Stock-level index tracking: a point-in-time benchmark, a long-only mandate and its attribution by sector

Carlo Hofer. November 2026. Code, notebooks and the tables every figure is drawn from: github.com/ochofer/stock-level-tracking. The dashboard: carlohofer.com/stock-level-tracking/dashboard/.

Summary

- Under the mandate the four covariance estimators realise 78.9 to 83.6 basis points a year of tracking error, and each one's forecast understates it. The bias statistic, realised over forecast, is 1.48 to 1.85, above the band of 0.8 to 1.25 fixed before the run.
- Index changes force 2.99 per cent a year of the Ledoit-Wolf path's 26.26 per cent of one-way turnover, or 11.4 per cent. On the Ledoit-Wolf estimator, loosening the cap on chosen turnover from 0.5 per cent a month to none raises it from 6.3 to 59.4 per cent a year, while the realised tracking error stays between 76.3 and 79.4 basis points a year. On the factor estimator chosen turnover rises from 6.2 to 14.9 per cent a year and the tracking error falls from 83.3 to 79.6.
- Month by month, allocation is the largest of the three Brinson-Fachler terms on the Ledoit-Wolf, factor and principal-component estimators: 16.4 to 16.7 basis points a month in absolute value on average, against 10.6 to 13.8 for selection and 1.5 to 2.2 for interaction. Over the 306 months the active return is 0.2 to 13.7 basis points a year before costs, within one Newey-West standard error of zero on every path, as is each of its three terms, and the six factors explain 2 to 13 per cent of its variance.
- The benchmark rebuilt from CRSP's list of index members reproduces CRSP's own value-weighted return of the same list with a tracking error of 11.5 basis points a year from 1926 to 2024. Its momentum and variance spreads have correlations of 0.888 and 0.954 with Kenneth R. French's large-cap spreads.

Terms used in this note, each with the one meaning it carries throughout.

universe: the point-in-time S&P 500, the stocks on CRSP's list of index members on the last trading day of the month before. **market capitalisation**: price times shares outstanding, at the end of the month before. **cap-weighted**: weighted by market capitalisation. CRSP's value-weighted return of the list is the cap-weighted return as CRSP computes it. **value-weighted**: cap-weighted. **benchmark**: the cap-weighted universe, each stock's weight its market capitalisation over the universe's total. **active weight**: a weight in the portfolio minus the same weight in the benchmark. **active return**: the portfolio's return minus the benchmark's, each over the month's names with a return. **basis point**: a hundredth of a percentage point. **tracking error**: the standard deviation of the monthly active return, times the square root of 12. The forecast is the tracking error the month's covariance matrix implies for the month's active weights, and the realised is the one measured afterwards. **excess return**: a return minus the one-month Treasury bill rate. **covariance matrix**: the table of the stocks' variances and of their covariances with each other. **eigenvalue**: the variance of one of the uncorrelated portfolios a covariance matrix splits into. The condition number is the largest eigenvalue over the smallest. **singular**: said of a matrix whose smallest eigenvalue is zero. **covariance estimator**: a way of estimating the covariance matrix from three years of daily returns: the sample covariance, the Ledoit-Wolf estimator, which shrinks the sample covariance towards a scaled identity matrix, a factor model on French's five factors and momentum, and a principal-component model with the benchmark as its first factor. The last three are the structured estimators. **band**: the range of the bias statistic within which a forecast counts as right, 0.8 to 1.25, fixed before the run. **shrinkage intensity**: the weight, from zero to one, that the Ledoit-Wolf estimator puts on the scaled identity matrix, set by the data. **estimation error**: the difference between an estimate and the quantity it estimates, here between a covariance matrix estimated from three years of returns and the one that holds in the month. **bias statistic**: the standard deviation of each month's active return divided by that month's forecast, over the months with a forecast above one millionth. It is one when the forecasts are right on average. **mandate**: the full set of constraints the portfolio obeys, constraint set C3 below. **tilt**: every Energy and Utilities stock capped at half its benchmark weight. **beta**: the slope of a return on the benchmark's return. **path**: one covariance estimator under one constraint set at one turnover cap, run month by month from the benchmark. **turnover**: a month's one-way trading as a share of the portfolio. Forced turnover is what index changes require, and chosen turnover the rest. Reweighting is the turnover that following the benchmark's own changes in share counts and its dividends takes, rebalancing the optimiser's own moves, and the overlap the trades of one step that the next undoes. **turnover cap**: the limit on chosen turnover, 2 per cent of the portfolio a month in the mandate. **Newey-West standard error**: the standard error of a mean that allows each month to be correlated with the next twelve. A plain standard error makes no such allowance. **autocorrelation**: the correlation of a series with its own past values. **attribution**: a split of the active return into parts that sum to it. **allocation**: the part from holding a sector at another weight than the benchmark's, the sector's active weight times the benchmark sector's return less the benchmark's. **selection**: the part from holding other stocks within a sector, the sector's benchmark weight times the portfolio sector's return less the benchmark sector's. **interaction**: the sector's active weight times that same difference. **Unclassified**: the twelfth sector, the members with no GICS sector, one to three securities a month. **factor**: a common source of return: the market, size, value, profitability and investment factors of Fama and French (2015), and momentum. **loading**: a stock's sensitivity to a factor, from a regression of its daily returns on the six factors over the three years before the month. **active exposure**: the members' loadings on a factor weighted by the active weights. A factor's contribution is its active exposure times its return, the residual is the active return less the contributions, and the factor share is one minus the residual's variance over the active return's. **look-ahead**: the use, in a decision dated at one time, of data published only later, named in the Other projects below. **Sharpe ratio**: a portfolio's mean excess return over its standard deviation, named in the Other projects below. **gate**: a test of the data fixed in advance, which stops the study if it is not met.

expectation: a test of a result fixed in advance, reported as a finding whether it is met or not. An expectation named for two or more estimators has one verdict, met only if it holds for each.

1. Introduction

An index-tracking mandate is judged by how far its portfolio strays from a cap-weighted benchmark, and the manager who runs it states that distance in advance as a forecast tracking error, computed from a covariance matrix of the benchmark's stocks. Yet the benchmark changes its members every month, a covariance matrix of about 500 stocks estimated from about 755 days of returns carries estimation error (Ledoit and Wolf, 2004) that an optimiser exploits (DeMiguel et al., 2009), and the client's constraints, bounds on single names and sectors, a cap on turnover, exclusions and tilts, decide how much the optimiser can do with the forecast. The question this note asks is therefore a practical one: under such a mandate, how close does the portfolio stay to the benchmark, how well does each of four common covariance estimators forecast the distance it realises, what is the turnover made of, and where does the active return come from by sector.

