Which risk model for an index mandate? Covariance estimators and tracking error on 49 US industries

Carlo Hofer. November 2026. Code and notebooks: github.com/ochofer/risk-models-for-index-tracking.

I ask which covariance estimator a long-only portfolio should use to track a cap-weighted benchmark of Ken French's 49 US industries under a mandate, from July 1979 to August 2026. Each month an optimiser, a routine that minimises the forecast tracking error, builds the portfolio with one of six covariance estimators, and I compare the tracking errors the optimised portfolios realise. I also judge the estimators as the literature usually does, by the volatility of the minimum-variance portfolio built from each, so that a difference from the usual results can be traced to the data or to the mandate. On monthly returns, the tracking error of the six-factor and the principal-component models exceeds that of the sample covariance of three years of daily returns by a statistically significant 7 to 9 basis points a year. The two models set the covariance between the industries' residuals to zero, and with half of that covariance included their excess is no longer statistically significant. When the four estimators built from daily returns use only the last three months of them, the two models' tracking error on daily returns falls below that of the sample covariance of the same months by a statistically significant margin. When weights may be negative, the two models' minimum-variance portfolios are more volatile than the sample covariance's by a statistically significant margin. When weights are long-only, they are about as volatile, with no statistically significant difference. Every estimator forecasts too little tracking error. The shortfall is statistically significant for every estimator before 2000, and from 2000 only for the sample covariance of 120 monthly returns.

Terms

Benchmark: the cap-weighted portfolio of French's 49 industry portfolios, which the fund tracks. **Mandate**: the rules the fund's portfolio follows: long-only weights, each industry other than Coal, Oil and Utilities within 2 percentage points of its benchmark weight (the bound), one-way turnover of at most 2 per cent a month (the turnover limit), the tilt, tobacco and weapons excluded, and a beta of one to the benchmark. **Tilt**: holding Coal, Oil and Utilities at no more than half their benchmark weight. **Tilt alone**: the bound, the turnover limit and the tilt, without the exclusions and the beta of one. **Optimiser**: the routine that each month chooses the long-only weights with the smallest forecast tracking error under the rules in force, starting from the previous month's weights. **Optimised portfolio**: the portfolios the optimiser builds month by month with one of the six covariance estimators, or the robust portfolio. Each is named by its estimator, so 'the six-factor model realises 93.4' means that the optimised portfolio built with it realises that tracking error. **Robust portfolio**: the portfolio whose largest forecast tracking error across five of the estimators is smallest. **Active weight**: an industry's portfolio weight minus its benchmark weight. **Active return**: the portfolio's return minus the benchmark's. **Tracking error**: the standard deviation of the active return, annualised by the square root of 12 on monthly returns and of 252 on daily returns. It is forecast before the month and realised after it. **Basis point**: a hundredth of a percentage point. **Covariance estimate**: the table of the variances and covariances of the industries' returns, from which the optimiser forecasts tracking error. **Shrinkage**: pulling an estimate part of the way towards a simpler target; the shrinkage intensity is the fraction of the way. **Structured models**: the six-factor model and the principal-component model, which build the covariance from six or five common sources of movement plus each industry's own variance. **Residual**: the part of a return a model does not explain. **Minimum-variance portfolio**: the fully invested portfolio with the lowest variance under a covariance estimate, unconstrained when weights may be negative and long-only when they may not. **Bias statistic**: the standard deviation of the realised active return divided by its forecast. It is 1 for an accurate forecast and above 1 when the forecast is too low. **Paired variance test**: the test of Ledoit and Wolf (2011) of whether two portfolios, measured on the same months or days, have the same variance. **Detectable difference**: the difference the paired test would find statistically significant at the 5 per cent level four times in five. **Holm's correction**: an adjustment of p-values for the number of comparisons read together (Holm, 1979). **Factor objective**: the sum of the portfolio's active exposures to value, profitability, investment and momentum, which the optimiser maximises in Section 3.7. **Budget**: the limit on the forecast tracking error under the factor objective, 75, 100 or 150 basis points a year. **Beta units**: the unit of an active exposure: the portfolio's beta to a factor minus the benchmark's. **Newey-West standard error**: a standard error of a mean of monthly figures that allows for correlation between months up to six months apart (Newey and West, 1987).

