Improved Tracking-Error Management for Active and Passive Investing

This Version: August 19, 2024

Gianluca De Nard, Olivier Ledoit, and Michael Wolf

Gianluca De Nard is a Postdoctoral Researcher at the University of Zurich, a Research Fellow at the NYU Stern Volatility and Risk Institute in New York, and Head of Quantitative Research at OLZ AG in Zurich.
gianluca.denard@econ.uzh.ch
Gessneralle 38, 8001 Zurich, Switzerland
+41 44 563 30 39

Olivier Ledoit is a Senior Research Associate at the University of Zurich and a Partner at AlphaCrest Capital Management in New York.
olivier.ledoit@econ.uzh.ch
Zürichbergstrasse 14, 8032 Zurich, Switzerland
+41 44 634 37 36

Michael Wolf is Professor of Econometrics and Applied Statistics at the University of Zurich and a Senior Fellow at ADIA Lab in Abu Dhabi.
michael.wolf@econ.uzh.ch
Zürichbergstrasse 14, 8032 Zurich, Switzerland
+41 44 634 50 96

Abstract

Tracking-error management is largely absent from the academic literature but ubiquitous in real life: Most portfolio managers are tied to a benchmark. Some of them aim to track a benchmark (such as the S&P 500), which is not necessarily a trivial task, since the benchmark often contain assets that are difficult or expensive to trade; in this case the objective is to *minimize* tracking error. Other managers aim to take on an active tilt without deviating too much from a benchmark; in this the objective is to *control* tracking error. In both cases managers need an estimator of the covariance matrix of many (excess) returns for their objective. This paper demonstrates the benefit of sophisticated shrinkage estimators (in conjunction with multivariate GARCH models) to this end, relative to the commonly used sample covariance matrix.

KEY WORDS: Covariance matrix estimation; multivariate GARCH; nonlinear shrinkage; portfolio selection; tracking error.

JEL CLASSIFICATION NOS: C13, C58, G11.

Three key takeaways:

- The paper addresses a variety of tracking-error problems that are relevant to portfolio managers such as the passive problem of tracking (or mimicking) a benchmark where the benchmark weights may be known or unknown, and the problem of implementing active tilts without deviating from a benchmark too much.
- The use of sophisticated shrinkage estimators of the covariance matrix, especially in conjunction with multivariate GARCH models, significantly improves tracking-error management compared to the commonly used sample covariance matrix and to naïve approaches that abstain entirely from estimating the covariance matrix.
- It is in the interest of portfolio managers who are tied to a benchmark to upgrade to sophisticated estimators of the covariance matrix in their fiduciary duty to adhere to best-practice for their investors.

Markowitz (1952) established quantitative methods in portfolio management by bringing to light the trade-off between risk (variance, volatility) and expected returns. Mathematics solves it by optimizing the amount of capital invested in each one of the various stocks in the universe under consideration. This, by itself, was a profound and long-overdue insight.

Markowitz's student Sharpe (1964) took it one step further by arguing that the variance must be measured not against the zero-return level or the risk-free rate, but against a passive value-weighted benchmark proxying for the so-called 'market portfolio'. If financial markets are efficient, then the benchmark is unbeatable, so it is up to self-proclaimed active investment managers to prove that they can beat it through special skills in selecting and timing stocks, and then assembling them judiciously into outperforming portfolios. Thus, the focus shifts from a Markowitz mean-variance trade-off to a Sharpe mean-tracking-error trade-off, relative to some benchmark index representing the investment universe of interest for the target investor or investor class.

Such is the real-world impact of these ideas that nowadays nearly all prospectuses of managed funds that seek to raise capital from investors must declare the specific index benchmark (among the many available) they are trying to beat or track. As Pastor et al. (2024, Section 5.4) point out, "the market share of indexing, relative to active management, has been steadily growing", making tracking-error management more important than ever. Indeed, the funds that do not abide are labeled "absolute return", which effectively banishes them to the "alternative investment" fringe, alongside hedge funds, cryptocurrencies, precious metals, etc. This typically limits them to small-percentage allocations of the overall wealth pie due to obvious risk concerns, tight regulatory oversight, and the natural inclination of financial-advisory platforms to err on the side of caution.

The present paper contributes to the tracking-error literature, whose place in academia is smaller than in the real world, by drawing from and adapting some cutting-edge research in the mean-variance literature, specifically in terms of estimating a key ingredient: the covariance matrix of many (excess) stock returns, whether unconditionally or conditionally. Because the investment universes involved are typically large (hundreds of stocks at least), shrinkage estimation will need to be applied, whether linearly or nonlinearly, in order to enhance accuracy and avoid the curse of dimensionality.
We study a wide and representative variety of tracking-error problems of practical interest:

1. tracking a benchmark whose weights are known using a restricted (smaller, or different) investment universe;
2. tracking a benchmark whose weights are unknown (where only its returns are observed);
3. taking on an 'active' tilt without deviating too much from a benchmark.

In turn, these different tracking-error problems are criss-crossed with realistic design choices:

a. implementing the above with or without a constraint outlawing short sales;
b. incorporating the tracking error either as a penalty term in the objective function, or as a constraint imposed upon the optimizer;
c. experimenting across a range of different universe sizes, up to a thousand.

The overall contribution of this paper, based on real-data backtest simulations, is that the more advanced covariance matrix technologies make a strongly valuable impact in the field of tracking error. We can safely single out the DCC-NL model of Engle et al. (2019), which combines multivariate GARCH with nonlinear shrinkage, as the most convincing all-around performer. However, in certain circumstances detected and highlighted in the empirical part of the paper, Quadratic Inverse Shrinkage (QIS, Ledoit and Wolf (2022b)), which is simpler because it is an unconditional model, can perform slightly better; and sometimes an even simpler model, the linear shrinkage formula of De Nard (2022), can be the laureate. But what is clearly established here is that continuing to use the 'textbook' sample covariance matrix in order to manage tracking error[1] is clearly suboptimal and outdated for all strategies that are benchmarked against a passive index. Quantitative portfolio managers who persevere in using the sample covariance matrix for tracking-error minimization in large universes may eventually face questions on whether they have been legally derelict in their fiduciary duty to adhere to best-practice for their investors.

Given that the main focus here is on the development of feasible investment strategies, we look at other considerations beyond just minimizing tracking error. One of them is ex ante optimism: the notion that ex post realized tracking error may be excessive due to in-sample overfitting. Here again, we find that shrinkage, especially of the DCC-NL type, mitigates the problem greatly, to the point that it becomes almost negligible in certain important backtest configurations. In terms of portfolio turnover and (when applicable) leverage, the same hierarchy between covariance matrix estimation techniques is reaffirmed overall.

Literature Review

Zenti and Pallotta (2002) highlight the importance of some themes that are key to the problem at hand:

1. the potential gap between ex ante and ex post tracking errors (which we carefully measure in this paper);
2. the need to take into account time-varying variances and correlations (as we do in the DCC-NL model described below, and which is the one that we champion);

[1] Depending on the objective, managing tracking error can mean either minimizing tracking error or controlling tracking error.

3. the need to recalibrate and rebalance on a regular basis (we do 21 trading days, which is near the middle of the span of frequencies that they deem worthy of consideration).

The two main limitations of their early paper are that: (i) they have a modest number of stocks in the universe (50, whereas we can easily go all the way up to 1000); and (ii) they treat portfolio selection as essentially an exogenous process, whereas, in fact, it typically involves a kind of mean-variance or mean-tracking error optimization that uses as input a somewhat erroneous covariance matrix estimator derived from the *same* historical data set as the EGARCH model used for tracking-error control, so the exogeneity hypothesis is hardly valid (cf. Exhibit 5 and the remarks surrounding it).

Jorion (2003) documents the advantages of taking into account overall portfolio variance, even in a tracking-error framework. This is exactly what we do in the sections on active portfolios. However his article is theoretical in nature, as evidenced by his sentence on page 72: "Expected returns are arbitrary and were chosen so as to satisfy the efficient-set parameter." By contrast, we buttress our theoretical developments with realistic backtests run on historical return data. Also, with only five assets in the investable universe, his dimensionality is too limited for practical purposes.

Ledoit and Wolf (2004) were the first to use (linear) shrinkage estimation of the large-dimensional covariance matrix for the purpose of minimizing tracking error. They focused only on the active part of the problem, whereas we also address the passive part, in two different ways. Also, their active component hinged on a forecast of expected returns that was somewhat manufactured, as it involved a deliberately controlled amount of look-ahead; by contrast, our active component uses the Jorion (2003) concept of a global-minimum-variance tilt (or constraint) that is fully backtestable and tradeable point-in-time. Finally, shrinkage technology for large-dimensional covariance matrices has matured considerably over the last two decades, particularly in the directions of (i) a better-suited linear-shrinkage target (De Nard, 2022), (ii) nonlinear shrinkage (Ledoit and Wolf, 2022b), and (iii) conditional heteroskedasticity (Engle et al., 2019), so we incorporate these three upgrades here.

Belhaj et al. (2005) have a narrower focus, the gap between ex ante estimated and ex post realized tracking errors, which can be called for brevity *in-sample optimism*. These authors explore the question largely from a theoretical point of view, but one of their great merits is that they highlight that the tracking-error problem differs between active and passive portfolio managers, a subtle yet essential distinction that we have incorporated into the very structure of our paper, both in the theoretical and in the empirical parts.

Basak et al. (2009) have the same narrow focus as above, but they escalate to more of a hands-on numerical solution: they use the jackknife to put a damper on in-sample optimism. In our paper, excessive optimism is a concern that we strenuously monitor and address. However, the preeminent objective of any benchmarked portfolio manager should be to find the covariance matrix estimator that most reduces ex post realized tracking error. One substantive contribution of ours is going beyond the standard factor models of Basak et al. (2009) to shrinkage techniques that not only reduce in-sample optimism but at the same time have the virtue of reducing ex-post tracking error.

Tracking Error in Portfolio Optimization

Passive: Minimum-Tracking-Error Portfolios

We start with the most common problem considered in this strand of literature. The goal is to track a known portfolio, also called the benchmark, made up from a given universe of stocks, which is typically large. Therefore, at any given point in time, one knows both the constituents of the portfolio and the corresponding vector of portfolio weights. Clearly, this a passive investment strategy.

Denote the vector of stock returns at time t by $r_t := (r_{t,1}, \ldots, r_{t,N})'$ where N expresses the size of the benchmark investment universe. Furthermore, denote the (known) vector of benchmark portfolio weights at time t by $w_{\text{BM},t}$. Depending on index management, N could in theory change as a function of t, but the notation N_t is too unwieldy, so with this caveat in mind we keep the notation N for the sake of simplicity.