The closest evidence comes from a companion study that ran a mandate of the same design on Kenneth R. French's 49 industry portfolios and found that five covariance estimators understate their own tracking error, by 22 to 61 per cent (Hofer, 2026). But an industry portfolio does not behave like a stock in the ways a mandate depends on. Industries do not enter or leave the benchmark, are never delisted, carry no single-name concentration and sit in no GICS sector, so the industry evidence leaves open whether the understatement, and the structure of the active return, carry over to the level at which a mandate is run. At the stock level the benchmark itself has to be built: a list of members as it stood at each date, their capitalisations at the start of each month, their returns with the returns of delisted stocks applied (Shumway, 1997), and their sectors as they were classified at the time.

This study builds that benchmark and runs the mandate on it. I take the members of the S&P 500 from CRSP's list, month by month from January 1926 to December 2024, weight each by its market capitalisation at the start of the month, and check the result three ways: against CRSP's own value-weighted return of the same list, against CRSP's count of members and their total value, and against French's large-cap portfolios sorted on momentum and on variance. From July 1999, when Compustat's GICS history begins, a long-only mandate minimises the forecast tracking error each month under four covariance estimators, with bounds per stock, company and sector, a cap on chosen turnover, the tilt, two excluded GICS sub-industries and a beta of one. Twenty-four paths run over the 306 months to December 2024. The turnover of each is split into the part that index changes force, the part that follows the benchmark's own reweighting and the part the optimiser chooses, and the active return of the four paths at a 2 per cent cap is attributed by sector with the method of Brinson and Fachler (1985) and by the five factors of Fama and French (2015) and momentum. Four gates on the data and nine expectations on the results were written into the code's constants, each dated before the run that judges it. The design measures the mandate's tracking error, its forecast and its parts, and only that: it does not estimate a return the mandate earns, since a portfolio built to minimise its tracking error is not built to earn one.

The S&P 500 between 1999 and 2024 is an informative setting for this question for three reasons. First, CRSP's list dates each membership by day, so the benchmark can be rebuilt with nothing unknown when each month begins, and CRSP's own file of the index supplies an answer key against which the rebuilt return can be checked. Second, the window holds the fall of 2000 to 2002, the financial crisis of 2008 and the shock of 2020, months in which volatilities and correlations moved far from their values over the three years before, which is where a forecast drawn from those three years is most likely to miss. Third, from 2014 the list holds companies with two share classes whose daily returns move almost together, a feature that tests a covariance estimator in a way that industry portfolios cannot.

Empirically, the rebuilt benchmark reproduces CRSP's value-weighted return of the same list with a tracking error of 11.5 basis points a year over 99 years, and the mandate's realised tracking error, 78.9 to 83.6 basis points a year on the four estimators, lies above every estimator's forecast. The bias statistics of 1.48 to 1.85 lie above the band of 0.8 to 1.25 fixed before the run, and the realised tracking error is highest in 2008, when the Ledoit-Wolf path realised 176 basis points against a mean forecast of 101. Index changes force 2.99 of the 26.26 per cent a year that the Ledoit-Wolf path turns over. By sector, allocation is the largest of the three terms month by month on each structured estimator, as

expected, while over the window the three terms are small beside their standard errors: on the Ledoit-Wolf path allocation and selection average 1.4 and 11.3 basis points a year, on Newey-West standard errors of 15.9 and 14.0. The estimates therefore indicate where the mandate's tracking error comes from without identifying a return it earns.

The note is structured as follows. Section 2 sets out the data layer, the covariance estimators, the mandate, the measures and the two attributions, with the expectations fixed for each. Section 3 reports the verdicts and the results in the order the design runs. Section 4 sets the results against the industry evidence and states the limits of the design.

2. Data and design

2.1 The universe and the benchmark

The universe in month *t* is the set of stocks on CRSP's list of index members on the last trading day of month *t* minus 1, so that nothing unknown when the month begins decides membership. Each member's monthly return comes from CRSP's monthly stock file. A delisting in the month is applied as Shumway (1997) proposes: CRSP's delisting return where it gives one, and where it does not, minus 30 per cent for a delisting for poor performance and zero for any other reason. A member's market capitalisation at the start of month *t* is its price times its shares outstanding at the end of month *t* minus 1, and the benchmark weights each member by it. A member's sector is the GICS sector in force on the last trading day of month *t* minus 1, found through CRSP's link to Compustat, and a member with no sector on that day is counted in a twelfth sector, Unclassified. The data are CRSP and Compustat via Wharton Research Data Services, accessed through Queen Mary University of London, from an extraction of 28 September 2026. CRSP ends on 31 December 2024 on this subscription, and so does the sample.

The index as S&P publishes it is not the benchmark, for two reasons that the data impose. Its total return is not in the extraction, and since 2005 the index weights each stock by the shares available to the public, its float, which cannot be rebuilt from CRSP's share counts. The benchmark is therefore the full-capitalisation universe of CRSP's list, and every tracking error in this note is measured against it.

2.2 The checks on the data

Three checks say whether the rebuilt benchmark is what it claims to be. First, its return, with dividends and price-only, is set beside CRSP's own value-weighted return of the same list. Second, its count of members and their total value are set beside CRSP's count and total value. Third, the universe is sorted each month into quintiles on momentum, the return compounded from month *t* minus 12 to month *t* minus 2, and on variance, the variance of daily returns over the 60 trading days before the month, both as French defines them, and the cap-weighted spread of the highest quintile less the lowest is set beside French's spread within his largest size quintile. French's quintile is cut at NYSE breakpoints and holds between about 100 and 530 names a month, so the two sets overlap without coinciding and a correlation near one is not expected.

Four gates and three expectations judge the data layer. The gates are that the benchmark's tracking error against CRSP's value-weighted return of the same list is at most 50 basis points a year (G1), that the usable window is at least 15 years (G2), that at least 90 per cent of the stocks removed from the universe from 1990 have a CRSP price within 10 trading days of their removal (G3), and that the market capitalisation is present in at least 90 per cent of membership months for at least 90 per cent of names (G4). A gate that is not met stops the study. The expectations are a tracking error of at most 20 basis points a year with a mean monthly difference within 2 basis points of zero, with dividends and price-only (E1), a count within 10 names of CRSP's in every month and a total value within 2 per cent in at least 95 per cent of months (E2), and, for each sort, a correlation of at least 0.80 with French's spread and a mean monthly difference within two standard errors of zero (E3).