1. Introduction

An index fund faces a narrow version of the old problem of estimation error in portfolio construction. Its mandate fixes most of the portfolio, and the one estimate its optimiser needs is a covariance matrix that forecasts tracking error, since the fund holds no view on returns. The question for such a fund is therefore whether the choice of covariance estimator changes the tracking error it realises once the mandate's constraints bind, and which estimator to use if it does.

Dom, Howard, Jansen and Lohre (2025) ask this question of the minimum-variance portfolio of the 500 largest US stocks. They find that, when the portfolio's weights may be negative, shrinkage, alone or combined with weights that favour recent returns, realises a lower volatility than the sample covariance, while RiskMetrics, which favours recent returns without shrinkage, does not. Most of this advantage disappears once the portfolio is long-only with a cap on

each stock's weight and a penalty on trading, although favouring recent returns still lowers volatility. Whether their result carries over to an index fund is open, for two reasons. First, a mandate constrains the portfolio in a different way, holding each position close to its benchmark weight and capping trading. Second, the quantity it minimises, the variance of the portfolio's return relative to the benchmark, is more than a hundred times smaller than the variance of the market, so the effect of an estimation error cannot be read across from one to the other.

I ask the same question of a long-only index-tracking mandate on French's 49 US industry portfolios. Each month from July 1979 to August 2026, an optimiser finds the long-only portfolio with the smallest forecast tracking error under the mandate, using one of six covariance estimators, or forms a robust portfolio built from five of them. The outcome is the tracking error the portfolio then realises, measured on monthly and on daily returns. Four further analyses bear on the same question. In the first, I include in the two structured models part of the covariance between the industries' residuals that they otherwise set to zero. In the second, the four estimators built from daily returns use the last 3, 6 or 12 months of them in place of three years. In the third, I judge the same estimates as Dom, Howard, Jansen and Lohre (2025) judge theirs, by the volatility of the minimum-variance portfolio built from each, so that a difference from their results can be traced to the data or to what the portfolio is asked to do. In the fourth, the optimiser seeks exposure to four factors, value, profitability, investment and momentum, within a limit on the forecast tracking error under the same mandate, as an enhanced index fund does, and I set the exposure the portfolio's betas forecast beside the exposure its returns realised.

French's 49 industries are well suited to test this question, for three reasons. First, the data are public and each figure is computed on a recorded version of French's files, so it can be rebuilt. Second, the industries' weights in the market can be rebuilt from French's own firm counts and firm sizes, and the rebuilt benchmark tracks his market return within 47 basis points a year, so the cost of the mandate is measured against a known reference. Third, the daily returns of the 49 industries cover the whole evaluation, 11,888 trading days, so each comparison can be measured a second time with far more observations. The mandate is of a realistic size: the tracking error it reaches, about 1 per cent a year, is the ex-ante limit of Robeco's Global Developed Enhanced Index Equities strategy (Robeco, 2024, p. 19).

Empirically, the choice of estimator changes the realised tracking error under the mandate by up to about 10 per cent. The four estimators built from three years of daily returns realise a tracking error of 84.7 to 93.4 basis points a year. The two structured models realise 7 to 9 basis points a year more than the sample covariance and than shrinkage, and each of these four differences is statistically significant at the 5 per cent level on a paired test after Holm's correction across nine comparisons. Measured on daily returns, the principal-component model realises 3.4 and the six-factor model 4.1 basis points a year more than the sample covariance. When the two models include half of the residual covariance between industries, their excess is 1.6 to 2.1 basis points a year on monthly returns and about zero on daily returns, neither statistically significant at the 5 per cent level on the paired test after Holm's correction. When the four estimators built from daily returns use only the last three months of them, the two models realise 2.2 to 2.3 basis points a year less than the sample covariance of the same months on daily returns, statistically significant at the 5 per cent level on the paired test after Holm's correction. On minimum-variance portfolios the structured models realise 0.95 to 1.58 percentage points a year more volatility than the sample covariance when weights may be negative. With long-only weights, no estimate's volatility differs from the sample covariance's on monthly returns by a margin that is statistically significant, and the data rule out a difference of more than about 0.4 percentage points a year for the four estimators built from daily returns and about 1.0 for the monthly sample covariance. Every estimator's forecast of the tracking error is too low: over 1979 to 2026 the realised active return is 22 to 39 per cent more volatile than forecast for the five estimators built from daily returns and 61 per cent for the sample covariance of 120 monthly returns. The shortfall is statistically significant at the 5 per cent level before 2000 for every estimator, and from 2000 only for the sample covariance of 120 monthly returns. The tilt alone is associated with an active return of +1.8 to +6.9 basis points a year, which is not statistically significant at the 5 per cent level. The data rule out a cost of the tilt of more than about 21 basis points a year.