Known Benchmark Weights

Often the benchmark (portfolio) contains some small and illiquid stocks that are difficult or expensive to trade in practice. As a result, even if one knows the constituents and the corresponding (portfolio) weights of the benchmark, it becomes demanding and costly to exactly trade it. Therefore, it is often desirable to use only a subset of $\tilde{N} < N$ stocks that are sufficiently easy and inexpensive to trade, which are called the "eligible stocks" or the "eligible universe". As an example consider the problem of tracking the Russell 2000 index ($N = 2000$) using only the $\tilde{N} = 500$ or $\tilde{N} = 1000$, say, most liquid stocks in the Russell 2000 universe. The goal is to track the benchmark 'as well as possible' using the subset of eligible stocks only, where 'as well as possible' is measured by the variance of the tracking error (that is, the difference between the feasible portfolio and the benchmark). If we denote the (conditional) covariance matrix of r_t by $\Sigma_{r,t}$ the portfolio-selection problem thus becomes

$$\min_{w} \quad (w - w_{\text{BM},t})'\Sigma_{r,t}(w - w_{\text{BM},t}) \tag{1}$$

$$\text{s.t.} \quad w'\mathbb{1} = 1 \ , \tag{2}$$

$$w_i = 0 \quad \forall i \in \text{ineligible} \ , \tag{3}$$

$$(w_i \geq 0) \quad \forall i \ . \tag{4}$$

Here, the symbol $\mathbb{1}$ in (2) denotes a conformable vector of ones and the constraint signifies that the feasible portfolio must be fully invested. The constraint (3) signifies that one can only invest in the subset of $\tilde{N}$ eligible stocks; clearly, this subset needs to be exogenously determined before the portfolio selection can take place. Finally, the optional constraint (4) signifies that the feasible portfolio must be long only; such a constraint would be typically in place but not necessarily always.

There might be a reduced-information setting where the portfolio managers only has access to the benchmark returns without knowing the portfolio weights of the benchmark. This can be considered a special case of having access to benchmark returns without even knowing the portfolio constituents necessarily, which is dealt with in the next section. Intuitively, in absence of the corresponding weights, knowing the portfolio constituents actually has limited benefit.

Unknown Benchmark Weights

At time t one observes $\tilde{x}_t := \tilde{r}_t - r_{\text{BM},t}$, where $\tilde{r}_t \subseteq r_t$ is the vector of returns that corresponds to the eligible universe and $r_{\text{BM},t}$ is the return on the benchmark whose portfolio weights are unknown (and where it does not really matter whether its constituents are known or unknown). The problem formulation then becomes

$$\min_{w} \quad w'\Sigma_{\tilde{x},t}w \tag{5}$$

$$\text{s.t.} \quad w'\mathbb{1} = 1 \ , \tag{6}$$

$$(w_i \geq 0) \ , \tag{7}$$

The resulting portfolios are sometimes referred to as benchmark-following or fund-mimicking portfolios, where the convention seems to be that the former term is generally used when the portfolio constituents are known whereas the latter is generally used when they are unknown; but, as stated before, knowledge of the portfolio constituents is actually irrelevant when their portfolio weights in the benchmark are unknown. In an extreme case, the benchmark might even belong (partly) to different asset classes, such as when one uses an eligible universe of stocks to mimic a multi-asset portfolio that contains, apart from stocks, also bonds, commodities, hedge funds, and cash, say. Obviously, the more 'distant' the investment universe of the benchmark is from the eligible universe of stocks, the more difficult the problem of mimicking the benchmark becomes.

Active: Strategies with Tracking-Error Considerations

We now turn attention to active investment strategies that are 'tied' to a given benchmark comprised of N stocks. For simplicity, we will here assume that all N stock are investable or eligible for the fund manager so that $\tilde{N} = N$.[2] The goal of the manager is to design an investment strategy that 'performs well' in an absolute sense, but does not 'deviate too much' from the benchmark.

Strategies with Tracking Error as Part of the Objective Function

One possible problem formulation includes the tracking error in the objective function:

$$\max_{w} \quad \delta \cdot \mathit{active} - (1-\delta) \cdot (w - w_{\mathrm{BM},t})' \Sigma_{r,t} (w - w_{\mathrm{BM},t}) \tag{8}$$

$$\text{s.t.} \quad w' \mathbb{1} = 1 \ , \tag{9}$$

$$(w_i \geq 0) \ , \tag{10}$$

where $\delta \in [0,1]$ is constant chosen by the investment manager. Here *active* denotes a measure for the active strategy that the manager would like to 'maximize'. In other words, the manager wants to maximize a convex combination of the 'quality' of the active portfolio and the negative of the variance of the tracking error relative to the benchmark. The parameter δ determines the weight / importance assigned to the active part and ranges from 0 (minimum-tracking-error portfolio) to 1 (active portfolio only, without tracking-error consideration).

For concreteness, and to abstain from having to estimate / forecast expected asset returns, in our empirical application below, we will focus on the global minimum variance (GMV) portfolio for the active portfolio which then results in $\mathit{active} \coloneqq -w' \Sigma_{r,t} w$. In this application then, the manager seeks to minimize a convex combination of the variance of the portfolio and the variance of the tracking error, subject to being fully invested and (if desired) being long only.

[2] This is not a necessary assumption, since one could also introduce in this context an eligible and an ineligible universe for the active manager.

Strategies with Tracking-Error Constraint

Another possible formulation includes the tracking error in the list of constraints rather than in the objective function:

$$\max_{w} \quad \textit{active} \tag{11}$$

$$\text{s.t.} \quad w'\mathbb{1} = 1 \ , \tag{12}$$

$$\sqrt{(w - w_{\text{BM},t})'\Sigma_{r,t}(w - w_{\text{BM},t})} \leq \tau \ , \tag{13}$$

$$(w_i \geq 0) \ , \tag{14}$$

where $\tau > 0$ is a constant chosen by the investment manager. In other words, the manager wants to maximize the 'criterion' of the active strategy, subject to being fully invested, (if desired) being long only, and to imposing an upper bound on the standard deviation of the tracking error relative to the benchmark. For concreteness, and to abstain from having to estimate / forecast expected asset returns, in our empirical application below, we will again focus on the GMV portfolio for the active portfolio which then results in $\textit{active} \coloneqq -w'\Sigma_{r,t}w$.

Tracking-Error Estimation

For any vector of portfolio weights w, the variance of the tracking error at time t is given by $(w - w_{\text{BM},t})'\Sigma_{r,t}(w - w_{\text{BM},t})$ for known benchmark weights, respectively by $w'\Sigma_{\tilde{x},t}w$ for unknown benchmark weights. Estimation of the tracking-error variance is 'reduced' to estimation of a covariance matrix $\Sigma_{r,t}$, respectively $\Sigma_{\tilde{x},t}$, which can be based on a history of past stock returns $\{r_t\}$, respectively tracking-error returns $\{\tilde{x}_t\}$.

Covariance Matrix Estimation

When N is small, such as $N = 30$ for the Dow-Jones 30 index, the sample covariance matrix would be a satisfactory choice. But when N is large, such as $N = 500$ for the S&P 500 index, $N = 2000$ for the Russell 2000 index, or even $N > 2000$ for the MSCI ACWI Index, the sample covariance matrix no longer works well. The problem of estimating a covariance matrix of asset returns when the size of the investment universe is large has been studied extensively in the literature. In this paper, we will make use of (non)linear shrinkage estimators that have been proposed by the authors over the last 20+ years and refer the reader not already familiar with these estimators to the overview paper of Ledoit and Wolf (2022a).

Empirical Analysis

Data and General Portfolio-Construction Rules

We download daily stock return data from the Center for Research in Security Prices (CRSP) starting on 01/01/1978 and ending on 12/31/2022. We restrict attention to stocks from the NYSE, AMEX, and NASDAQ stock exchanges.

For simplicity, and in line with much of the literature, we adopt the common convention that 21 consecutive trading days constitute one (trading) 'month'. The out-of-sample period ranges from 07/02/1982 through 12/30/2022, resulting in a total of 479 months (or 10,059 days). All portfolios are updated monthly.[3] We denote the investment dates by $k = 1, \ldots, 479$. At any investment date k, an $N \times N$, respectively an $\tilde{N} \times \tilde{N}$ covariance matrix is estimated based on the most recent 1260 daily (raw) returns, r_t, respectively daily eligible excess returns, $\tilde{x}_t$, which roughly corresponds to using five years of past data.

We consider investment universes up to $N = 1000$. For a given combination (k, N), the investment universe is obtained as follows. We find the set of stocks that have an almost complete return history over the most recent $T = 1260$ days as well as a complete return 'future' over the next 21 days.[4]

From the remaining set of stocks, we then pick the largest N, respectively largest $\tilde{N}$, stocks (as measured by their market capitalization on investment date k) as our investment universe, respectively eligible investment universe. The ineligible investment universe consists of the M smallest stocks where $M \in \{0, 1, \ldots, N - \tilde{N}\}$. In this way, the (entire, eligible and ineligible) investment universe changes relatively slowly from one investment date to the next.

There is a great advantage in having a well-defined rule that does not involve drawing stocks at random, as such a scheme would have to be replicated many times and averaged over to give stable results. As far as rules go, the one we have chosen seems the most reasonable because it avoids so-called "penny stocks" whose behavior is often erratic; also, high-market-cap stocks tend to have the lowest bid-ask spreads and the highest depth in the order book, which allows large investment funds to invest in them without breaching standard safety guidelines. Finally, the benchmark indices are often market-cap weighted, stressing the importance of large caps while reducing the impact of small caps for tracking-error management.

[3]Monthly updating is common practice to avoid an unreasonable amount of turnover and thus transaction costs. During a month, from one day to the next, we hold number of shares fixed rather than portfolio weights; in this way, there are no transactions during a month.

[4]The first restriction allows for up to 2.5% of missing returns over the most recent 1260 days, and replaces missing values by zero. The latter, 'forward-looking' restriction is not feasible in practice but is commonly used in the literature. Although it might affect (in a minor way) absolute performance due to survivorship bias, it does not systematically affect relative performance of various methods.

Competing Covariance Matrix Estimators

To compute the ex ante tracking error, by which we mean the tracking error implied by a given estimator of the covariance matrix, the following estimators are included in our analysis:

- **S:** the **s**ample covariance matrix.
- **L:** the **l**inear shrinkage estimator of De Nard (2022); that is, we use the constant-variance-covariance (CVC) shrinkage target with a constant variance on the diagonal and a constant covariance on the off-diagonal.
- **NL:** the **n**on**l**inear shrinkage estimator of Ledoit and Wolf (2022b), that is, we use the quadratic-inverse shrinkage (QIS) estimator.
- **DCC-NL:** the multivariate GARCH model of Engle et al. (2019) where the unconditional correlation matrix is estimated via **n**on**l**inear shrinkage (QIS). We use an 'averaged-forecasting' approach as proposed by De Nard et al. (2021) and De Nard et al. (2022).

We also include the following naïve investment strategies that abstain from any estimation of the tracking error (and thus from any estimation of the covariance matrix):

- **VW-E:** the **v**alue-**w**eighted portfolio of the **e**ligible universe, based on market cap.
- **EW-E:** the **e**qually-**w**eighted portfolio of the **e**ligible universe.