2.3 The covariance estimators

For each month of the evaluation window, the covariance matrix of the members' daily excess returns over the three years before the month is estimated in four ways: the sample covariance, the Ledoit-Wolf estimator (Ledoit and Wolf, 2004), a factor model on French's five factors and momentum, and a principal-component model with the benchmark as its first factor and four further components. The four estimators are copied unchanged from the companion study's code, and a test checks each copied file's SHA-256. A member that lacks a return on more than five days of the window takes

a row from a model fitted on its own days, so that each matrix is positive semi-definite by construction. With about 490 members on about 755 days the sample matrix exists and is poorly conditioned, and the first clause of E4 expects that in every month its condition number is at least 10 times that of the Ledoit-Wolf matrix.

2.4 The mandate

The mandate holds the universe close to the benchmark under constraints a long-only client would set, and its constraints come in four cumulative sets. C0 is long-only. C1 adds a bound per stock, each within 50 basis points of its benchmark weight, the same bound on the sum of a company's share classes, and a band per GICS sector, each within 2 percentage points of its benchmark weight. C2 adds a cap of 2 per cent a month on chosen turnover. C3, the mandate, adds the tilt, a weight of zero in two GICS sub-industries, Tobacco and Aerospace and Defense, and a beta of one to the benchmark. In a set with the tilt the sector band applies to the other sectors only, because half of Energy's benchmark weight is above 2 percentage points in 268 of the 306 months, and a sub-industry is wider than a list of controversial weapons. Each month the optimiser chooses the weights that make the forecast tracking error as small as the constraints allow, starting from a reference portfolio: last month's portfolio after last month's returns, with the stocks that left the universe sold and the entrants bought at their benchmark weight. Forced turnover runs to the reference portfolio and chosen turnover from it, so the cap limits only what the optimiser chooses.

Twenty-four paths run month by month over the evaluation window, each starting as the benchmark: the four estimators under the four sets at the 2 per cent cap, and the Ledoit-Wolf and factor estimators under the mandate at caps of 0.5, 1 and 4 per cent and with no cap, which trace the frontier of tracking error against turnover. Costs are 10 basis points per unit of total one-way turnover, forced turnover included, with 20 as the sensitivity, and returns are gross of fees. The expectations on the mandate are that the sample estimator's realised tracking error is above the Ledoit-Wolf estimator's (E4, second clause), that for the Ledoit-Wolf and factor estimators the mean forecast does not rise as the cap loosens and the realised tracking error at the 4 per cent cap is below that at 0.5 (E5), that each structured estimator realises 50 to 150 basis points a year with a bias statistic between 0.8 and 1.25 (E6), and that forced turnover is less than a third of the Ledoit-Wolf path's total (E7).

2.5 The two attributions

The active return of the four paths at the 2 per cent cap is split each month in two ways. By sector, with the method of Brinson and Fachler (1985), allocation, selection and interaction are computed for each of the twelve sectors, and summed over the sectors the three equal the month's active return. Both sides are taken over the month's names with a return, each rescaled to one, as the committed active return is. A sector that the portfolio does not hold, none of its stocks with a return above a weight of one millionth, has a selection of zero and keeps the rest of its part as interaction, so that the identity holds exactly. The monthly terms are summed over months arithmetically, with no geometric linking, and reported as annual averages with Newey-West standard errors at 12 lags (Newey and West, 1987). By factor, the attribution of the companion study, copied unchanged, multiplies each month's active exposures to the six factors, from the factor estimator's loadings of the month, by French's monthly factor returns, and leaves a residual that closes the sum.

Two expectations judge the attributions. E8 expects the three terms to sum to the active return within 1e-9 in every month of the four paths, and, on each of the Ledoit-Wolf, factor and principal-component paths, the mean absolute monthly allocation to exceed the mean absolute monthly selection, with the interaction's the smallest of the three, each month's figure being the sum over the twelve sectors. The reason fixed before the run is that the mandate's active return comes from the tilt, an underweight of two sectors, while the stocks within sectors are chosen to keep the tracking error small. E9 expects contributions and residual to sum to the active return within 1e-9 in every month, and the mean active exposure to the market factor to lie within 0.05 of zero on each of the four paths, because the mandate holds its beta at one.

3. Results

3.1 The verdicts

Table 1 reports the four gates and the nine expectations with the figure each verdict rests on. The four gates are met, so the study runs on the full design. Of the expectations, E1, E2, E3, E7, E8 and E9 are met, E5 and E6 are not, and E4 is met in its second clause and not in its first. Each is reported below in the order the design runs, in the same words whether it is met or not.

Table 1. The four gates and nine expectations, fixed before each run, and what the run found

	Fixed before the run	*What the run found*	*Verdict*
G1	Tracking error of the benchmark against CRSP's own return of the same list, at most 50 bp a year	11.5 bp a year	Met
G2	Usable window at least 15 years	99 years	Met
G3	At least 90 per cent of removals with a CRSP price within 10 trading days	100 per cent of 836	Met
G4	Capitalisation in at least 90 per cent of membership months for at least 90 per cent of names	99.5 per cent of 1,912 names	Met
E1	That tracking error at most 20 bp a year and the mean difference within 2 bp, with dividends and price-only	11.5 and 11.6 bp a year, −0.18 and −0.16 bp	Met
E2	Count within 10 names of CRSP's in every month, value within 2 per cent in at least 95 per cent of months	1,188 and 1,186 of 1,188 months	Met
E3	Momentum and variance spreads: correlation at least 0.80 with French's, mean difference within two standard errors	0.888 and 0.954, 1.15 and 0.29 standard errors	Met
E4, first clause	Sample condition number at least 10 times the Ledoit-Wolf one in every month	86 of 306 months	Not met
E4, second clause	Under the mandate, the sample estimator's realised tracking error above the Ledoit-Wolf one	80.6 against 78.9 bp a year	Met
E5	On the frontier, the mean forecast does not rise as the cap loosens, and the realised tracking error is lower at 4 per cent than at 0.5	holds for the factor estimator, fails for Ledoit-Wolf (79.0 against 76.3 bp a year)	Not met
E6	Realised tracking error 50 to 150 bp a year and bias statistic 0.8 to 1.25, each structured estimator	tracking error 78.9 to 83.6, bias 1.48 to 1.85	Not met
E7	Forced turnover less than a third of the total, Ledoit-Wolf path	2.99 of 26.26 per cent a year, 11.4 per cent	Met
E8	The three terms sum to the active return, and allocation is the largest in absolute value and interaction the smallest, each structured estimator	largest gap 9.6e-17, allocation 16.4 to 16.7, selection 10.6 to 13.8, interaction 1.5 to 2.2 bp a month	Met
E9	Contributions and residual sum to the active return, and the mean exposure to the market within 0.05 of zero on every path	largest gap 8.7e-19, exposure −0.0006 to +0.0001	Met

Source: CRSP and Compustat via Wharton Research Data Services, Kenneth R. French Data Library and author's calculations. results/20260928/verdicts_milestone1.json to verdicts_milestone4.json.