2. Data and design

All data come from Ken French's library and are downloaded when the code runs. Every figure here is computed on the files French built from the CRSP database of August 2026. Returns are monthly excess returns, and the daily returns of the 49 industries and of the factors are used for the covariance estimates and for the daily measurement.

The benchmark is the cap-weighted portfolio of the 49 industries, each industry weighted by its number of firms times its average firm size from French's files. Over the 686 months from July 1969 it tracks French's market return within 46.7 basis points a year, with a correlation of 0.9996. Each month from July 1979 to August 2026, 566 months, an optimiser chooses the long-only weights with the smallest forecast tracking error against this benchmark, starting from the previous month's weights as the month's returns have moved them. The mandate holds each industry other than Coal, Oil and Utilities within 2 percentage points of its benchmark weight, the bound, allows one-way turnover of at most 2 per cent a month, the turnover limit (relaxed only in a month in which no portfolio meets it together with the other rules, 2 to 4 of the 566 months under the mandate), holds those three at no more than half their benchmark weight, excludes tobacco and weapons, and sets the portfolio's beta to the benchmark at one. To separate the cost of the tilt from the rest of the mandate, I also run the same optimiser with the tilt alone, without the exclusions and the beta constraint.

I compare seven optimised portfolios, built with six covariance estimators, each estimated only from data before the month, and the robust portfolio, built from five of them. The first four are built from the daily returns of the three years before the month. The first is the sample covariance, called the daily sample covariance where it is set beside the monthly one. The second is Ledoit-Wolf shrinkage, which pulls the sample covariance towards a target with the same average variance and no covariances, by an amount the data choose (Ledoit and Wolf, 2004). The third is a six-factor model, which builds the covariance from each industry's exposures to the market, size, value, profitability, investment and momentum factors plus its own variance. The fourth is a principal-component model, which does the same with the benchmark and the four largest principal components of the part of the returns the benchmark does not explain. Both structured models set the covariance between the residuals of different industries to zero. The fifth is the sample covariance of the last 120 monthly returns, the monthly sample covariance for short. The sixth is RiskMetrics, an exponentially weighted covariance of five years of daily returns with a half-life of 252 trading days. The robust portfolio uses the first five.

I judge each optimised portfolio by its realised tracking error, its mean forecast, its bias statistic, its active return, its turnover and its trading cost at 50 basis points per unit of turnover. I compare the realised tracking errors of two optimised portfolios with the paired variance test, in its bootstrap form, on the months the two share, with blocks of twelve months, and on the 11,888 trading days, with blocks of 252 days. Each portfolio's daily active returns are built from the month's weights, drifting with the industries' daily returns until the month's last day. I apply Holm's correction where more than one comparison is read together. The standard error of a mean active return is the realised tracking error divided by the square root of the 47 years. Finally, each month's active return is split into the contributions of the six factors and a residual: an industry's exposures to the factors come from a regression of its daily returns on the factors over the previous three years, and the portfolio's active exposure to a factor is the sum of its active weights times those exposures.

For the minimum-variance comparison I build, each month, the unconstrained and the long-only minimum-variance portfolio of each of the six covariance estimators, hold each for the month, and compare each with the sample covariance's on monthly excess returns and on daily total returns, with Holm's correction within each of the four sets of five comparisons. Under the factor objective, the optimiser seeks the largest sum of the portfolio's active exposures to value, profitability, investment and momentum in place of the smallest forecast tracking error, under every rule of the mandate and a budget of 75, 100 or 150 basis points a year on the forecast tracking error.