Benchmarks

In any application we need to select a benchmark relative to which we want to manage tracking error. We include the following three benchmarks in our analysis.

- **VW:** the long-only **v**alue-**w**eighted portfolio.
- **EW:** the long-only **e**qually-**w**eighted portfolio.
- **Markowitz:** the "130/30" long-short Markowitz portfolio with momentum signal: for a detailed description see Appendix A.

For the applications with *known* benchmark weights, in each case the constituents of the benchmark are all $N = 1000$ stocks in the investment universe (at any given investment date k). On the other hand, for the applications with *unknown* benchmark weights, in each case the constituents of the benchmark are the $M = 200$ smallest stocks in terms of market capitalization (at any given investment date k).

Performance Measures

- **TE:** We measure the overall ex post **t**racking **e**rror by computing the standard deviation of the 10,059 out-of-sample portfolio returns in excess of the benchmark; similarly, we measure the ex post tracking error for month k by computing the same measure of the 21 out-of-sample returns during the month only. In each case, we then multiply by $\sqrt{252}$ to annualize. We compute the ex ante tracking error for month k as $\sqrt{(\hat{w}_k - w_{\text{BM},k})'\hat{\Sigma}_{r,k}(\hat{w}_k - w_{\text{BM},k})}$, respectively $\sqrt{\hat{\tilde{w}}_k'\hat{\Sigma}_{\tilde{x},k}\hat{\tilde{w}}_k}$, where we use the same estimator of the covariance matrix that was used for constructing the portfolio (that is, for constructing $\hat{w}_k$, respectively $\hat{\tilde{w}}_k$) in the first place, and then multiply by $\sqrt{252}$ to annualize. The portfolio $\hat{w}_k$ corresponds to known benchmark weights, and $\hat{\tilde{w}}_k$ to unknown ones.
- **MR:** We compute the **m**ean **r**atio of the monthly ex ante to the ex post tracking error as $\frac{1}{479}\sum_{k=1}^{479}$ ex ante TE_k/ex post TE_k.
- **MD:** We compute the **m**ean **d**ifference between the monthly ex ante and ex post tracking errors as $\frac{1}{479}\sum_{k=1}^{479}(\text{ex ante TE}_k - \text{ex post TE}_k)$, and then multiply by $\sqrt{252}$ to annualize.
- **MAD:** We compute the **m**ean **a**bsolute **d**ifference of the monthly ex ante vs. the ex post tracking error as $\frac{1}{479}\sum_{k=1}^{479}|\text{ex ante TE}_k - \text{ex post TE}_k|$, and then multiply by $\sqrt{252}$ to annualize.
- **SD:** We compute the **s**tandard **d**eviation of the 10,059 out-of-sample returns, and then multiply by $\sqrt{252}$ to annualize.
- **SD*:** We compute the **s**tandard **d**eviation of the out-of-sample returns for all days where the constraint (13) is fulfilled for all competitors, and then multiply by $\sqrt{252}$ to annualize.
- **Success:** We compute the success rate of the one-month ex post tracking error vs. its constraint (13): $\frac{1}{479}\sum_{k=1}^{479}\mathbb{1}_{\{\text{ex post TE}_k\leq\tau\}}$, where $\mathbb{1}_{\{\cdot\}}$ denotes the indicator function.
- **Obj:** We compute the realized **obj**ective of (8) as $\sqrt{\delta\cdot\text{SD}^2 + (1-\delta)\cdot\text{TE}^2}$.
- **TO:** We compute average (monthly) **t**urn**o**ver as $\frac{1}{478}\sum_{k=1}^{478}||\hat{w}_{k+1} - \hat{w}_k^{\text{hold}}||_1$, where $||\cdot||_1$ denotes the L^1 norm and $\hat{w}_k^{\text{hold}}$ denotes the vector of the 'hold' portfolio weights at the end of month k.[5]
- **GL:** We compute average (monthly) excess **g**ross **l**everage as $\frac{1}{479}\sum_{k=1}^{479}||\hat{w}_k||_1 - 1$.

[5] The vector $\hat{w}_k^{\text{hold}}$ is determined by the initial vector of portfolio weights, $\hat{w}_k$, together with the evolution of the various prices of the N stocks in the portfolio during month k.

Passive Minimum-Tracking-Error Portfolios

Known Benchmark Weights

The problem formulation of passive minimum-tracking-error portfolios with known benchmark weights is described above in (1–4). Exhibit 1 presents the results on ex post tracking error (TE) which can be summarized as follows:

- For any scenario, as expected, TE decreases as the size of the eligible universe, $\tilde{N}$, increases: A larger universe of eligible stocks makes it easier to track the benchmark.
- Being able to short stocks is often (weakly) beneficial: The differences range from 'zero' to 'substantial', where the most substantial differences are observed for the Markowitz benchmark.

Ex post tracking errors									
	Long-short				Long-only				
	S	L	NL	DCC-NL	Naïve	S	L	NL	DCC-NL
	Value-weighted benchmark								
$\tilde{N} = 100$	1.92	1.90**	1.92	**1.88*** **	2.82	1.92	1.90**	1.92	**1.88*** **
$\tilde{N} = 500$	0.37	0.33***	0.32***	**0.31*** **	0.73	0.34	0.32***	0.31***	**0.31*** **
$\tilde{N} = 800$	0.23	0.13***	0.10***	**0.10*** **	0.22	0.19	0.17***	0.17***	**0.17*** **
	Equally-weighted benchmark								
$\tilde{N} = 100$	5.57	5.49***	5.51***	**5.43*** **	6.61	5.52	5.49*	5.52	**5.43*** **
$\tilde{N} = 500$	2.21	2.03***	1.98***	**1.93*** **	2.95	2.04	1.99***	1.98***	**1.93*** **
$\tilde{N} = 800$	1.02	0.89***	0.84***	**0.81*** **	1.22	0.87	0.85***	0.84***	**0.81*** **
	Markowitz benchmark								
$\tilde{N} = 100$	10.00	9.78***	9.74***	**9.36*** **	14.61	10.30	10.28	10.28	**9.87*** **
$\tilde{N} = 500$	6.51	5.80***	5.62***	**5.54*** **	14.18	7.12	7.06***	7.06***	**6.76*** **
$\tilde{N} = 800$	4.77	3.53***	3.29***	**3.20*** **	14.31	5.91	5.84***	5.84***	**5.63*** **

Exhibit 1. Annualized ex post tracking-error numbers of minimum-tracking-error portfolios, in percent. All numbers are based on the 10,059 daily out-of-sample excess returns from 07/02/1982 until 12/30/2022. For any row, the lowest (and thus best) number appears in **bold face** and significant outperformance over S is denoted by asterisks: *** denotes significance at the 0.01 level; ** denotes significance at the 0.05 level; and * denotes significance at the 0.1 level.

- TE is generally best (that is, lowest) for the the value-weighted benchmark, followed by the equally-weighted benchmark, followed by the Markowitz benchmark.
- The overall ranking, from best to worst, is DCC-NL, NL, L, S and Naïve, where Naïve corresponds to VW-E for the VW benchmark and to EW-E for the EW and Markowitz benchmark. Therefore, taking into account the covariance matrix is crucial for minimum-tracking-error portfolios.
- With the single exception of $\tilde{N} = 100$ for the value-weighted benchmark, all shrinkage estimators outperform the sample covariance matrix.
- In particular, DCC-NL gives the uniformly lowest TEs across all scenarios, which is also always lower than the corresponding TE based on S with statistical significance.

We also study whether one estimator delivers a lower ex post tracking error than another estimator with statistical significance. To reduce a multiple testing problem and since a major goal of this paper is to show that using sophisticated shrinkage estimators (in conjunction with multivariate GARCH models) improve minimum-tracking-error portfolios, we restrict attention to three comparisons: (i) S with L, (ii) S with NL, and (iii) S with DCC-NL. For a given scenario, a two-sided p-value for the null hypothesis of equal TE is obtained by the prewhitened HAC_{PW} method described in Ledoit and Wolf (2011, Section 3.1).[6] With the exception of some 'smaller' L and NL minimum-tracking-error portfolios ($\tilde{N} = 100$), the outperformance of the shrinkage estimators over S is always statistically significant and, arguably, economically meaningful as well.

In sum, upgrading from the sample covariance matrix to a shrinkage estimator of the covariance matrix is clearly beneficial in terms of reducing ex post tracking error, with the best choice being DCC-NL.

Exhibit 1 presents ex post tracking errors over the entire out-of-sample period. As a robustness check, we can in addition study monthly ex post tracking errors. To this end we restrict attention to long-short portfolios based on $\tilde{N} = 800$ stocks tracking the value-weighted benchmark. The upper panel of Exhibit 2 plots the time series of monthly ex post tracking errors for S and DCC-NL.[7] It can be seen that DCC-NL series lies 'uniformly' below the S series, so that the outperformance of DCC-NL over S is not restricted to certain periods but holds throughout. The lower panel of Exhibit 2 compares the distributions of the monthly ex post tracking errors by lining up the corresponding violin plots.[8] The benefit of L, NL,

[6] As the out-of-sample size is very large at 10,059, there is no need to use the computationally more involved bootstrap method described in Ledoit and Wolf (2011, Section 3.2), which is preferred for small sample sizes.

[7] We only include two covariance matrix estimator in this plot to keep the plot visually 'digestable'.

[8] A violin plot visually summarizes the distribution of a data set by combining the corresponding box plot with a (rotated) kernel density estimator; note that within the box of the box plot the sample median is indicated by a horizontal line whereas the sample average is indicated by a diamond.

and DCC-NL relative to S can clearly be seen, since their distributions lie ‘below’ that of S; the overall winners are NL and DCC-NL. In passing, it is also noteworthy that the naïve VW-portfolio (that is, the value-weighted portfolio of the eligible universe) actually performs better according this metric than the tracking portfolio based on the sample covariance matrix.

