contains at most the most recent 512 complete observations. Its final target observation is dated $t - 100$, so that the corresponding forward return ends at t and no target uses prices after the forecast date, *i.e.*, the method is causal. We require at least 63 complete observations; before then, the forecast is set to zero.

Within each training window, feature j is divided by its sample (equal weight) standard deviation d_j (with a small numerical floor). Features are not centered, and no intercept is included. Let $\tilde{z}_\tau \in \mathbf{R}^p$ denote this scaled feature vector. We choose the coefficient matrix $B_t \in \mathbf{R}^{3\times p}$ by solving the convex quadratic optimization problem

$$\text{minimize} \quad \textstyle\sum_\tau \rho_\tau \left\| B\tilde{z}_\tau - r_\tau^{\text{forw}} \right\|_2^2 + \lambda \|B\|_F^2, \tag{4}$$

with variable $B \in \mathbf{R}^{3\times p}$, where the normalized observation weights ρ_τ decay exponentially with a half-life of 252 trading days. We use regularization parameter value $\lambda = 10$. The annualized return forecast is then

$$\alpha_t^{\text{ann}} = B_t \tilde{z}_t \in \mathbf{R}^3,$$

where B_t is the solution of (4) and $\tilde{z}_t$ is the scaled feature vector on day t. Only features observed throughout the training window and on day t are used. For the monthly portfolio decision in (3), we put this forecast on the same 21-trading-day scale as the cash return by using $\alpha_t = (21/252)\alpha_t^{\text{ann}}$. The simple Markowitz forecast is put on the same scale by multiplying its exponentially weighted average daily return by 21. This simple forecast model is a ridge-regularized linear forecast, re-fit each trading day using a rolling sample.

B.3 Forecast performance

The return forecasts used by the simple Markowitz and Markowitz portfolios are shown in figure 6, alongside the realized 100-day forward return (the target for the forecast method). The trajectories are shown in their original annualized units. The trailing average is smooth, while the regression forecast responds to a broader set of market information. The two forecasts are generally, but not always, in agreement. It is easy to identify times when both forecasts are poor, *e.g.*, have the wrong sign compared to the realized 100-day forward returns.

We evaluate the forecast performance by computing the cosine similarity between the forecast at time t and the realized 100-day forward return at time t. The results are shown in figure 7. The good news is that these are generally positive. From comparison with figure 1, we see that the periods with large cosine similarity correspond to periods with large cumulative returns.