3.2 The data layer

The rebuilt benchmark reproduces CRSP's own figures closely. The universe holds 1,936 securities over 2,064 membership spells, 440,724 member-months from January 1926 to December 2024, and its count is within 10 names of CRSP's own count on the same day in all 1,188 months, so the usable window is the whole 99 years (G2, E2). Its tracking error against CRSP's value-weighted return of the same list is 11.5 basis points a year with dividends and 11.6 price-only, with mean monthly differences of minus 0.18 and minus 0.16 basis points (G1, E1), and over the evaluation window the tracking error with dividends is 15.2. Its total value is within 2 per cent of CRSP's in 1,186 of the 1,188 months. All 836 removals from 2 January 1990 have a CRSP price within 10 trading days (G3), and 1,902 of 1,912 names have a market capitalisation in at least 90 per cent of their membership months (G4). Of the 799 delisting months of members, CRSP gives the delisting return in 798, and the minus 30 per cent of Shumway (1997) is used once.

Two comparisons with no verdict bear on how the list was assembled. The benchmark's price-only return against the index's own price return has a tracking error of 93.6 basis points a year over the months to December 2004, when the index weighted each stock by its full capitalisation, and 22.4 from January 2005, when it weights by float. The gap

before 2005 is therefore evidence on membership independent of CRSP's list, removals included. From July 1999 at least 99.4 per cent of the universe's members and at least 99.0 per cent of its capitalisation have a GICS sector in every month, which is why the evaluation window runs the 306 months from July 1999.

Against French's portfolios, the universe's spreads track his large-cap spreads without coinciding with them (E3). Over the 1,172 months from January 1927 to December 2024 with a value in both, the momentum spread has a correlation of 0.888 with French's, and their mean monthly difference is minus 10.4 basis points on a standard error of 9.0, or 1.15 standard errors. Over the 417 months from April 1990 the variance spread has a correlation of 0.954 and a mean difference of plus 2.7 basis points on 9.4, or 0.29 standard errors. With no verdict, the momentum correlation is 0.846 over the 359 months to March 1957, when the universe held about 90 names, and 0.927 over the 813 months from April 1957, consistent with the two sets overlapping more once the universe holds about 500 names.

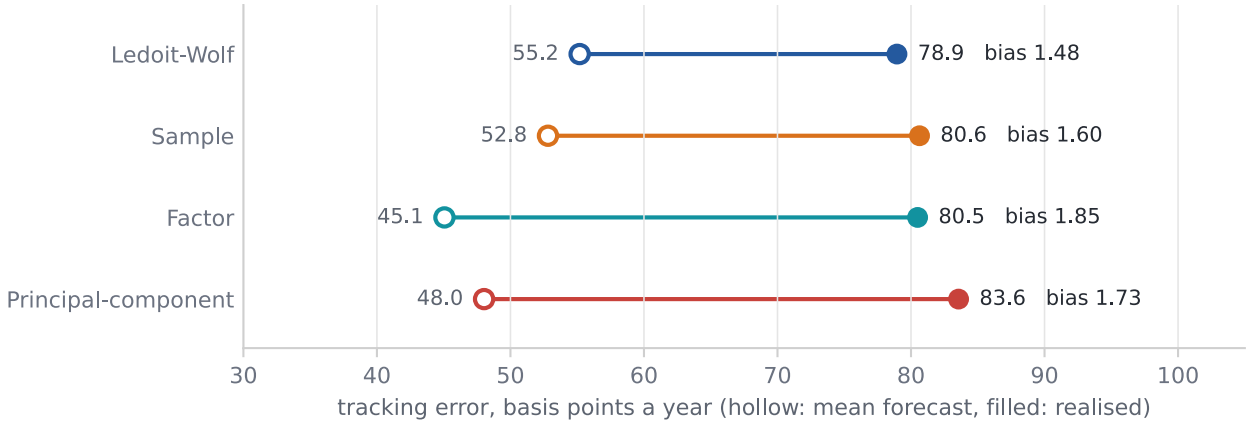
3.3 The covariance estimators

The first clause of E4 is not met. The sample matrix's condition number is at least 10 times the Ledoit-Wolf matrix's in 86 of the 306 months, all of them from 2017, and the ratio is 5.4 at the median and 2.8 at the least. With about 490 members on about 755 days the sample matrix is not singular, and the Ledoit-Wolf shrinkage intensity lies between 1.1 and 7.1 per cent, so the shrunk matrix stays close to the sample one. The months that reach the level have a cause found after the run. From 2014 the universe holds up to five companies with two share classes each, and where both classes are complete members their daily returns have a correlation of 0.98 at the median, so the difference between the two is a portfolio with almost no variance and sets the smallest eigenvalue of the sample matrix. With one share class for each company the level is reached in 30 of the 306 months, all from 2020 to 2022, and the median ratio is 5.0. This check carries no verdict, and the basis of E4 stays as it was written before the run.

3.4 Tracking error under the mandate

Figure 1 and Table 2 report the tracking error the mandate forecast and realised. Under the first three constraint sets nothing pushes the portfolio off the benchmark, and it stays there in every month: the largest absolute active weight of the twelve paths is 2.4e-6. Under the mandate at the 2 per cent cap the realised tracking error is 78.9 basis points a year on the Ledoit-Wolf estimator, 80.6 on the sample estimator, 80.5 on the factor estimator and 83.6 on the principal-component estimator, against mean forecasts of 55.2, 52.8, 45.1 and 48.0. The sample estimator's realised figure lies above the Ledoit-Wolf estimator's, as the second clause of E4 expects, but the difference of 1.7 basis points is smaller than the sampling error of either figure over 306 months, about 3, so the clause is met without separating the two estimators.