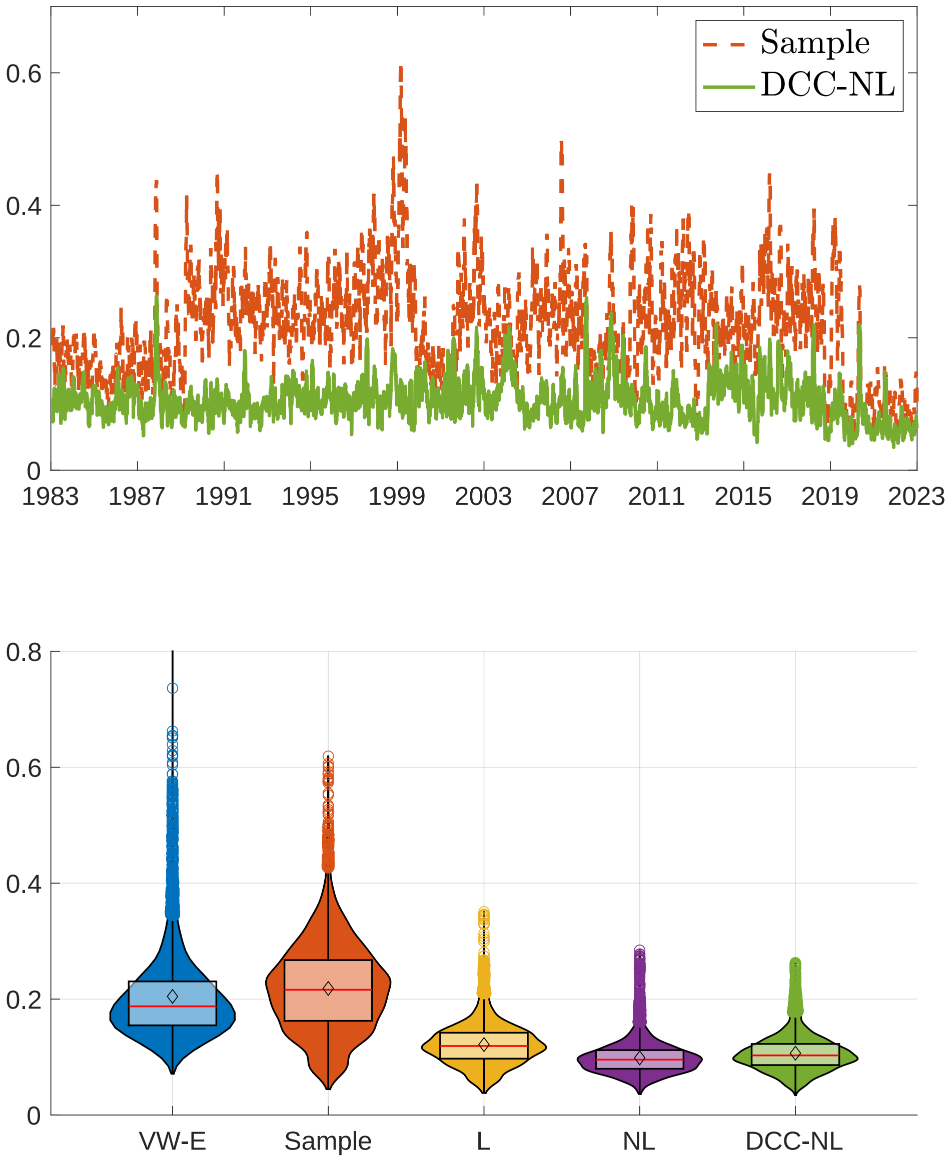


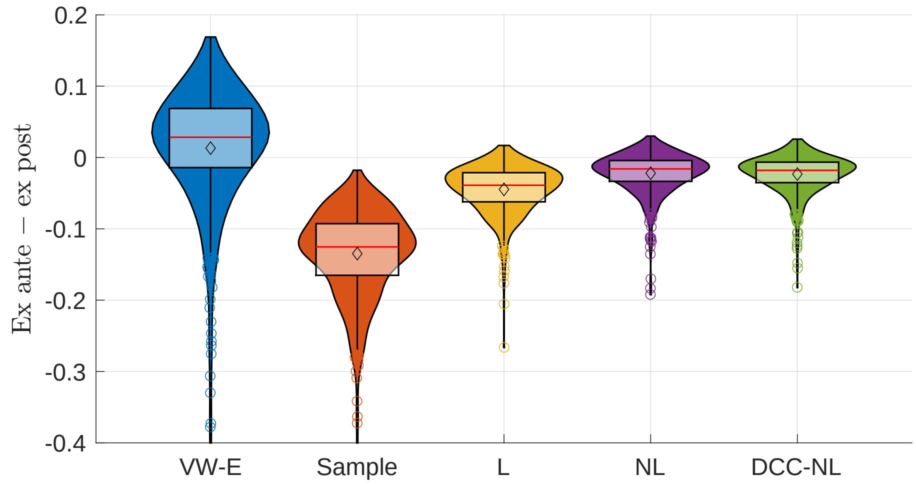
Exhibit 2. Monthly ex post tracking-errors (upper panel) and their distribution (lower panel) for long-short minimum-tracking-error portfolios ($\tilde{N} = 800$) tracking the value-weighted benchmark, in percent.

We next compare monthly ex ante with ex post tracking errors by looking at the mean ratio (MR), the mean difference (MD), and the mean absolute difference (MAD) for each scenario in Exhibit 3. First, the mean ratio should ideally be equal to one, with values less than one indicating that the ex ante estimator tends to be 'optimistic' regarding the ex post realization. It can be seen that all methods are generally optimistic and that optimism increases with $\tilde{N}$; overall, S performs the worst whereas NL and DCC-NL perform the best. Second, the mean difference should ideally be equal to zero, with values less than zero indicating that the ex ante estimator tends to be 'optimistic' regarding the ex post realization. It can be seen that all methods are generally optimistic and that optimism increases with $\tilde{N}$; overall, S performs the worst whereas NL and DCC-NL perform the best. Third, the mean absolute difference should be small; overall, S performs the worst whereas DCC-NL performs the best.

Monthly ex ante vs. ex post tracking-errors of long-short portfolios												
	Mean ratio				Mean difference				Mean absolute difference			
	S	L	NL	DCC-NL	S	L	NL	DCC-NL	S	L	NL	DCC-NL
	Value-weighted benchmark											
$\tilde{N} = 100$	**1.01**	1.08	1.13	0.96	−0.15	−0.04	**0.04**	−0.18	0.47	0.47	0.49	**0.36**
$\tilde{N} = 500$	0.63	0.81	**0.91**	0.84	−0.14	−0.08	**−0.05**	−0.06	0.15	0.09	0.07	**0.07**
$\tilde{N} = 800$	0.39	0.67	0.82	**0.82**	−0.13	−0.04	−0.02	**−0.02**	0.13	0.05	0.03	**0.03**
	Equally-weighted benchmark											
$\tilde{N} = 100$	**1.00**	1.04	1.08	0.95	−0.50	−0.30	**−0.18**	−0.62	1.38	1.38	1.43	**1.10**
$\tilde{N} = 500$	0.63	0.79	**0.88**	0.82	−0.87	−0.53	**−0.37**	−0.43	0.88	0.57	0.48	**0.46**
$\tilde{N} = 800$	0.44	0.64	0.78	**0.82**	−0.57	−0.34	−0.22	**−0.18**	0.57	0.34	0.24	**0.20**
	Markowitz benchmark											
$\tilde{N} = 100$	0.94	**1.03**	1.14	0.93	−1.33	−0.59	**0.28**	−1.20	2.37	2.35	2.55	**1.91**
$\tilde{N} = 500$	0.59	0.83	**1.00**	0.87	−2.75	−1.28	**−0.28**	−0.96	2.75	1.50	1.29	**1.19**
$\tilde{N} = 800$	0.35	0.70	0.95	**0.90**	−2.99	−1.13	**−0.35**	−0.47	2.99	1.18	0.78	**0.68**

Exhibit 3. Monthly ex ante vs. ex post tracking errors of long-short minimum-tracking-error portfolios, in percent. All numbers are based on 10,059 daily out-of-sample excess returns from 07/02/1982 until 12/30/2022. For any row, the best performer appears in **bold face**.

Exhibit 4 provides a graphical illustration for long-short portfolios based on $\tilde{N} = 800$ eligible stocks tracking the value-weighted benchmark. The upper panel shows violin plots for the monthly differences between ex ante and ex post tracking errors whereas the lower panels shows violin plots for the corresponding absolute differences. Again, the benefit of upgrading from the sample covariance matrix to shrinkage estimators is visually apparent, wiht NL and DCC-NL performing the best.



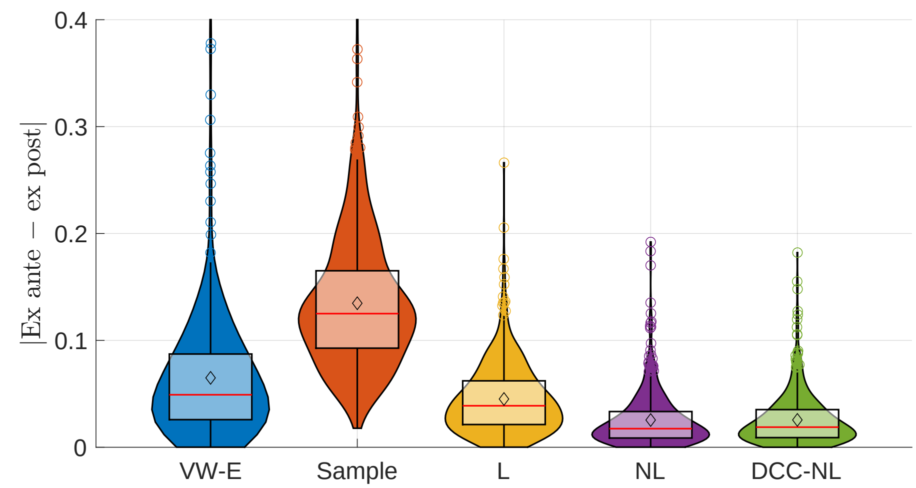


Exhibit 4. Violin plots of monthly ex ante vs. ex post tracking-error differences (upper panel) and absolute differences (lower panel) for long-short minimum-tracking-error portfolios ($\tilde{N} = 800$) tracking the value-weighted benchmark, in percent.

So far, the comparisons of ex ante and ex post tracking errors have been carried out "within methods". This means that after a method has been used to construct a portfolio, the same method has been used to compute the corresponding ex ante tracking error. Naturally, doing so generally yields 'optimistic' results, especially for large $\tilde{N}$.[9]

Alternatively, one can look at the ability of a given method to evaluate (or estimate) the ex post tracking error of an arbitrary portfolio.[10] This question is addressed in Exhibit 5

[9]The situation is akin to asking someone to pick a team he/she thinks will win a sports championship, say the Premier League, and then asking the same person to predict the number of points this team will have by the end of the season.

[10]The situation is akin to asking someone to pick a team he/she thinks will win a sports championship, say

where in the left panels we use the sample covariance matrix to compute ex ante tracking errors (for the five portfolios considered) and in the right panels we use DCC-NL to compute ex ante tracking errors. The upper two panels show that whereas S is most optimistic for its own portfolio (as expected) it is also optimistic for the shrinkage-based portfolios, although to a lesser extent, and pessimistic for the VW-E portfolio. On the other hand, DCC-NL tends to be quite realistic for all five portfolios, though of course somewhat optimistic for its own portfolio. In addition, the two lower panels demonstrate that the absolute mean difference is lower in distribution for all five portfolios when upgrading from the sample covariance matrix to DCC-NL. This exercise clearly demonstrates that there is a benefit in using a more accurate estimator of the covariance matrix not only for portfolio construction but also for portfolio evaluation.