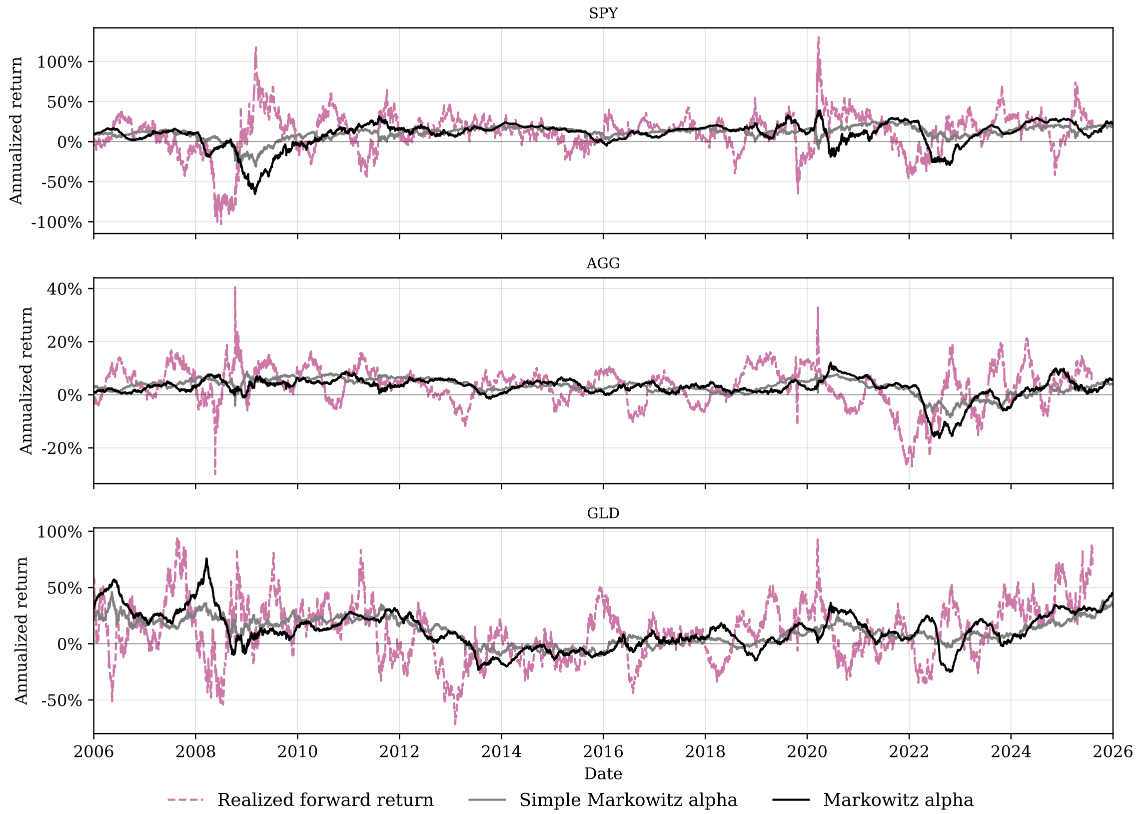


Figure 6: The two return forecasts and the realized 100-day forward return r_t^{forw}, plotted at the forecast date t, one panel per asset. The forecasts and realized returns are shown in annualized units, without normalization.

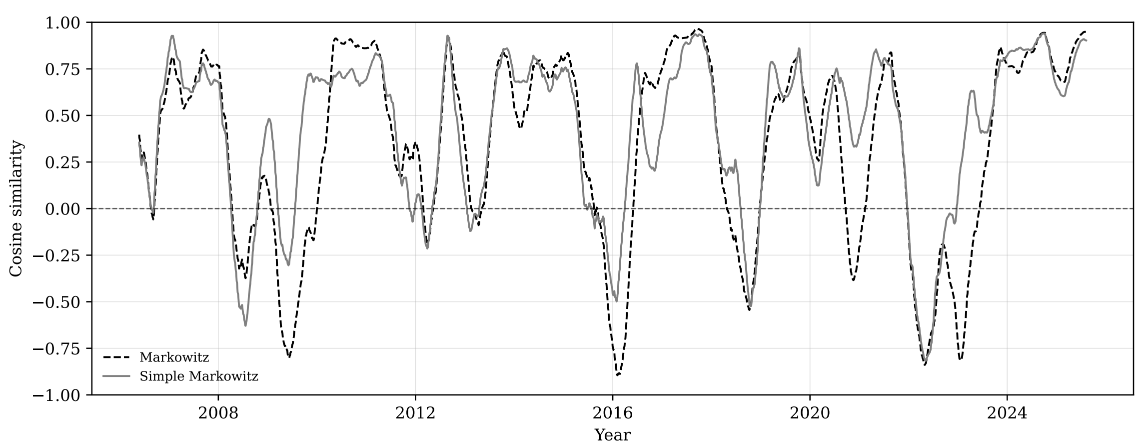


Figure 7: Rolling 100-day mean of the cosine similarity between the return forecast α_t and the realized 100-day forward return $r_t^{(h)}$.

Portfolio	Return	95% interval	Sharpe	95% interval	Max DD	95% interval
Markowitz	11.6%	[7.5%, 15.7%]	1.08	[0.62, 1.57]	18.1%	[9.9%, 26.1%]
Simple Markowitz	10.4%	[6.4%, 14.5%]	0.91	[0.48, 1.38]	15.8%	[10.5%, 25.9%]
50/30/20 VC	7.8%	[4.7%, 10.8%]	0.82	[0.39, 1.26]	15.9%	[8.5%, 22.4%]
60/40 VC	7.1%	[4.0%, 10.1%]	0.71	[0.29, 1.16]	16.9%	[9.1%, 23.9%]
50/30/20	9.1%	[4.7%, 13.3%]	0.70	[0.27, 1.17]	27.1%	[12.7%, 36.2%]
60/40	8.2%	[3.4%, 12.8%]	0.56	[0.13, 1.04]	33.7%	[14.9%, 41.9%]

Table 8: Point estimates and bootstrap 95% intervals for the return, Sharpe ratio, and maximum drawdown of the six portfolios.

Portfolio	Mean of Δ	95% interval	$P(\Delta \leq 0)$
Simple Markowitz	+0.17	$[-0.06, +0.40]$	0.068
50/30/20 VC	+0.26	$[-0.06, +0.61]$	0.057
60/40 VC	+0.37	$[-0.02, +0.78]$	0.033
50/30/20	+0.38	$[-0.01, +0.79]$	0.027
60/40	+0.52	$[+0.07, +0.99]$	0.011

Table 9: Difference Δ between the Sharpe ratio of the Markowitz portfolio and that of each of the other portfolios, paired across bootstrap replications, with its 95% interval and the bootstrap probability that it is not positive.