3. Results

3.1 The estimators under the mandate

Table 1 reports the seven optimised portfolios under the mandate. Without the tilt, under the bound and the turnover limit alone, the portfolio realises a tracking error of at most 6.5 basis points a year, so almost all of the mandate's tracking error comes from the tilt. Under the full mandate the sample covariance realises a tracking error of 84.7 basis points a year, shrinkage 84.8, RiskMetrics 84.3, the robust portfolio 87.4, the principal-component model 91.5, the six-factor model 93.4 and the monthly sample covariance 95.8. The four estimators built from three years of daily returns lie within 8.8 basis points a year of one another. The tilt alone accounts for 94 to 95 per cent of these tracking errors, and the exclusions and the beta constraint add 4 to 5 basis points a year.

Table 1. The seven optimised portfolios under the mandate, 1979-07 to 2026-08 (566 months): realised tracking error, forecast calibration, active return and trading

Optimised portfolio	*Realised tracking error, bp a year*	*Mean forecast, bp a year*	*Bias statistic*	*Bias, calibrated*	*Active return, mandate, bp a year (s.e.)*	*Active return, tilt alone, bp a year (s.e.)*	*Mandate minus tilt alone, bp a year (s.e.)*	*One-way turnover, % a month*	*Trading cost at 50 bp, bp a year*
Sample covariance, 3 years daily	84.7	59.8	1.26	1.11	−6.9 (12.3)	+1.8 (11.7)	−8.7 (3.8)	1.31	7.8
Ledoit-Wolf shrinkage, 3 years daily	84.8	61.9	1.22	1.11	−6.4 (12.4)	+2.8 (11.7)	−9.1 (3.8)	1.23	7.4
Six-factor model, 3 years daily	93.4	60.8	1.37	1.17	−1.0 (13.6)	+6.9 (12.9)	−8.0 (3.4)	0.86	5.1
Principal components, 3 years daily	91.5	58.3	1.39	1.15	−5.8 (13.3)	+3.9 (12.6)	−9.7 (3.8)	0.96	5.8
Sample covariance, 120 months	95.8	55.0	1.61	1.12	−2.4 (13.9)	+6.6 (13.2)	−9.0 (4.7)	1.54	9.2
Robust (worst case of the five)	87.4	67.8	1.09	1.06	−4.8 (12.7)	+5.6 (12.0)	−10.4 (4.4)	1.06	6.4
RiskMetrics (exponentially weighted), 5 years daily	84.3	60.4	1.23	1.08	−5.3 (12.3)	+2.7 (11.6)	−7.9 (3.9)	1.44	8.6

Tracking error is the standard deviation of the monthly active return times the square root of 12. The bias statistic is the standard deviation of the monthly active return divided by the same month's forecast, so it weights each month by its own forecast and differs from the ratio of the realised column to the forecast column. In the calibrated column, each month's forecast is scaled by the same portfolio's bias over the previous 60 months, with nothing from the future. The standard error of an active return is its realised tracking error over the square root of 47 years. That of the mandate minus the tilt alone is a Newey-West standard error with six lags on the paired monthly differences. The benchmark itself turns over 0.39 per cent a month one way, which at 50 basis points per unit of turnover costs an index fund 2.3 basis points a year.

Table 2 reports the paired tests, which indicate that the choice of estimator changes the realised tracking error by up to about 10 per cent. Measured on monthly returns, the principal-component model realises 6.8 basis points a year more than the sample covariance, with a 95 per cent interval of 2.2 to 11.5, and the six-factor model 8.8 more, with an interval of 4.7 to 12.9. Each difference is statistically significant at the 5 per cent level on the paired test after Holm's correction, both across the six comparisons of Table 2 and across nine comparisons: the six against the sample covariance and the three pairs among shrinkage and the two structured models. Against shrinkage the two differences are 8.6 and 6.7, also statistically significant. Shrinkage itself differs from the sample covariance by 0.2 basis points, which is not statistically significant, and the data rule out a difference of more than about half a basis point: on three years of daily returns the two are almost the same estimate, with a median shrinkage intensity of 0.012, i.e., shrinkage moves the estimate 1.2 per cent of the way to its target. The monthly sample covariance realises 11.1 basis points a year more than the daily one, which is not statistically significant on monthly returns, where the test has too little power to detect a difference of that size: it would detect one of 27.7 four times in five. RiskMetrics realises 0.4 basis points less than the sample covariance (interval −2.3 to +1.5), which is not statistically significant either. With 566 months the test has too little power to detect a reduction of the size Dom, Howard, Jansen and Lohre (2025) report for RiskMetrics on their constrained portfolio, long-only with a cap on each stock's weight and a penalty on trading (GMV CON in their Table 3), 2.7 per cent, which is 2.28 basis points here.