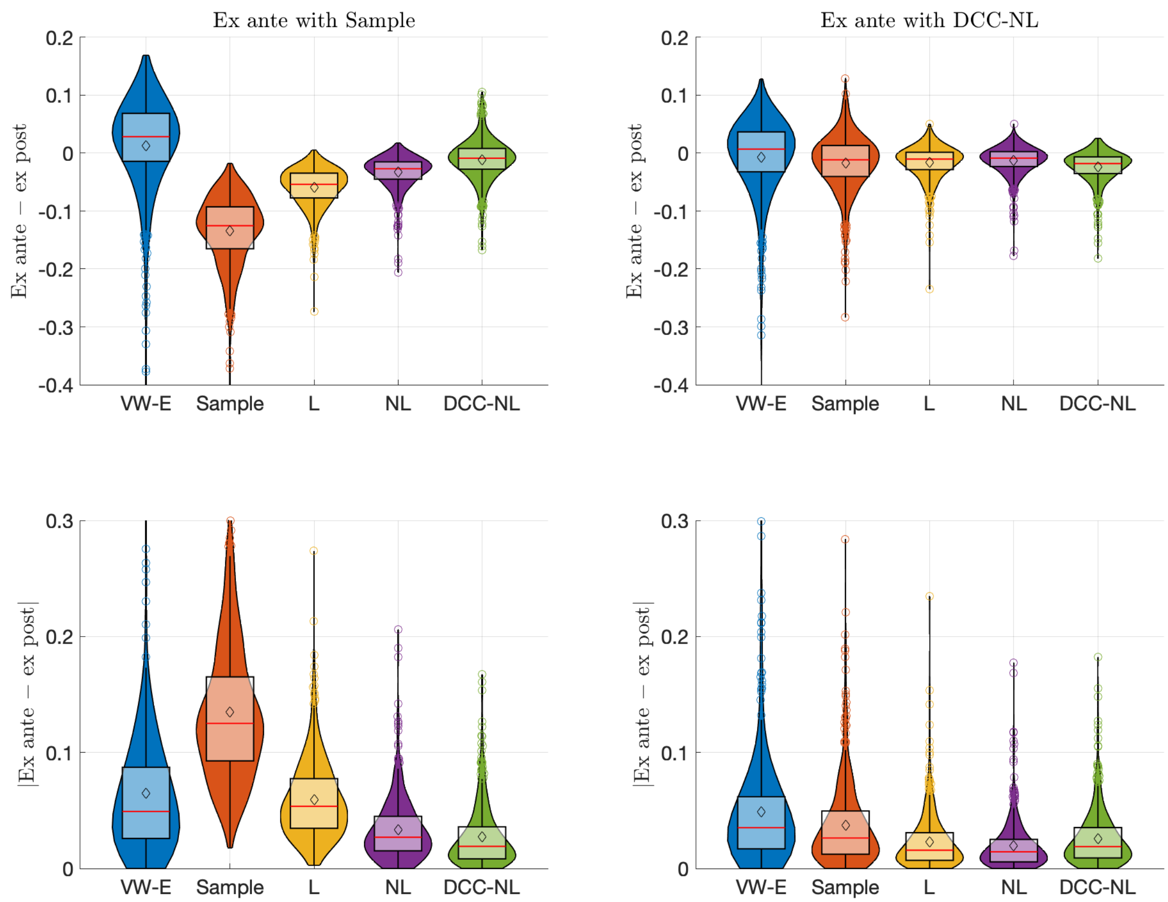


Exhibit 5. Violin plots of monthly ex ante vs. ex post tracking-errors of long-short minimum-tracking-error portfolios ($\tilde{N} = 800$) tracking the value-weighted benchmark, in percent. The ex ante tracking errors in the panels on the left (right) are computed with the sample covariance matrix (DCC-NL).

the Premier League, and then asking *another* person to predict the number of points this team will have by the end of the season.

We next turn to the additional performance measures of turnover and gross leverage. The results for long-short portfolios are presented in Exhibit 6 and can be summarized as follows:

- On balance, NL generates the least turnover and S generates the highest turnover. The comparison between L and DCC-NL is not so clear but at least for large $\tilde{N}$ it holds that DCC-NL generates less turnover than L despite being a dynamic model.
- On balance, NL generates the least gross leverage and S generates the highest gross leverage. The comparison between L and DCC-NL is not so clear but at least for large $\tilde{N}$ it holds that DCC-NL generates less gross leverage than L.

Characteristics of long-short portfolios								
	Turnover				Gross Leverage			
	S	L	NL	DCC-NL	S	L	NL	DCC-NL
	Value-weighted benchmark							
$\tilde{N} = 100$	16.13	12.64	**10.36**	20.97	0.90	0.06	**0.03**	0.23
$\tilde{N} = 500$	13.29	9.89	**7.96**	9.74	0.71	0.06	**0.00**	0.01
$\tilde{N} = 800$	12.78	8.32	**6.91**	7.56	1.73	0.07	**0.00**	0.02
	Equally-weighted benchmark							
$\tilde{N} = 100$	34.57	25.05	**19.24**	53.81	44.66	21.41	**10.28**	25.24
$\tilde{N} = 500$	57.97	37.16	**24.06**	38.07	50.27	20.80	**6.46**	14.86
$\tilde{N} = 800$	49.74	31.77	**20.10**	21.96	10.97	1.89	**0.07**	0.52
	Markowitz benchmark							
$\tilde{N} = 100$	68.05	48.86	**39.06**	116	148	94.12	**77.32**	84.17
$\tilde{N} = 500$	217	130	**93.93**	146	324	193	137	**133**
$\tilde{N} = 800$	330	168	**124**	137	364	178	124	**116**

Exhibit 6. Additional performance measures of long-short minimum-tracking-error portfolio, in percent. All numbers are based on 479 monthly weight vectors from 07/02/1982 until 12/30/2022. For any row, the lowest (and thus best) number appears in **bold face**.

As a final robustness check we re-run the numbers of Exhibit 1, which are the ones of most interest, for the long-short portfolios but with a shorter estimation window: Instead of using the past $T = 1260$ days to estimate a covariance matrix, we now use the past $T = 504$ respectively $T = 252$ days only; in other words, instead of using (roughly) five years of past data, we now use (roughly) two respectively one year of past data only. Intuitively, a shorter estimation window should favor the 'static' shrinkage estimators L and NL relative to the 'dynamic' shrinkage estimator DCC-NL. However, as shown in Exhibit 7, DCC-NL continues to yield the best results across all scenarios and also for shorter estimation windows. Again, all shrinkage estimators consistently outperform the sample covariance matrix and the naïve strategies.

Ex post tracking errors of long-short portfolios								
	$T = 504$				$T = 252$			
	S	L	NL	DCC-NL	S	L	NL	DCC-NL
	Value-weighted benchmark							
$\tilde{N} = 100$	1.99	1.91***	1.92***	**1.90***	2.21	1.92***	1.92***	**1.92*****
$\tilde{N} = 500$	0.67	0.34***	0.32***	**0.31*****	1.22	0.34***	0.33***	**0.33*****
$\tilde{N} = 800$	1.08	0.13***	0.10***	**0.10*****	1.71	0.12***	**0.10*****	0.11***
	Equally-weighted benchmark							
$\tilde{N} = 100$	5.75	5.50***	5.50***	**5.48*****	6.45	5.54***	5.53***	**5.53*****
$\tilde{N} = 500$	3.35	2.10***	2.00***	**1.95*****	2.59	2.14***	2.07***	**2.03*****
$\tilde{N} = 800$	1.11	0.91***	0.84***	**0.82*****	0.96	0.90***	0.86***	**0.85*****
	Markowitz benchmark							
$\tilde{N} = 100$	10.53	10.07***	10.03***	**9.91*****	12.21	10.70***	10.65***	**10.56*****
$\tilde{N} = 500$	7.50	6.21***	5.88***	**5.86*****	8.77	6.79***	6.54***	**6.44*****
$\tilde{N} = 800$	6.06	3.73***	3.43***	**3.38*****	6.70	3.96***	3.77***	**3.72*****

Exhibit 7. Annualized ex post tracking-errors of minimum-tracking-error portfolios with two and one years of estimation-window length, in percent. All numbers are based on 10,059 daily out-of-sample excess returns from 07/02/1982 until 12/30/2022. For any row, the lowest (and thus best) number appears in **bold face** and significant outperformance over S is denoted by asterisks: *** denotes significance at the 0.01 level; ** denotes significance at the 0.05 level; and * denotes significance at the 0.1 level.

Unknown Benchmark Weights

The problem formulation of passive minimum-tracking-error portfolios with unknown benchmark weights is described above in (5–7). Arguably, this formulation is less relevant in practice compared to the case of known benchmark weights. Therefore, we relegate the corresponding results to Appendix B to save space.

Active Portfolios

Remember that for the active portfolio strategies we assume that every stock is investable, so that $\tilde{N} = N$ always, and that the tracking error is only a 'component' of the investor's objective now.

Tracking Error as Part of the Objective Function

The problem formulation is described in (8–10) with *active* standing for the negative of the variance of the active portfolio. As stated before, in this application then, the manager seeks to minimize a convex combination of the variance of the portfolio and the variance of the tracking error, subject to being fully invested and (if desired) to being long only. Therefore, to make the results easier to digest and to put them on the same 'scale' as the tracking-error results above, Exhibit 8 reports the square root of the ex post convex combination of the two variances, which corresponds to taking the square root of the *negative* of the objective function (8):

$$\sqrt{\delta \cdot w'\Sigma_{r,t}w + (1-\delta) \cdot (w - w_{\mathrm{BM},t})'\Sigma_{r,t}(w - w_{\mathrm{BM},t})} \tag{15}$$

In this way, smaller numbers are now also better.

It can be seen that the best results are uniformly achieved by DCC-NL. The overall ranking, from best to worst, is DCC-NL, NL, L, and S. All shrinkage estimators benefit from the ability to go short (that is, long-short vs. long-only), whereas going short is mostly harmful for the sample covariance matrix due to its larger estimation error, especially in larger dimensions. Note that for long-short portfolios the outperformance of shrinkage estimators over the sample covariance matrix is remarkable. However, for long-only, only shrinkage estimators in conjunction with DCC can consistently and markedly outperform the sample covariance matrix.

For $\delta = 100\%$, which corresponds to the global minimum variance (GMV) portfolio without any tracking-error consideration, the out-of-sample standard deviation is statistically significant lower for shrinkage estimators, at least consistently for long-short GMV portfolios.[11]

[11] Arguably, portfolios without tracking-error consideration are not of leading interest in this paper, but the result is certainly noteworthy in passing.