C Statistical significance

All confidence intervals reported here come from a stationary block bootstrap [53], with a mean block length of 21 trading days and 10,000 replications applied on the portfolio returns time series. The stationary bootstrap preserves dependence within randomly selected runs of consecutive observations. So if the time-series has a short-memory autocorrelation, that structure is approximately retained in each replication. Within each replication we sample the time series of returns for each portfolio (and the fed funds rate series) in the same manner (dictated by the index series produced by the bootstrap), so the contemporaneous correlation between the portfolios is preserved, and the Sharpe ratio differences reported below are paired.

Table 8 gives the point estimate and the 95% interval for the return, Sharpe ratio, and maximum drawdown of each of the six portfolios.

Table 9 gives the difference between the Sharpe ratio of the Markowitz portfolio and that of each of the other five portfolios, with its 95% interval and the bootstrap probability that the difference is not positive.

The intervals are wide, as is expected for Sharpe ratios estimated over twenty years. The advantage of the Markowitz portfolio over the 60/40 fixed-weight benchmark—the principal benchmark this paper sets out to beat—is significant at the 5% level: its interval excludes zero. The interval for the 50/30/20 fixed-weight benchmark narrowly includes zero.

The advantages over the volatility-controlled benchmarks and over the simple Markowitz portfolio are of comparable magnitude in point estimate—the difference over the 60/40 VC benchmark, +0.37, is very nearly the difference over the 50/30/20 benchmark, +0.38—but their intervals include zero, so they are not significant at the 5% level on this sample. (The intervals are two-sided, so significance at this level corresponds to $P(\Delta \leq 0) \leq 0.025$.) Twenty years of data cannot resolve differences of this size. These last comparisons are among portfolios we propose here: the volatility-controlled benchmarks and the simple Markowitz portfolio are constructions of this paper rather than established methods, so they bear on which of our portfolios an investor should prefer, and not on whether we improve on the status quo. We therefore do not claim that the Markowitz portfolio is significantly better than its volatility-controlled counterparts, only that it ranks above them consistently. We note that these intervals describe sampling variation only. They do not account for the specification search discussed in appendix E.

Table 2 gives the Sharpe ratios of the six portfolios over the four consecutive five-year subperiods, and over the full period. The Markowitz portfolio has the highest Sharpe ratio in two of the four subperiods, and over the full period. In 2011–2016 it does not: its Sharpe is 0.86, below the simple Markowitz portfolio, at 0.94, and both of the 60/40 portfolios, at 0.99. Its margin over the volatility-controlled benchmarks is widest in the first subperiod, which contains the 2008 crisis, and is narrower afterwards. Excluding the 2008 and 2009 calendar years entirely and splicing the remaining daily returns, the Sharpe ratios are 1.11 for the Markowitz portfolio, 0.95 for the simple Markowitz portfolio, 0.93 for 50/30/20 VC, 0.83 for 60/40 VC, 0.87 for 50/30/20, and 0.74 for 60/40, so the ordering is essentially preserved—only 50/30/20 moves above 60/40 VC—but the margin narrows.

Lag	Return	Volatility	Sharpe	Max DD	Mean DD	Turnover
0 days	11.6%	9.0%	1.08	18.1%	3.0%	284.4%
1 day	11.6%	9.0%	1.08	18.1%	2.9%	284.8%
5 days	11.5%	9.0%	1.07	17.6%	2.9%	285.3%
21 days	10.9%	9.0%	1.00	15.1%	3.3%	284.2%

Table 10: Performance of the Markowitz portfolio when its return forecast is lagged.

Lag	Return	Volatility	Sharpe	Max DD	Mean DD	Turnover
0 days	11.6%	9.0%	1.08	18.1%	3.0%	284.4%
1 day	10.9%	8.9%	1.01	18.8%	2.8%	268.8%
5 days	10.1%	10.2%	0.81	30.6%	3.6%	274.5%
21 days	9.4%	10.3%	0.73	23.0%	4.4%	285.4%

Table 11: Performance of the Markowitz portfolio when both its return forecast and covariance estimate are lagged.

D Lagged information

In this section we experiment with the effect on performance of lagging the information provided to the Markowitz portfolio by one day, one week, and one month. In the first experiment we lag α_t, but not Σ_t, with the results shown in table 10. We can see that lags in α_t barely affect the portfolio performance, with only a modest degradation with full month lag. In summary, our simple return forecast α_t is not sensitive to lag.

In our second experiment, we lag both α_t and Σ_t, with the results shown in table 11. Here too a one day lag has little effect, but longer lags do degrade performance more than when α_t alone is lagged. Our conclusion is that in our Markowitz portfolio, a reactive covariance forecast appears to be substantially more important than a reactive return forecast. This is consistent with the covariance estimate's short 11-day window, which is intended to respond quickly to changing market risk.