Figure 1. Forecast and realised tracking error under the mandate, four covariance estimators, July 1999 to December 2024

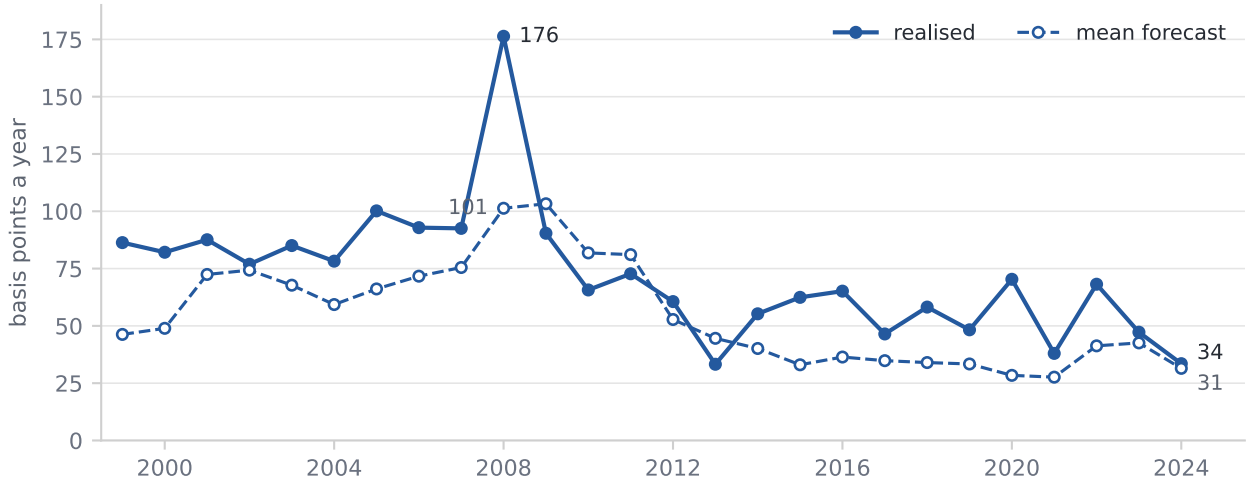


Basis points a year. Hollow: the mean of the monthly forecasts. Filled: the realised tracking error. Simulated, monthly rebalancing, before fees. Source: results/20260928/paths_summary.json, drawn by notebook 12.

E6 is not met for any of the three structured estimators. Their realised tracking error lies inside the range of 50 to 150 basis points a year, at 78.9, 80.5 and 83.6, but the bias statistic, 1.48 on the Ledoit-Wolf estimator, 1.85 on the factor estimator and 1.73 on the principal-component estimator, lies above the band's upper end of 1.25 for all three, and the sample estimator's, with no verdict, is 1.60. Every estimator's forecast therefore understates the mandate's realised

tracking error, and the understatement varies over time. On the Ledoit-Wolf path the realised tracking error by year runs from 33 basis points in 2013 to 176 in 2008, when the mean forecast was 101, and the forecast lies below the realised figure in 22 of the 26 years, above it only in 2009, 2010, 2011 and 2013, as Figure 2 shows, so the understatement is not confined to one crisis.

Figure 2. Realised and forecast tracking error by year, the Ledoit-Wolf estimator under the mandate, 1999 to 2024



Basis points a year. The realised tracking error of each calendar year and the mean of its monthly forecasts. The year 1999 holds six months. Simulated, before fees. Source: results/20260928/paths_by_year.csv, drawn by notebook 12.

Table 2. The twelve paths under the mandate, July 1999 to December 2024

Path	*Realised tracking error*	*Mean forecast*	*Bias statistic*	*Active return, gross*	*Newey-West standard error*	*Active return, net at 10 bp*	*Total turnover, % a year*	*Chosen turnover, % a year*	*Names held, mean*
Sample, cap 2 per cent	80.6	52.8	1.60	+13.7	18.6	+11.1	26.3	24.2	375
Ledoit-Wolf, cap 2 per cent	78.9	55.2	1.48	+12.7	18.2	+10.1	26.3	24.2	388
Factor, cap 2 per cent	80.5	45.1	1.85	+8.7	17.5	+7.1	15.1	13.2	473
Principal-component, cap 2 per cent	83.6	48.0	1.73	+0.2	17.4	−1.7	19.3	17.3	459
Ledoit-Wolf, cap 0.5 per cent	76.3	64.5	1.26	+7.2	17.2	+6.3	8.8	6.3	427
Ledoit-Wolf, cap 1 per cent	78.4	58.3	1.39	+13.4	18.2	+11.9	14.6	12.2	404
Ledoit-Wolf, cap 4 per cent	79.0	54.3	1.52	+10.1	18.3	+5.4	47.4	45.7	383
Ledoit-Wolf, cap none	79.4	54.1	1.53	+11.3	18.9	+5.2	61.0	59.4	381
Factor, cap 0.5 per cent	83.3	46.6	1.82	+7.2	17.5	+6.3	8.6	6.2	474
Factor, cap 1 per cent	82.3	45.3	1.85	+8.7	17.6	+7.4	12.6	10.5	474
Factor, cap 4 per cent	79.6	45.0	1.85	+8.4	17.5	+6.7	16.5	14.6	473
Factor, cap none	79.6	45.0	1.85	+8.5	17.5	+6.8	16.7	14.9	473

Tracking error, active return and its standard error in basis points a year, turnover one-way in per cent a year. Simulated, monthly rebalancing, gross of fees. The net active return is the gross less 10 basis points per unit of one-way turnover, forced turnover included. Source: results/20260928/paths_summary.json.

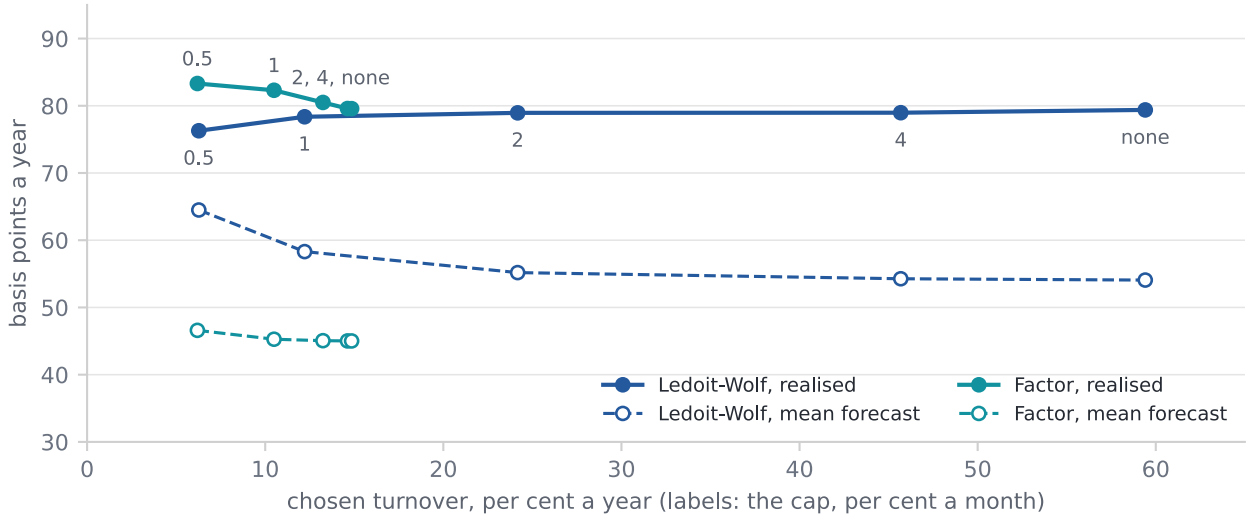
The active return is small beside its standard error on every path. The mean active return is 12.7, 13.7, 8.7 and 0.2 basis points a year gross of costs on the Ledoit-Wolf, sample, factor and principal-component estimators, each within one Newey-West standard error of zero (17.4 to 18.6), and 10.1, 11.1, 7.1 and minus 1.7 net of costs at 10 basis points. On the Ledoit-Wolf estimator the cap on chosen turnover binds in every month, and the portfolio holds 388 names on average and 333 at the least.