Ex post tracking errors								
	Long-short				Long-only			
	S	L	NL	DCC-NL	S	L	NL	DCC-NL
	$\delta = 100\%$							
$\tilde{N} = 100$	12.93	12.90	12.78***	**12.66*** **	13.54	13.62	13.52**	**13.18*** **
$\tilde{N} = 500$	10.47	9.69***	9.53***	**9.11*** **	11.53	11.50***	11.50***	**9.84*** **
$\tilde{N} = 800$	12.94	8.59***	8.40***	**7.26*** **	10.51	10.52	10.61	**7.89*** **
	$\delta = 75\%$							
$\tilde{N} = 100$	12.93	12.90	12.78	**12.66**	12.65	12.71	12.64	**12.44**
$\tilde{N} = 500$	11.00	10.59	10.51	**10.30**	11.21	11.20	11.20	**10.41**
$\tilde{N} = 800$	12.37	10.02	9.92	**9.41**	10.63	10.63	10.67	**9.40**
	$\delta = 50\%$							
$\tilde{N} = 100$	7.08	6.56	6.94	**6.65**	11.18	11.20	11.16	**11.09**
$\tilde{N} = 500$	10.33	10.14	10.10	**10.00**	10.30	10.30	10.30	**10.00**
$\tilde{N} = 800$	10.98	9.85	9.81	**9.58**	9.99	9.99	10.02	**9.54**
	$\delta = 25\%$							
$\tilde{N} = 100$	3.84	3.54	3.29	**3.47**	8.52	8.52	8.51	**8.49**
$\tilde{N} = 500$	8.15	8.09	8.08	**8.05**	8.12	8.11	8.11	**8.04**
$\tilde{N} = 800$	8.34	7.98	7.96	**7.89**	8.00	8.00	8.00	**7.89**

Exhibit 8. Realizations of objective function (15). All numbers are based on 10,059 daily out-of-sample returns from 07/02/1982 until 12/30/2022. For any row, the lowest (and thus best) number appears in **bold face**. For $\delta = 100\%$, significant outperformance over S is denoted by asterisks.

Exhibit 9 presents violin plots of monthly realizations of the objective function for $\delta \in \{0.75, 0.5, 0.25\}$. It can be seen that for any value of δ using DCC-NL yields a distribution that lies 'below' the corresponding distribution when using the sample covariance matrix. Therefore, updating from the sample covariance matrix to a sophisticated shrinkage estimator holds also remarkable overall improvement for active managers that control for benchmark deviations.

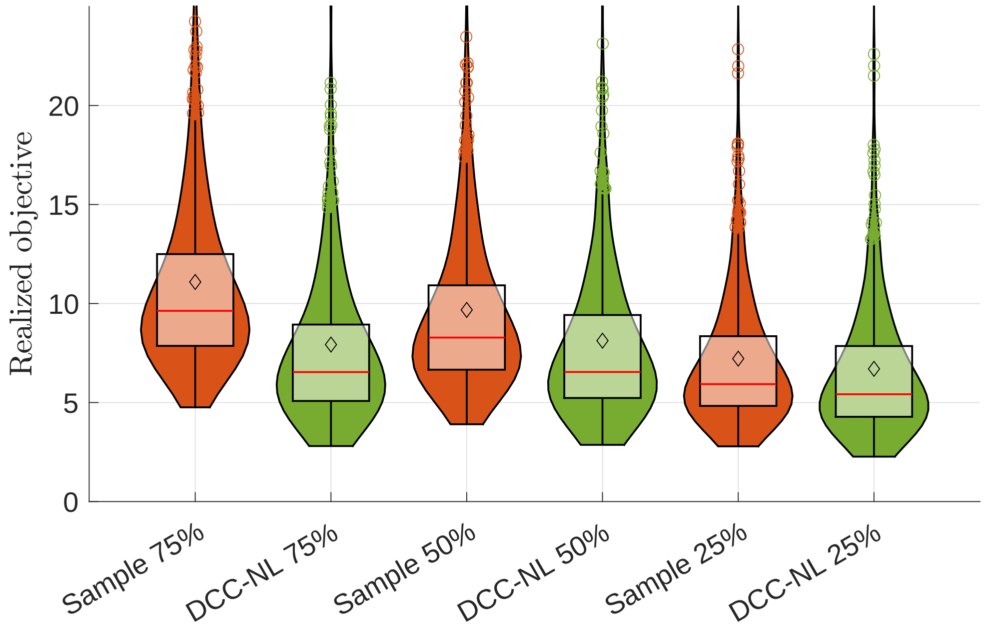


Exhibit 9. Monthly realizations of objective function (15) for long-short portfolios, in percent. The number after the name "Sample" or "DCC-NL" indicates the value of the δ parameter in the objective function.

Tracking-Error Constraint

The problem formulation is described in (11–14) with *active* standing for the negative of the variance of the active portfolio. As stated before, in this application then, the manager seeks to minimize the variance of the portfolio, subject to an upper bound τ of the standard deviation / variance of the tracking error, as well as being fully invested and (if desired) to being long only. In this section we show empirical results for GMVs without, with 10%, with 5%, and with 1% (variance of the) tracking error constraints.

Note that this important practice example is more difficult to evaluate, as we cannot just look at the out-of-sample objective number, being SD. We need also to take into account if, or how many times, the tracking error constraint (13) is actually fulfilled out-of-sample. For fair comparison, we report in Exhibit 10 not only the out-of-sample standard deviation (SD) numbers, but also the SD for all days where the constraint is fulfilled across all competitors (SD*), the ex post tracking error (TE), the mean absolute difference between the monthly ex ante vs. ex post tracking errors (MAD), and the success rate of the one-month ex post tracking error vs. its constraint; all being defined above.

Long-short GMV portfolio with TE constraint												
	$N = 100$				$N = 500$				$N = 1000$			
	S	L	NL	DCC-NL	S	L	NL	DCC-NL	S	L	NL	DCC-NL
	GMV											
SD	12.93	12.90	12.78	**12.66**	10.47	9.69	9.53	**9.11**	12.94	8.59	8.40	**7.26**
TE	14.13	**13.09**	13.83	13.26	16.16	15.06	14.94	**13.75**	19.13	15.58	15.05	**14.13**
	GMV with 10% TE constraints											
SD	13.41	13.42	**13.35**	13.85	11.34	**11.15**	11.35	12.39	11.01	**10.84**	10.95	12.03
SD*	9.96	9.93	9.91	**9.70**	8.31	8.23	8.33	**8.15**	7.30	7.96	8.07	**7.85**
TE	9.88	9.89	9.78	**9.07**	9.70	10.29	9.75	**8.69**	9.56	9.98	9.64	**8.77**
MAD	3.71	3.71	3.76	**2.50**	3.81	3.77	3.85	**2.85**	3.85	3.73	3.84	**2.98**
Success	71.74	70.97	73.16	**76.12**	74.45	71.14	74.75	**80.29**	75.18	72.32	75.43	**79.40**
	GMV with 5% TE constraint											
SD	15.15	**15.03**	15.17	15.50	14.19	**13.91**	14.11	14.60	13.86	**13.75**	13.87	14.16
SD*	10.72	10.64	10.71	**10.45**	9.74	9.61	9.71	**9.55**	9.51	9.29	9.37	**9.20**
TE	5.06	5.27	5.01	**4.63**	4.92	5.25	5.00	**4.58**	5.23	5.18	4.97	**4.68**
MAD	1.88	1.85	1.89	**1.22**	1.89	1.87	1.91	**1.43**	1.86	1.85	1.88	**1.49**
Success	70.11	66.86	70.69	**74.75**	73.06	69.00	72.86	**76.90**	67.94	69.03	72.45	**75.24**
	GMV with 1% TE constraint											
SD	17.52	**17.49**	17.52	17.60	17.03	**16.98**	17.02	17.11	16.89	**16.85**	16.91	16.94
SD*	12.02	12.00	12.02	**11.94**	10.67	10.63	10.66	**10.58**	10.96	10.91	10.95	**10.84**
TE	1.02	1.06	1.01	**0.93**	1.03	1.06	0.99	**0.91**	1.17	1.06	0.99	**0.95**
MAD	0.38	0.37	0.38	**0.25**	0.36	0.37	0.39	**0.29**	0.35	0.37	0.39	**0.30**
Success	70.09	66.77	70.84	**73.71**	70.36	68.56	73.33	**77.11**	58.01	68.39	73.24	**74.14**

Exhibit 10. Performance measures for various estimators of the long-short global minimum variance portfolio with tracking-error constraints, in percent. All numbers are based on 10,059 daily out-of-sample returns from 07/02/1982 until 12/30/2022. For any row, the best performer appears in **bold face**.

Exhibit 10 presents the results for the long-short portfolios which can be summarized as follows.

- In the absence of a tracking-error constraint, all shrinkage estimators consistently and markedly outperform the sample covariance matrix estimation in terms of SD and TE. With the single exception of $N = 100$, DCC-NL has the lowest SD and TE numbers.
- Across all TE constraints and performance measures, with the single exception of SD, DCC-NL is the clear winner.
- In terms of SD, L performs the best. However, this comes as no surprise as the success rate of L is also among the lowest.
- On the contrary, DCC-NL has the highest success rates ranging from 73.7% to 80.3%. If we control for the fulfilment of the TE constraint, DCC-NL has the lowest out-of-sample standard deviation.
- Additionally, DCC-NL has the lowest ex post tracking error numbers and smallest differences of ex ante vs. ex post tracking error numbers. Finally, only DCC-NL has uniformly lower TE numbers compared to its TE constraint across all scenarios.

To visualize the results, the upper panel of Exhibit 11 plots the time series of monthly ex post tracking errors for S and DCC-NL. It can be seen that the DCC-NL series are consistently closer to, and also more frequently below, the TE constraint threshold. Noteworthy is also the robustness and accuracy of the DCC-NL ex post TE series, by not overshooting too much during periods of financial turmoil, but also not undershooting too much during calmer periods. Therefore, the outperformance of DCC-NL over S is not restricted to certain periods but holds throughout. The lower panel of Exhibit 11 compares the distributions of the deviations of the monthly ex post tracking errors (TE constraint) by lining up the corresponding violin plots. The benefit of DCC-NL relative to S can clearly be seen, since its distributions are more concentrated around zero, as well as the larger mass lies below zero, as that of S.

The long-only results are qualitatively similar and therefore not reported here to save space, but they are available upon request. In certain scenarios Quadratic Inverse Shrinkage (QIS, Ledoit and Wolf (2022b)) or the linear shrinkage formula of De Nard (2022), which are both static estimators, can be the winners for some performance measures (by only a small margin though).