E Hyper-parameter sensitivity and choice

Sensitivity to backtest years. Our simulations and analyses are for the specific span from 2006–2026, which includes several market disruptions and multiple market conditions. When other date ranges are chosen, the performance numbers vary a bit, but the overall conclusions are almost always the same: the volatility-controlled portfolios have better performance than the fixed-weight benchmarks, and the Markowitz portfolios have even better performance.

Hyper-parameter sensitivity. We first address the question of sensitivity of our results for the Markowitz portfolio to the hyper-parameters. The Markowitz portfolio has very few hyper-parameters to choose. They are target volatility (which we take as 7%), the allowed deviation of the relative weights from $(0.5, 0.3, 0.2)$ (which we take as one), the trailing period for the covariance estimate (which we take as 11 days), and the hyper-parameters in our return forecast. Our choice of features, and their processing (listed in table 7) seem to be unremarkable. The hyper-parameters in the return forecast are the target horizon (which we take as 100 days), the half-life used to fit the regression matrix B (which we take as 252 days), and the ridge regularization parameter λ (which we take to be 10).

We analyze the performance for varying target volatilities in §3.4; the results are what we would expect, and the high level conclusions remain. The limit on relative weight deviation has a small and positive effect on performance, and could be removed without much change. The covariance estimate window does affect the performance, though smoothly. The three hyper-parameters in the return forecast have a smooth effect on performance. As for choice of features, we can drop several without much decrease in performance. (And we have no doubt that adding others could improve performance.) In summary, the performance of the Markowitz portfolio is not particularly sensitive to the choice of the (few) hyper-parameters. Large changes do, however, affect the performance.

Choice of hyper-parameters. Another question we address is *when* we choose the hyper-parameters, by doing simulations on past data. (As usual, hyper-parameters are crudely gridded, say by factors of 2.) For a fixed 2006 start, the current specification remains close to the best tested one-at-a-time neighboring specification at every annual endpoint from 2011 through 2025. A longer alpha-fitting half-life of 315 days performs slightly better over these endpoints, but improves the full-period Sharpe ratio by only 0.01. We also compare the fixed specification against annual walk-forward selection, in which, at each year end, the specification is chosen by Sharpe ratio over the trailing window, using only data preceding the deployment year, and is then applied for the following year. The evaluation window begins at the first year at which a selection can be made, so the two selection rules are evaluated over different windows. Table 12 gives the results. The trailing five-year rule sometimes chose neighboring covariance-window or decay values. Thus, annual retuning did not improve out-of-sample performance, and the results appear locally robust, although shorter estimation windows do not always select exactly the same parameters.

Selection rule	Evaluation window	Fixed specification	Walk-forward selection
Trailing five years	2011–2026	1.00	0.93
Trailing ten years	2016–2026	1.08	1.00

Table 12: Sharpe ratio of the fixed specification and of annual walk-forward selection, for two selection windows. The two rules are evaluated over different windows, since the evaluation window begins at the first year at which a selection can be made.

Deflated Sharpe ratio. We account for the specification search using the deflated Sharpe ratio of Bailey and López de Prado [3, 2]. The documented sweep contains 17 specifications, whose annualized Sharpe ratios have a standard deviation of 0.07, which gives an expected maximum Sharpe ratio, under the null hypothesis of no skill, of 0.12. The resulting deflated Sharpe ratio for the Markowitz portfolio exceeds 0.99, and remains above 0.99 when the assumed number of trials is raised to 1000, which lifts the expected maximum under the null only to 0.22. This calculation assumes that the trials are the documented one-at-a-time neighbors of the chosen specification. These are highly correlated with one another, so the dispersion of the trial Sharpe ratios—and hence the deflation—is likely understated.

F Comparison with risk-based portfolios

Both our volatility-controlled and our optimization-based portfolios size their positions using an estimate of risk, so it is natural to ask whether their performance simply reflects a generic benefit of risk control, which is shared by a large family of well known risk-based allocation rules. In this appendix we address this question by comparing the Markowitz portfolio against seven standard alternatives, implemented in our setting.

Each benchmark holds the same three assets (SPY, AGG, and GLD) plus cash, is long only and unlevered, rebalances monthly, uses the same covariance estimate Σ_t described in §2.4, and pays the same trading cost, over the same 2006–2026 period. Each is evaluated using the metrics of §2.6. The comparison is therefore an apples-to-apples one: the benchmarks differ from our portfolios only in how they turn the same data into weights.