3.5 The cap on turnover

Figure 3 reports the frontier of tracking error against chosen turnover, and E5 is not met. For both estimators the mean forecast falls as the cap loosens, from 64.5 basis points a year at 0.5 per cent to 54.1 with no cap on the Ledoit-Wolf estimator and from 46.6 to 45.0 on the factor estimator, so the first part holds. The realised tracking error at the 4 per

cent cap is 79.6 on the factor estimator, below its 83.3 at 0.5 per cent, and 79.0 on the Ledoit-Wolf estimator, above its 76.3, so the second part fails for Ledoit-Wolf, and E5, one verdict over the two, is not met.

Figure 3. Tracking error against chosen turnover at five caps, the Ledoit-Wolf and factor estimators under the mandate, July 1999 to December 2024



Realised (filled) and mean forecast (hollow) tracking error in basis points a year, against chosen turnover in per cent a year. The labels give the cap in per cent a month, and caps whose turnovers lie within 2 per cent a year of each other share a label. Simulated, before fees. Source: results/20260928/paths_summary.json, drawn by notebook 12.

On the Ledoit-Wolf estimator the frontier is close to flat in what the mandate realises. Across the five caps the Ledoit-Wolf estimator's realised tracking error lies between 76.3 and 79.4 basis points a year, while its chosen turnover runs from 6.3 to 59.4 per cent a year. The factor estimator chooses at most 14.9 per cent a year, and the 2 per cent cap binds on it in 13.7 per cent of months, against every month on Ledoit-Wolf. On the Ledoit-Wolf estimator the extra turnover therefore buys a lower forecast and no lower realised tracking error, which is consistent with the optimiser trading on the estimation error in the forecast. On the factor estimator, which trades far less, the realised tracking error falls with the forecast, from 83.3 to 79.6.

In a month where the other constraints need more chosen turnover than the cap allows, the cap gives way to the least turnover they need. This happens in the first month of every capped path under the mandate, July 1999, when the portfolio leaves the benchmark for the mandate and needs 6.41 per cent, and in 6 further months at the 0.5 per cent cap on the Ledoit-Wolf estimator and 3 on the factor estimator. So on the Ledoit-Wolf estimator at caps of 0.5, 1 and 2 per cent, where the cap binds in every month, chosen turnover a year lies a little above twelve times the cap.

3.6 What the turnover is made of

E7 is met. On the Ledoit-Wolf path at the 2 per cent cap forced turnover is 2.99 per cent a year of a total of 26.26, or 11.4 per cent, below the third fixed before the run. Reweighting is 3.67, rebalancing 24.94 and the overlap 5.34. A portfolio that holds the benchmark turns over 5.78 per cent a year, forced turnover 2.80, reweighting 3.71 and no rebalancing, so nearly all of what the mandate trades beyond the benchmark's own turnover is the optimiser's choice.

3.7 The attribution by sector

E8 is met. The largest gap between the sum of the three terms and the active return, in any month of the four paths, is 9.6e-17. The mean absolute monthly allocation, selection and interaction are 16.6, 13.8 and 1.8 basis points on the Ledoit-Wolf path, 16.7, 10.6 and 1.5 on the factor path and 16.4, 12.3 and 2.2 on the principal-component path, and, with no verdict, 16.6, 14.4 and 1.9 on the sample estimator's path. Month by month, then, the mandate's sector weights move its return as much as or more than the stocks it holds within a sector, as the reasoning behind E8 expected.

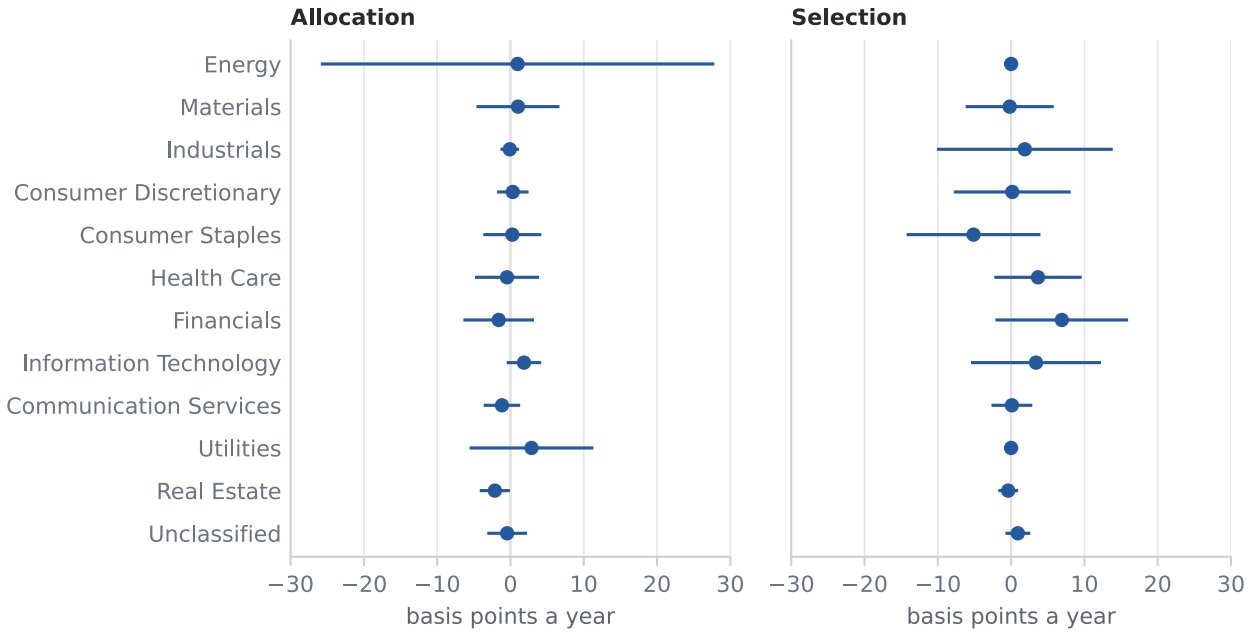
Table 3. The three terms over the window, four paths under the mandate at the 2 per cent cap, July 1999 to December 2024