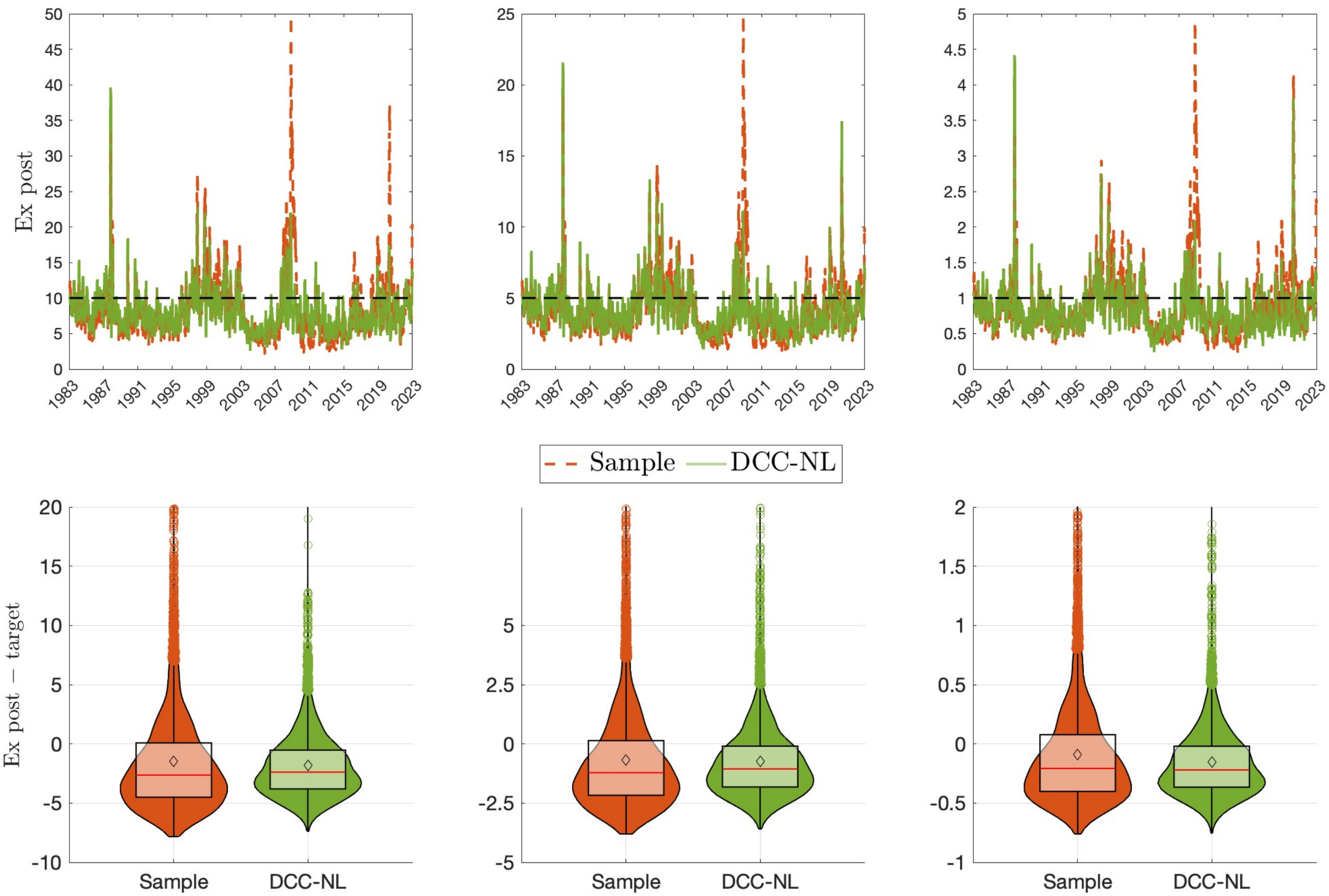


Exhibit 11. Rolling-window monthly ex post tracking-error numbers (upper panel) and ex ante vs. ex post tracking-error differences (lower panel) for the $N = 500$ long-short minimum variance portfolio with various ex ante tracking-error constraints, in percent. The panels on the left are with 10%, the panels in the middle are with 5%, and the panels on the right are with 1% tracking-error constraint.

In sum, upgrading from the sample covariance matrix to a shrinkage estimator of the covariance matrix is clearly beneficial also in the case of benchmark-following or fund-mimicking portfolios with unknown benchmark weights. As expected, the TE numbers are (much) larger for unknown benchmark weights compared to known benchmark weights, since the benchmarks now now based on the $M = 200$ smallest (as well as ineligible) stocks only, so that in each scenario the eligible universe and the benchmark universe are disjoint.

Consequently, DCC-NL consistently and markedly outperforms the sample covariance matrix, which allows active managers to implement their strategies more effectively while controlling tracking error. For example, benchmarked managers often introduce a buffer on their ex ante tracking-error estimation to control for in-sample optimism and prediction errors, especially when the TE constraints have to be fulfilled / are (legally) binding. Thanks to the improved accuracy of DCC-NL, potential buffers on TE constraints can be reduced or even removed: Reductions, or even removals, of such buffers allow active managers to be closer to their investment strategies of interest.

Conclusion

In this paper we have studied the benefit of using improved estimators of the covariance matrix in the context of tracking-error management by putting ourselves in the (bespoke) shoes of a portfolio manager who is tied to a benchmark. Whereas such managers are typically ignored in the academic literature, they form the majority in real life.

The passive managers among them face the seemingly trivial task of tracking an index, such as the S&P 500, Russell 2000 or the MSCI World. In practice, the task is not trivial because an index typically contains many stocks and some of them are difficult or expensive to trade; therefore, the task is to track the index as close as possible using a subset of its constituents only. A more challenging case is to mimic a benchmark or fund with unknown weights, or even unknown or ineligible constituents.

The active managers among them develop their own strategies designed to have attractive risk-return properties but are not allowed / willing to deviate too much from a specified benchmark.

All of them, therefore, need an estimator of the covariance matrix of many (excess) returns at the portfolio-selection stage. We have demonstrated the benefit of shrinkage estimators to this end and can safely single out the dynamic DCC-NL model of Engle et al. (2019) as the most convincing all-around performer. At any rate, the obvious message is that portfolio managers who aim to minimize respectively control tracking error (that is, passive respectively active managers tied to a benchmark) need to abandon the sample covariance matrix to avoid eventually facing questions on whether they have been legally derelict in their fiduciary duty to adhere to best-practice for their investors.

References

Basak, G. K., Jagannathan, R., and Ma, T. (2009). Jackknife estimator for tracking error variance of optimal portfolios. *Management Science*, 55(6):990–1002.

Belhaj, R., Maillard, D., and Portait, R. (2005). Portfolio management under constraints on ex-ante and ex-post tracking-errors. *Bankers, Markets & Investors*, (79).

De Nard, G. (2022). Oops! I shrunk the sample covariance matrix again: Blockbuster meets shrinkage. *Journal of Financial Econometrics*, 20(4):569–611.

De Nard, G., Engle, R. F., Ledoit, O., and Wolf, M. (2022). Large dynamic covariance matrices: Enhancements based on intraday data. *Journal of Banking and Finance*, 138:106426.

De Nard, G., Ledoit, O., and Wolf, M. (2021). Factor models for portfolio selection in large dimensions: The good, the better and the ugly. *Journal of Financial Econometrics*, 19(2):236–257.

Engle, R. F., Ledoit, O., and Wolf, M. (2019). Large dynamic covariance matrices. *Journal of Business & Economic Statistics*, 37(2):363–375.

Jegadeesh, N. and Titman, S. (1993). Returns to buying winners and selling losers: Implications for stock market efficiency. *Journal of Finance*, 48(1):65–91.

Jorion, P. (2003). Portfolio optimization with constraints on tracking error. *Financial Analysts Journal*, 59(5):70–82.

Ledoit, O. and Wolf, M. (2004). Honey, I shrunk the sample covariance matrix. *Journal of Portfolio Management*, 30(4):110–119.

Ledoit, O. and Wolf, M. (2011). Robust performance hypothesis testing with the variance. *Wilmott Magazine*, September:86–89.

Ledoit, O. and Wolf, M. (2022a). The power of (non-)linear shrinking: A review and guide to covariance matrix estimation. *Journal of Financial Econometrics*, 20(1):187–218.

Ledoit, O. and Wolf, M. (2022b). Quadratic shrinkage for large covariance matrices. *Bernoulli*, 28(3):1519–1547.

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

Pastor, L., Stambaugh, R. F., and Taylor, L. A. (2024). Green tilts. NBER Working Paper 31320, National Bureau of Economic Research.

Sharpe, W. F. (1964). Capital asset prices: A theory of market equilibrium under conditions of risk. *Journal of Finance*, 19(3):425–442.

Zenti, R. and Pallotta, M. (2002). Are asset managers properly using tracking error estimates? *Journal of Asset Management*, 3(3):279–289.

Zhao, Z., Ledoit, O., and Jiang, H. (2023). Risk reduction and efficiency increase in large portfolios: Leverage and shrinkage. *Journal of Financial Econometrics*, 21(1):73–105.

A Markowitz Portfolio with Momentum Signal

We now turn attention to a 'full' Markowitz portfolio with a signal.

By now researchers have established a large number of variables that can be used to construct a return-predictive signal. For simplicity and reproducibility, we use the well-known momentum factor (or just momentum for short) of Jegadeesh and Titman (1993). For a given investment period k and a given stock, the momentum is the geometric average of the previous 252 returns on the stock but excluding the most recent 21 returns; in other words, one uses the geometric average over the previous 'year' but excluding the previous 'month'. Collecting the individual momentums of all the N, respectively M, stocks contained in the entire, respectively ineligible, universe yields the return-predictive signal, denoted by m_t.

Given the estimator of the covariance matrix of stock returns $\hat{\Sigma}_{r,t}$ and the gross-exposure parameter γ, the Markowitz mean-variance efficient portfolio based on a return predictive signal m_t is formulated as:

$$\min_{w} w'\Sigma_{r,t}w \tag{16}$$

$$\text{subject to} \quad w'm_t = b_t \ , \tag{17}$$

$$w'\mathbb{1} = 1 \ , \ \text{and} \tag{18}$$

$$|w|'\mathbb{1} \leq \gamma, \tag{19}$$

where b_t is a selected target expected return and $|\cdot|$ returns the absolute value of all vector elements.

To be consistent among the discussed competitors, we use always the sample covariance matrix for $\hat{\Sigma}_{r,t}$ to compute the benchmark portfolio, but note that DCC-NL outperforms also in this setting; see De Nard et al. (2022).

For the empirical analysis in this paper we set the target expected return b_t equal to the expected return of the equally-weighted portfolio among the top-quintile stocks (according to their momentums) and the gross-exposure parameter γ equal to 1.6. Therefore, we focus on the so-called "130/30" long-short portfolio.

For more details about mean-variance portfolios with leverage constraint and the risk reduction as well as efficiency increase in large portfolios due to shrinkage estimators see Zhao et al. (2023).