F.1 Methods

On each rebalance day, each of the seven methods produces a fully invested vector of relative weights $\eta \in \mathbf{R}^3$, satisfying $\eta \geq 0$ and $\mathbf{1}^T\eta = 1$. Let $\nu_t \in \mathbf{R}^3$ be the vector of estimated asset volatilities, $(\nu_t)_i = (\Sigma_t)_{ii}^{1/2}$, and let

$$\mathrm{RC}_i(\eta) = \frac{\eta_i(\Sigma_t\eta)_i}{\eta^T\Sigma_t\eta}, \quad i = 1, 2, 3,$$

be the fraction of portfolio variance contributed by asset i; these fractions sum to one, though an individual fraction can be negative when the covariance estimate has negative off-diagonal entries. The seven methods are the following.

- *Equal weight.* $\eta = (1/3)\mathbf{1}$, the $1/N$ rule, which is a surprisingly strong benchmark in practice [21].
- *Inverse volatility.* $\eta_i \propto 1/(\nu_t)_i$, sometimes called naïve risk parity [54, 42].
- *Equal risk contribution.* η is chosen so that $\mathrm{RC}_i(\eta) = 1/3$, *i.e.*, each asset contributes the same share of portfolio variance [43, 54].
- *50/30/20 risk budget.* η is chosen so that $\mathrm{RC}(\eta) = (0.5, 0.3, 0.2)$, *i.e.*, *risk*, rather than capital, is split in the proportions of the 50/30/20 benchmark [16, 55].
- *Global minimum variance.* η minimizes $\eta^T\Sigma_t\eta$ [33, 20].
- *Maximum diversification.* η maximizes the diversification ratio $(\nu_t^T\eta)/(\eta^T\Sigma_t\eta)^{1/2}$, *i.e.*, the ratio of the weighted average asset volatility to the portfolio volatility [19].
- *Black–Litterman.* The return forecast is shrunk toward the returns implied by an equilibrium portfolio [8, 34]. We take the 50/30/20 benchmark as the equilibrium portfolio, giving the implied returns $\pi_t = \delta\Sigma_t w^{\mathrm{tgt}}$ with risk aversion $\delta = 2.5$, and use the same return forecast α_t as our Markowitz portfolio (appendix B) as absolute views on all three assets. With prior covariance $\tau\Sigma_t$ and diagonal view covariance

$\Omega_t = \tau\,\mathbf{diag}((\Sigma_t)_{11}, (\Sigma_t)_{22}, (\Sigma_t)_{33})$, *i.e.*, view uncertainty proportional to each asset's variance, and $\tau = 0.05$, the posterior mean return is

$$\bar{\mu}_t = \left((\tau\Sigma_t)^{-1} + \Omega_t^{-1}\right)^{-1} \left((\tau\Sigma_t)^{-1}\pi_t + \Omega_t^{-1}\alpha_t\right).$$

The weights η then maximize the mean-variance utility $\bar{\mu}_t^T\eta - (\delta/2)\,\eta^T\Sigma_t\eta$. (Here Σ_t and α_t are expressed in annual units, since δ and τ are not scale free.)

The first six methods use only the covariance estimate Σ_t; only Black–Litterman uses a forecast of returns, and it uses exactly the forecast that our Markowitz portfolio uses. The first two are given in closed form; the others are small smooth problems over the simplex, which we solve numerically, warm started at the previous month's solution.

None of these methods holds cash; each prescribes only the relative weights of the three assets. To compare them with our portfolios at a common level of targeted risk, we also form a volatility-controlled version of each, exactly as in (1): on each rebalance day the weights are

$$\hat{w}_t = \min\left\{\frac{\sigma^{\rm tgt}}{(\eta^T\Sigma_t\eta)^{1/2}},\ 1\right\}\eta,$$

with the same target $\sigma^{\rm tgt} = 7\%$ used throughout the paper, and the remainder held in cash. We report both the fully invested and the volatility-controlled variants.

F.2 Results

Table 13 gives the metrics for the seven benchmarks, in both variants, with the Markowitz portfolio repeated from table 1 for reference. Table 14 gives their average relative weights, along with the average cash weight of the volatility-controlled variants. (The relative weights do not depend on the variant, since scaling toward the volatility target does not change them.)