Path	*Allocation, bp a year*	*Standard error*	*Selection, bp a year*	*Standard error*	*Interaction, bp a year*	*Standard error*	*Allocation, mean absolute, bp a month*	*Selection, mean absolute, bp a month*	*Interaction, mean absolute, bp a month*	*Sector-months not held*
Ledoit-Wolf	+1.4	15.9	+11.3	14.0	0.0	2.0	16.56	13.81	1.76	16
Sample, no verdict	+1.3	16.1	+12.3	14.7	+0.1	2.2	16.63	14.37	1.86	22
Factor	+5.5	16.3	+4.0	9.3	−0.8	1.5	16.74	10.63	1.50	0
Principal-component	+3.5	15.4	−3.9	9.6	+0.6	2.5	16.41	12.33	2.21	0

Annual averages with Newey-West standard errors in basis points a year, and E8's mean absolute monthly terms in basis points a month, each month's figure the sum over the twelve sectors. Source: results/20260928/brinson_summary.json and verdicts_milestone4.json.

Over the window the three terms are small beside their standard errors. Table 3 shows that on the Ledoit-Wolf path the active return of 12.7 basis points a year splits into allocation 1.4, selection 11.3 and interaction 0.0, with Newey-West standard errors of 15.9, 14.0 and 2.0, so none is distinguishable from zero. Allocation is 1.3, 5.5 and 3.5 on the sample, factor and principal-component paths, and selection 12.3, 4.0 and minus 3.9, and no total on any path lies one standard error from zero. Figure 4 breaks the Ledoit-Wolf path's allocation and selection down by sector. On this path 3 of the 36 sector terms, interaction included, which the figure does not show, lie two standard errors or more from zero, and 1 to 4 on the other paths, against, at the 5 per cent level, 1.8 on average if the tests were independent.

Figure 4. Allocation and selection by sector over the window, the Ledoit-Wolf path under the mandate, July 1999 to December 2024



Annual averages in basis points a year, each with a bar of two Newey-West standard errors on each side. Interaction is in Table 3's source file. Simulated, gross of fees. Source: results/20260928/brinson_summary.json, drawn by notebook 12.

The two tilted sectors show how the attribution reads the tilt. In Energy and Utilities selection is within a tenth of a basis point a year of zero on all four paths, because a sector whose stocks all sit at the same fraction of their benchmark weights earns the benchmark sector's return, and the tilt's cap at half binds on almost every stock: on the Ledoit-Wolf path the stock sits at half its benchmark weight, within one millionth, in 19,431 of the 19,433 stock-months in the two sectors. Their whole effect is therefore allocation, 1.0 basis points a year for Energy and 2.9 for Utilities on the Ledoit-Wolf path, on standard errors of 13.4 and 4.2, so over the window the tilt neither cost nor earned a return that the data can distinguish from zero. A sector is not held in 22 months on the sample estimator's path and in 16 on the Ledoit-Wolf path, all of them from 2022, and 66 member-months have no return and are left out of both sides.

3.8 The attribution by factor

E9 is met. Contributions and residual sum to the active return within 8.7e-19 in every month, and the mean active exposure to the market factor lies between −0.0006 and +0.0001 on the four paths. Table 4 reports the six contributions, their sum, the residual and the active return, each with the plain standard error the copied attribution computes, with no allowance for correlation from one month to the next. On the Ledoit-Wolf path the six contributions sum to −9.7 basis points a year and the residual to 22.4, on plain standard errors of 4.4 and 14.9. The six factors explain 2 to 13 per cent of the variance of the active return on the four paths, so the mandate's active risk is mostly sector and stock risk that these factors do not describe.

Table 4. The factor attribution, four paths under the mandate at the 2 per cent cap, July 1999 to December 2024

	Ledoit-Wolf	*Sample*	*Factor*	*Principal-component*
Market, contribution	−1.7 (0.5)	−1.6 (0.5)	−1.4 (0.5)	−1.2 (0.7)
Size, contribution	+0.7 (1.2)	+0.6 (1.2)	+0.9 (1.2)	+2.1 (1.9)
Value, contribution	−2.0 (2.5)	−1.9 (2.5)	−1.8 (1.5)	−5.4 (4.6)
Profitability, contribution	−4.7 (3.5)	−4.9 (3.4)	−2.3 (2.4)	−6.5 (5.5)
Investment, contribution	−1.3 (1.5)	−1.1 (1.5)	−0.8 (1.3)	−1.7 (2.5)
Momentum, contribution	−0.8 (2.2)	−0.7 (2.1)	0.0 (1.7)	+0.7 (1.8)
The six together	−9.7 (4.4)	−9.6 (4.4)	−5.4 (3.3)	−12.0 (5.7)
Residual	+22.4 (14.9)	+23.4 (15.3)	+14.1 (15.8)	+12.3 (15.4)
Active return	+12.7 (15.6)	+13.7 (16.0)	+8.7 (15.9)	+0.2 (16.5)
Factor share, per cent	8.9	8.6	2.1	12.9
Market, mean active exposure	−0.0006	−0.0006	+0.0001	−0.0002
Size, mean active exposure	+0.0055	+0.0053	+0.0054	+0.0056
Value, mean active exposure	−0.0068	−0.0067	−0.0019	−0.0100
Profitability, mean active exposure	−0.0071	−0.0071	−0.0041	−0.0162
Investment, mean active exposure	+0.0045	+0.0047	+0.0006	+0.0002
Momentum, mean active exposure	−0.0028	−0.0028	−0.0023	−0.0018

Contributions, their sum, the residual and the active return in basis points a year with plain standard errors in brackets, the factor share in per cent, and the mean active exposures. Source: Kenneth R. French Data Library, results/20260928/factor_attribution_summary.json.