B Passive Portfolios: Unknown Benchmark Weights

Exhibit 12 presents the results on ex post tracking error which can be summarized as follows:

- For any scenario, as expected, TE decreases as the size of the eligible universe, $\tilde{N}$, increases: A larger universe of eligible stocks makes it easier to track the benchmark.
- Being able to short stocks is not necessarily beneficial. In particular, the results for the sample covariance matrix are uniformly worse for long-short compared to long-only, and sometimes by a pronounced margin. This is due to the larger estimation error of S as well as worse turnover and leverage numbers compared to the shrinkage estimators. On the other hand, the shrinkage estimators enjoy a 'robustness' property in this regard, since for any given scenario the difference is generally small. On balance, NL is best in the long-short case whereas DCC-NL is best in the long-only case.
- TE is generally best (that is, lowest) for the the value-weighted benchmark, followed by the equally-weighted benchmark, followed by the Markowitz benchmark.
- With the single exception of $\tilde{N} = 100$, all shrinkage estimators outperform the sample covariance matrix. The outperformance is statistically significant and economically meaningful.
- As expected, the results are worse for unknown benchmark weights compared to known benchmark weights, though it should be kept in mind that the benchmarks now are based on the $M = 200$ smallest stocks only so that in each scenario the eligible universe and the benchmark universe are disjoint.

In sum, upgrading from the sample covariance matrix to a shrinkage estimator of the covariance matrix is clearly beneficial in terms of reducing ex post tracking error also when the benchmark weights are unknown, with the best choice being NL in the long-short case and DCC-NL in the long-only case.

The remainder of this appendix re-runs the remaining previous analyses for known benchmark weights but now for the case of unknown benchmark weights. Since detailed results can be found in the exhibits below, we restrict ourselves to a brief executive summary in words.

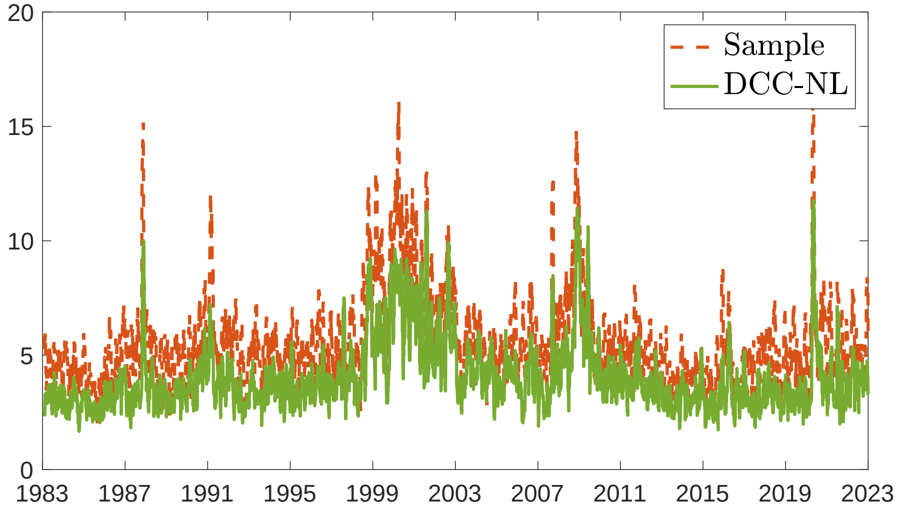
Ex post tracking errors									
	Long-short				Long-only				
	S	L	NL	DCC-NL	Naïve	S	L	NL	DCC-NL
	Value-weighted benchmark								
$\tilde{N} = 100$	9.11	9.01***	**9.01*** **	9.29	11.61	9.00	9.00	**8.98*** **	9.02
$\tilde{N} = 500$	6.46	5.77***	**5.62*** **	5.82***	9.98	5.71	5.67***	5.65***	**5.60*** **
$\tilde{N} = 800$	5.95	4.39***	**4.20*** **	4.33***	9.53	4.29	4.28**	4.25***	**4.18*** **
	Equally-weighted benchmark								
$\tilde{N} = 100$	9.14	9.05***	**9.05*** **	9.33	10.56	9.03	9.03	**9.01*** **	9.06
$\tilde{N} = 500$	6.49	5.80***	**5.66*** **	5.86***	7.60	5.74	5.71***	5.68***	**5.64*** **
$\tilde{N} = 800$	5.97	4.41***	**4.22*** **	4.35***	6.09	4.32	4.30***	4.27***	**4.21*** **
	Markowitz benchmark								
$\tilde{N} = 100$	13.03	**12.87*** **	12.89***	12.97	15.49	12.93	12.91	12.89***	**12.63*** **
$\tilde{N} = 500$	11.88	10.73***	**10.42*** **	10.75***	14.56	10.57	10.52***	10.51***	**10.45*** **
$\tilde{N} = 800$	13.79	10.31***	**9.72*** **	10.03***	14.37	9.86	9.83***	9.82***	**9.77*** **

Exhibit 12. Annualized ex post tracking-error numbers for minimum-tracking-error portfolios, in percent. All numbers are based on 10,059 daily out-of-sample excess returns from 07/02/1982 until 12/30/2022. For any row, the lowest (and thus best) number appears in **bold face** and significant outperformance over S is denoted by asterisks: *** denotes significance at the 0.01 level; ** denotes significance at the 0.05 level; and * denotes significance at the 0.1 level.

The upper panel of Exhibit 13 demonstrates (again) that DCC-NL yields monthly ex post tracking errors that are 'uniformly' below those based on the sample covariance matrix over time; the lower panel of the exhibit demonstrate (again) that all shrinkage estimators yield monthly ex post tracking errors whose distribution lies well below that of the monthly ex post tracking errors based on the sample covariance matrix. These results are (again) only for the long-short portfolios based on $\tilde{N} = 800$ eligible stocks tracking the value-weighted benchmark.

Exhibit 13 shows the 'optimism' and 'accuracy' of monthly ex ante tracking errors by comparing them to ex post tracking errors. First, as the eligible universe and the benchmark universe are disjoint, not only the TE numbers are higher, but also the optimism problem is larger and the accuracy is lower for the unknown benchmark weights setting. Second, again it is seen that, on balance, using the sample covariance matrix leads to the most optimistic and least accurate results, although now it is L instead of DCC-NL that yields the overall best results (according to all metrics). Note that also the turnover and gross leverage numbers for long-short portfolios are qualitatively similar to those of Exhibit 6 for known benchmark weights.

As in the case of known benchmark weights seen previously, all the results are robust to shorter estimation window lengths.



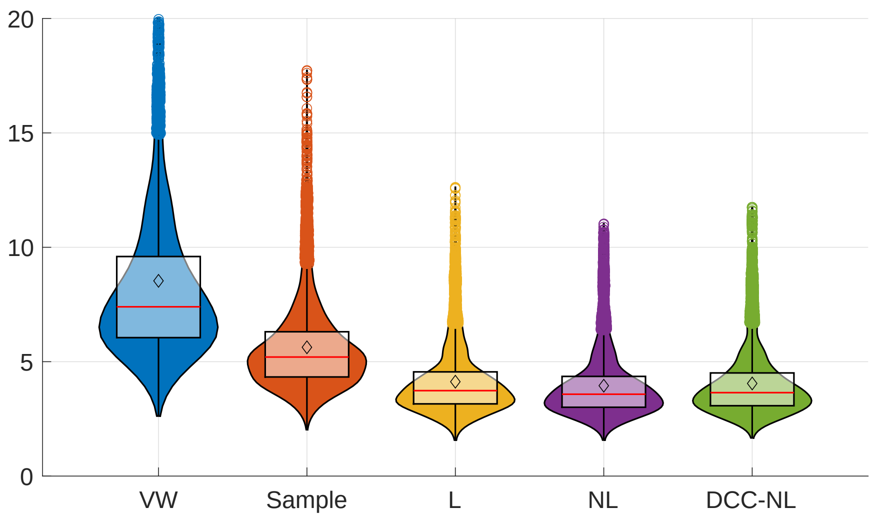


Exhibit 13. Monthly ex post tracking-error numbers (upper panel) and their distribution (lower panel) for long-short portfolios ($\tilde{N} = 800$) tracking the value-weighted benchmark, in percent.

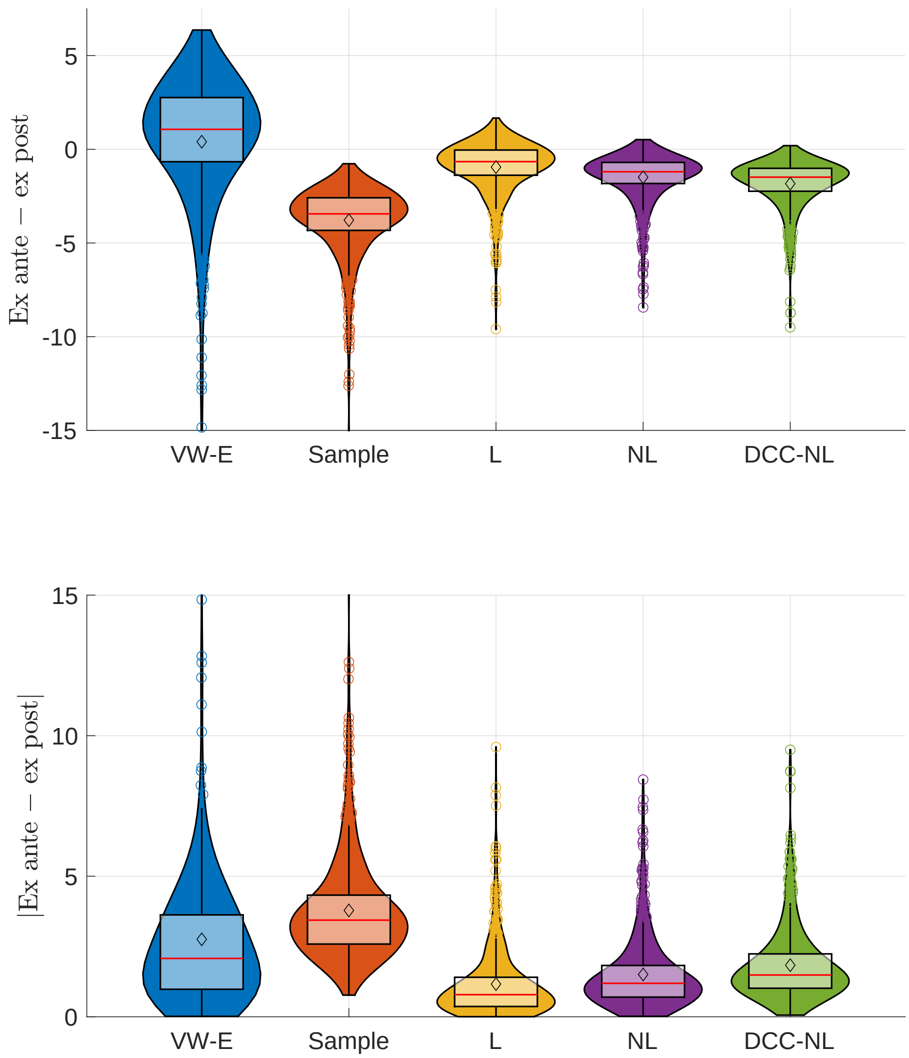


Exhibit 14. Violin plots of monthly ex ante vs. ex post tracking-errors for long-short minimum-tracking-error portfolios ($\tilde{N} = 800$) tracking the value-weighted benchmark, in percent.