Risk control alone does not explain the performance. At their standard specifications, every one of the seven methods lands well below the Markowitz portfolio. With volatility control, their Sharpe ratios range from 0.60 to 0.85, against 1.08 for the Markowitz portfolio. The best of them, the 50/30/20 risk budget portfolio, at 0.85, is only slightly above our own 50/30/20 volatility-controlled benchmark, at 0.82; indeed the four best methods are, for practical purposes, tied with it. The picture is much the same without volatility control, where the Sharpe ratios range from 0.63 to 0.80. It is worth stating this comparison the other way around. In their conventional fully invested form, every one of the seven methods falls below our 50/30/20 volatility-controlled benchmark; the highest of them, inverse volatility, reaches 0.80 against its 0.82. The four that do edge past it are those given our volatility-control overlay. Sizing positions by risk, however it is done, gets an investor to roughly the same place as the simple volatility-controlled benchmark of §2.3, and no further.

The comparison is not an artifact of the risk level. Because these methods cannot lever up, several of them realize volatilities well below the 7% target—as low as 4.9% for the global minimum variance portfolio—so we compare them on Sharpe ratio rather than return. The Markowitz portfolio's Sharpe ratio is in any case not very sensitive to the risk

Portfolio	Return	Volatility	Sharpe	Max DD	Mean DD	Turnover
Markowitz	11.6%	9.0%	1.08	18.1%	3.0%	284.4%
Volatility-controlled, 7% target						
50/30/20 risk budget	6.6%	5.6%	0.85	14.7%	1.8%	178.9%
Equal risk contribution	6.2%	5.3%	0.83	14.7%	2.0%	158.5%
Equal weight	7.7%	7.2%	0.82	14.2%	2.6%	74.2%
Maximum diversification	6.6%	5.9%	0.82	15.8%	2.1%	276.3%
Inverse volatility	6.1%	5.4%	0.80	15.1%	2.0%	126.1%
Black–Litterman	8.5%	8.6%	0.77	16.2%	3.1%	270.4%
Global minimum variance	4.7%	4.9%	0.60	16.2%	1.8%	161.8%
Fully invested						
50/30/20 risk budget	7.3%	7.1%	0.77	18.7%	1.9%	189.9%
Equal risk contribution	6.4%	6.9%	0.67	19.4%	2.0%	168.1%
Equal weight	8.7%	9.3%	0.73	23.8%	2.9%	13.5%
Maximum diversification	6.9%	6.7%	0.75	22.0%	2.3%	278.3%
Inverse volatility	6.6%	6.1%	0.80	17.0%	1.9%	120.7%
Black–Litterman	11.4%	15.2%	0.63	27.7%	6.1%	315.3%
Global minimum variance	5.1%	5.3%	0.63	17.1%	1.8%	157.8%

Table 13: Performance metrics for the seven risk-based benchmarks, in the volatility-controlled (middle) and fully invested (bottom) variants, with the Markowitz portfolio for reference (top). Within each panel the methods are ordered by the Sharpe ratio of the volatility-controlled variant.

Portfolio	SPY	AGG	GLD	Cash
Markowitz	48.3%	23.4%	28.3%	15.1%
50/30/20 risk budget	27.5%	58.7%	13.7%	5.4%
Equal risk contribution	20.5%	63.0%	16.5%	3.2%
Equal weight	33.4%	33.2%	33.4%	12.9%
Maximum diversification	23.0%	61.6%	15.4%	2.8%
Inverse volatility	18.7%	64.2%	17.1%	2.1%
Black–Litterman	44.9%	12.1%	43.0%	33.0%
Global minimum variance	10.8%	84.9%	4.4%	1.2%

Table 14: Average relative weights of the three assets, and average cash weight of the volatility-controlled variant, for the seven risk-based benchmarks and the Markowitz portfolio, averaged over all trading days in the evaluation window. The relative weights are the same for both variants.

level (figure 4b): run at a 3% target it realizes 4.8% volatility with a Sharpe ratio of 0.92, and at a 5% target it realizes 7.1% volatility with a Sharpe ratio of 1.07. So at each benchmark's own realized volatility, the Markowitz portfolio still has a substantially higher Sharpe ratio.

Where the risk-based methods lose. Five of the six purely risk-based methods (all but equal weight) allocate using the covariance estimate alone, and are therefore drawn toward the lowest volatility asset. As table 14 shows, they hold between 59% and 85% of their risky value in AGG, whose annualized return over the period is 3.1% (table 1). Their returns—4.7% to 6.6% with volatility control, and 5.1% to 7.3% fully invested—reflect this: they diversify risk well, but they have no mechanism for tilting toward the assets that are expected to perform well. The Markowitz portfolio, by contrast, holds a far more balanced average mix, 48.3/23.4/28.3, and—as discussed in §3.2—derives its performance from varying that mix over time rather than from its average.