3.9 What the estimates exclude

Taken together, the estimates bound the mandate's active return without locating one. The active return of every path under the mandate lies within one Newey-West standard error of zero, and with standard errors of 17.2 to 18.9 basis points a year over 306 months, a mean active return smaller than about twice that in absolute value cannot be told apart from none. The terms of the attribution over the window are of the same order and carry the same bound. What the estimates do establish is the size and the source of the tracking error: about 80 basis points a year, understated by every forecast, moved month by month more by sector weights than by the stocks held within sectors, and the make-up of the turnover, of which index changes force 11.4 per cent.

4. Discussion

The natural comparison is with the companion study on French's 49 industries, which ran a mandate of the same design on the same four estimators and a fifth from monthly returns and found realised tracking errors of 84.7 to 93.4 basis points a year on its four estimators from daily returns and bias statistics of 1.22 to 1.61 on its five estimators (Hofer, 2026). At the stock level the realised figures are of the same order, 78.9 to 83.6, and the forecasts understate them by more, with bias statistics of 1.48 to 1.85. The finding that the forecast falls short of the realised tracking error therefore carries over from industries to stocks, and on these estimates it widens, plausibly because a stock-level matrix has more dimensions in which the optimiser can find an estimation error to trade on, as the Ledoit-Wolf frontier of Figure 3 suggests.

For a manager who runs such a mandate, the forecast is the figure reported to the client before the month, and the realised the one measured afterwards. The evidence therefore indicates that a risk budget set on a three-year covariance forecast of this kind would have been exceeded on average, with bias statistics of 1.48 to 1.85, and by most in the

months when volatilities rose, as in 2008. It also indicates that on the Ledoit-Wolf estimator a looser turnover cap lowers the forecast without lowering what is realised, while on the factor estimator, which trades far less, the realised tracking error falls from 83.3 to 79.6. The attribution adds that the mandate's month-to-month deviation comes more from sector weights than from the stocks held within sectors, the underweight of Energy above all, whose allocation has a standard error of 13.4 basis points a year against 4.2 or less for every other sector, and that over 25 years neither the sector weights nor the stocks held within them leave a result that the data distinguish from zero.

At the same time, the design has limits, each set by a feature of the data or the method. First, every figure of the mandate is a simulation, with costs proportional to turnover and no market impact, since the extraction holds no trading data. Second, the benchmark weights CRSP's list by full capitalisation, and the index has weighted by float since 2005, which CRSP's share counts cannot rebuild, so the tracking errors are measured against the full-capitalisation benchmark and would differ against the index itself. Third, the evaluation window starts in July 1999 because Compustat's GICS history begins on 30 June 1999, so the mandate is not observed before then. Fourth, the attribution sums its monthly terms arithmetically, so a year's terms add to the sum of its monthly active returns, which differs from the active return compounded over the year, and another ordering of the definitions would move part of the interaction into allocation or selection. Finally, the factor attribution reads every path through loadings estimated on daily returns over the three years before the month, applied to monthly factor returns, so a loading measured with error moves part of a contribution into the residual.

What would narrow these limits is clear from their causes. Measuring the tracking error against the index as published would require its float-adjusted weights, which are not in this extraction. Testing whether a calibrated forecast removes the understatement would require a calibration fixed before the run, such as the companion study's scaling of each forecast by the path's own realised-to-forecast ratio over the previous 60 months, and run on these paths. And separating the optimiser's trading on estimation error from its trading on genuine changes in risk would require a forecast built from data the optimiser does not see.

Against this evidence, the narrow claim of this note is that a long-only mandate held close to a cap-weighted benchmark of US large caps, under a client's bounds, a tilt and a turnover cap, realised about 80 basis points a year of tracking error from 1999 to 2024 under each of four common covariance estimators, that each estimator's forecast understated it, with bias statistics of 1.48 to 1.85, and that its month-to-month deviation came more from sector weights than from the stocks held within sectors, while none of its active return over the period can be distinguished from zero. The evidence therefore suggests that the forecast tracking error of such a mandate is worth reporting beside a measure of its own past accuracy, which is what the bias statistic supplies.

References

Brinson, G. P. and Fachler, N. (1985). Measuring non-U.S. equity portfolio performance. *Journal of Portfolio Management* 11(3), 73 to 76. https://doi.org/10.3905/jpm.1985.409005

DeMiguel, V., Garlappi, L. and Uppal, R. (2009). Optimal versus naive diversification: How inefficient is the 1/*N* portfolio strategy? *Review of Financial Studies* 22(5), 1915 to 1953. https://doi.org/10.1093/rfs/hhm075

Fama, E. F. and French, K. R. (2015). A five-factor asset pricing model. *Journal of Financial Economics* 116(1), 1 to 22. https://doi.org/10.1016/j.jfineco.2014.10.010

Hofer, C. (2026). Optimal versus naive diversification: a replication, and what a tracking-error mandate changes. Note, October 2026. https://www.carlohofer.com/optimal-vs-naive-diversification/note/B-P_note_2026-10.pdf

Ledoit, O. and Wolf, M. (2004). A well-conditioned estimator for large-dimensional covariance matrices. *Journal of Multivariate Analysis* 88(2), 365 to 411. https://doi.org/10.1016/S0047-259X(03)00096-4

Newey, W. K. and West, K. D. (1987). A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica* 55(3), 703 to 708. https://doi.org/10.2307/1913610

Shumway, T. (1997). The delisting bias in CRSP data. *Journal of Finance* 52(1), 327 to 340. https://doi.org/10.1111/j.1540-6261.1997.tb03818.x

Other projects

Optimal versus naive diversification. Under an index-tracking mandate on French's 49 industries, five covariance estimators understate their own tracking error, by 22 to 61 per cent. The replication of DeMiguel, Garlappi and Uppal (2009) reproduces 29 of the 30 Sharpe ratios that carry a band. Page, dashboard, note, code

Climate risk as an equity signal. An ownership graph linking 328 listed companies to the 5,115 tracked assets they own. Of 19 apparent ownership changes checked by hand, 12 had no transaction behind them. Page, note, code

Volatility-managed portfolios: a replication, the real-time critique, and the months after the sample. Volatility estimation and position sizing on Ken French's equity factors, backtested out of sample and without look-ahead, replicating Moreira and Muir (2017) and Cederburg, O'Doherty, Wang and Yan (2020). Scaled by the inverse of last month's realised variance, a factor has a Sharpe ratio above the original's beyond its 95 per cent interval for one of six factors, momentum, and the 136 months after Moreira and Muir's sample cannot tell any difference from zero. Page, dashboard, note, code

Rules-based portfolio. Runs a real-money portfolio under rules published before the first order: two UCITS ETFs at 70/30, a monthly cycle with pre-trade checks, look-through exposure and ex-ante risk on a public dashboard. Page, dashboard, guide, code