The return forecast is necessary but not sufficient. The Black–Litterman portfolio uses the same return forecast α_t as our Markowitz portfolio, and it does earn a higher return than the purely risk-based methods, 8.5% with volatility control and 11.4% without. But it earns those returns at a much higher risk: fully invested it realizes 15.2% volatility with a 27.7% maximum drawdown, and even with volatility control its Sharpe ratio is 0.77. Trading off risk against return in a utility function, with the forecast shrunk toward an equilibrium prior, is simply less effective here than imposing a hard constraint on the estimated volatility and an ℓ_1 trust region around the 50/30/20 mix (2), as in (3). The lift we report therefore comes neither from the return forecast alone nor from risk control alone, but from the combination. Black–Litterman is the only benchmark considered here with free hyper-parameters; we sweep them in §F.3, and its advantage over the purely risk-based methods survives, but the ranking against the Markowitz portfolio does not change.

Turnover. The Markowitz portfolio has the highest turnover of the portfolios in table 13, at 284.4%. But turnover does not explain the ranking: the maximum diversification and Black–Litterman portfolios trade nearly as much, at 276.3% and 270.4%, with Sharpe ratios of 0.82 and 0.77; and all of the figures reported here are net of the same trading costs.

The covariance estimate does not favor our method. All seven methods use the covariance estimate Σ_t of §2.4, whose 11-day window we chose for our own portfolios, so one might ask whether that choice handicaps them. We therefore re-ran every method, in both variants, over a grid of trailing covariance windows of 11, 21, 63, 126, and 252 trading days, *i.e.*, 70 specifications in all. Table 15 gives the Sharpe ratio of the volatility-controlled variant.

For every one of the seven methods the volatility-controlled variant attains its highest Sharpe ratio at the 11-day window—the window our own portfolios use—and longer, more conventional windows make every method worse, monotonically or nearly so; the shared covariance estimate helps the comparators, it does not handicap them. Taking the best cell of the whole grid, selected in sample, the highest Sharpe ratio any of these methods reaches

Method	11 days	21 days	63 days	126 days	252 days
50/30/20 risk budget	0.85	0.75	0.73	0.68	0.68
Equal risk contribution	0.83	0.77	0.74	0.69	0.70
Equal weight	0.82	0.78	0.72	0.69	0.67
Maximum diversification	0.82	0.78	0.75	0.69	0.71
Inverse volatility	0.80	0.74	0.73	0.70	0.69
Black–Litterman	0.77	0.70	0.66	0.66	0.70
Global minimum variance	0.60	0.45	0.42	0.37	0.37

Table 15: Sharpe ratio of the volatility-controlled variant of each of the seven risk-based benchmarks, for trailing covariance windows of 11 to 252 trading days. The methods are ordered as in table 13.

is 0.85, against 1.08 for the Markowitz portfolio; in the fully invested variant the best cell anywhere in the grid is 0.80, for inverse volatility at the 11-day window.

F.3 Tuning Black–Litterman

The six purely risk-based methods have no free hyper-parameters beyond the covariance estimate, which they share with our portfolios. Black–Litterman does, so we sweep them and report its best in-sample performance.

We first note that the scale τ has no effect at all. With prior covariance $\tau\Sigma_t$ and view covariance $\Omega_t = \tau\,\mathbf{diag}(\Sigma_t)$, as in §F.1, the factor τ cancels from the posterior mean, leaving

$$\bar{\mu}_t = \left(\Sigma_t^{-1} + \mathbf{diag}(\Sigma_t)^{-1}\right)^{-1}\left(\Sigma_t^{-1}\pi_t + \mathbf{diag}(\Sigma_t)^{-1}\alpha_t\right),$$

where $\mathbf{diag}(\Sigma_t)$ denotes the diagonal part of Σ_t. (In simulation, $\tau = 0.005$, 0.05, and 0.5 give identical results to six significant figures.) To vary the confidence placed in the views we therefore introduce a multiplier $\omega > 0$, taking $\Omega_t = \omega\tau\,\mathbf{diag}(\Sigma_t)$, which gives

$$\bar{\mu}_t = \left(\Sigma_t^{-1} + \omega^{-1}\,\mathbf{diag}(\Sigma_t)^{-1}\right)^{-1}\left(\Sigma_t^{-1}\pi_t + \omega^{-1}\,\mathbf{diag}(\Sigma_t)^{-1}\alpha_t\right).$$

As $\omega \to 0$ the posterior mean approaches the return forecast α_t, and as $\omega \to \infty$ it approaches the equilibrium returns π_t; the canonical specification is $\omega = 1$.

We sweep the risk aversion δ over ten values from 0.5 to 100 and the view confidence ω over nine values from 10^{-4} to 100, giving 90 specifications, each run in both variants. Table 16 reports the best of them, alongside the canonical specification and the Markowitz portfolio.

Three things are worth noting. First, even at its best specification—selected with full hindsight over the entire simulation, on the metric being compared—Black–Litterman reaches a Sharpe ratio of 0.95 with volatility control and 0.94 fully invested, against 1.08 for the Markowitz portfolio. Over all 90 specifications the volatility-controlled Sharpe ratio ranges from 0.70 to 0.95, and the fully invested one from 0.57 to 0.94; all of them remain below 1.08. The comparison is a conservative one, since the hyper-parameters of the Markowitz portfolio are not selected this way (appendix E).

Specification	Return	Volatility	Sharpe	Max DD	Mean DD	Turnover
Markowitz	11.6%	9.0%	1.08	18.1%	3.0%	284.4%
Volatility-controlled, 7% target						
Canonical, $\delta = 2.5$, $\omega = 1$	8.5%	8.6%	0.77	16.2%	3.1%	270.4%
Best, $\delta = 20$, $\omega = 10^{-4}$	10.1%	8.7%	0.95	19.1%	3.4%	327.4%
Fully invested						
Canonical, $\delta = 2.5$, $\omega = 1$	11.4%	15.2%	0.63	27.7%	6.1%	315.3%
Best, $\delta = 15$, $\omega = 10^{-4}$	12.3%	11.0%	0.94	24.3%	4.4%	328.1%

Table 16: Black–Litterman under its canonical specification and under the best of the 90 specifications swept, selected in-sample on the Sharpe ratio, with the Markowitz portfolio for reference.

Second, the best specifications are those with ω near zero, *i.e.*, those in which the equilibrium prior is switched off entirely and the posterior mean is just our return forecast α_t. The shrinkage that Black–Litterman is designed to provide does not help here; what the portfolio gains, it gains from the forecast it is given.

Third, what remains once the prior is switched off is a difference in formulation. With $\omega \to 0$ the benchmark chooses a fully invested portfolio maximizing $\alpha_t^T \eta - (\delta/2)\eta^T \Sigma_t \eta$ and then scales it toward the volatility target, whereas our Markowitz portfolio chooses the cash weight and the asset mix jointly, subject to a hard constraint on estimated volatility and an L_1 trust region around 50/30/20. That difference is worth about 0.13 of Sharpe ratio here.

F.4 Caveats

These are implementations of the seven methods within the restrictive setting of this paper: three assets, long only, no leverage, and monthly rebalancing. Risk parity in particular is normally applied to a broader universe and with leverage, which is how it reaches equity-like returns; denied leverage, it is confined to the low-volatility corner of the universe. Our claim is therefore the narrow one: for an investor working under the constraints we consider, and using the same data, these methods do not reach the performance of the Markowitz portfolio.

G Simulation assumptions

Trading cost. Our simulations charge a spread of 5 basis points on each trade, which we consider conservative for the three highly liquid ETFs we use. To check that our results do not rest on this, we re-run the six portfolios at 10 and 20 basis points; table 17 gives the Sharpe ratios. At each spread, the optimization-based portfolios use that same spread in their objective and the simulation charges it to realized value. At four times the assumed cost the Markowitz portfolio still returns 11.1%, at a Sharpe ratio of 1.03, and the ranking of the six portfolios is unchanged at every cost level. The annually rebalanced benchmarks are essentially unaffected, since they barely trade.

Portfolio	5 bp	10 bp	20 bp
Markowitz	1.08	1.07	1.03
Simple Markowitz	0.91	0.92	0.89
50/30/20 VC	0.82	0.81	0.80
60/40 VC	0.71	0.71	0.69
50/30/20	0.70	0.70	0.70
60/40	0.56	0.56	0.56

Table 17: Sharpe ratios of the six portfolios at trading cost rates of 5, 10, and 20 basis points. The paper uses 5 basis points throughout.

The cash rate. Our simulations accrue cash at the effective federal funds rate, which an individual investor cannot obtain directly; the practical substitute is a Treasury bill fund, which yields slightly less. We therefore re-run the simulations with cash accruing at the federal funds rate less 25 basis points annualized, while still measuring the Sharpe ratio in excess of the federal funds rate itself. The optimization-based portfolios also use this lower cash return in their objective and therefore re-optimize their allocations. The simple Markowitz Sharpe ratio falls by 0.002, and the volatility-controlled portfolios fall by less than 0.007. The Markowitz portfolio instead increases its risky allocation and its realized Sharpe ratio rises by 0.014 in this historical simulation; this is a reallocation effect, not a direct benefit from receiving less interest. The two fixed-weight benchmarks are unaffected, since they hold no cash. The direct effect is small because even the Markowitz portfolio holds only around 15% of its value in cash on average.