Table 2. Paired tests of realised tracking errors under the mandate, 1979-07 to 2026-08, on 566 months and on 11,888 trading days: each optimised portfolio against the sample covariance of three years of daily returns, basis points a year

	On monthly returns					*On daily returns*					
Optimised portfolio	*Tracking error*	*Difference*	*95% interval*	*Detectable difference*	*Holm p-value*	*Tracking error*	*Difference*	*95% interval*	*Detectable difference*	*Holm p-value*	*Variance, monthly over daily*
RiskMetrics (exponentially weighted), 5 years daily	84.3	−0.38	−2.30 to +1.55	2.69	1.000	73.9	−0.59	−0.95 to −0.23	0.51	0.014	1.30
Ledoit-Wolf shrinkage, 3 years daily	84.8	+0.17	−0.19 to +0.53	0.50	1.000	74.3	−0.15	−0.29 to −0.01	0.19	0.035	1.30
Principal components, 3 years daily	91.5	+6.83	+2.16 to +11.49	6.21	0.025	77.9	+3.41	+2.43 to +4.40	1.37	0.0030	1.38
Six-factor model, 3 years daily	93.4	+8.79	+4.74 to +12.85	5.57	0.0030	78.6	+4.14	+2.72 to +5.57	1.87	0.0030	1.41
Sample covariance, 120 months	95.8	+11.11	−9.98 to +32.19	27.71	1.000	87.2	+12.74	+11.03 to +14.44	2.34	0.0030	1.21
Robust (worst case of the five)	87.4	+2.70	−13.61 to +19.02	20.01	1.000	76.5	+2.05	+0.86 to +3.24	1.59	0.0075	1.30

The variance test of Ledoit and Wolf (2011) in its bootstrap form: a studentised circular block bootstrap of the difference, with 2,000 resamples, in blocks of 12 months on monthly returns and of 252 trading days on daily returns. The sample covariance of three years of daily returns realises 84.7 basis points a year on monthly returns and 74.5 on daily returns, and its variance on monthly returns is 1.29 times its variance on daily returns. The difference is the estimate's tracking error minus the sample covariance's, and the interval is its 95 per cent interval. The detectable difference is the one the test would find statistically significant at the 5 per cent level four times in five. The Holm p-value is corrected for the six comparisons of each frequency. Across nine comparisons on monthly returns, these six and the three pairs among the other daily estimators, the two structured models' differences from the sample covariance and from shrinkage are each statistically significant at the 5 per cent level. The last column, variance on monthly over daily returns, is the portfolio's variance on monthly returns, per year, divided by its variance on daily returns, per year. It is 1 when the daily active returns are uncorrelated from one day to the next.

3.2 Months and days

Measured on the 11,888 trading days, the comparisons are far more precise: the standard error of RiskMetrics' difference from the sample covariance falls from 0.91 to 0.17 basis points. The second half of Table 2 reports them. On daily returns the principal-component model realises 3.4 and the six-factor model 4.1 basis points a year more than the sample covariance, each statistically significant at the 5 per cent level after Holm's correction across the six comparisons. The monthly sample covariance realises 12.7 more and the robust portfolio 2.1 more, and both differences are statistically significant at that level. RiskMetrics realises 0.59 less (interval −0.95 to −0.23), which is statistically significant. That comparison sets RiskMetrics' five years of daily returns against the sample covariance's three. Against the sample covariance on the same five years, as the paper compares them, RiskMetrics realises 0.37 less (−0.88 to +0.14), which is not statistically significant. The data rule out a reduction as large as the one Dom, Howard, Jansen and Lohre report for RiskMetrics on their constrained portfolio, 2.7 per cent, which is 2.01 basis points a year on daily returns. Shrinkage realises 0.15 less, also statistically significant. On a more cautious standard error, which allows the two portfolios' variances to drift apart over the years and is computed from the means of 47 consecutive blocks of 252 trading days, the differences of the two structured models and of the monthly sample covariance stay statistically significant after Holm's correction, and those of RiskMetrics, shrinkage and the robust portfolio do not.

The daily tracking errors are lower than the monthly ones, 73.9 to 87.2 basis points a year under the mandate against 84.3 to 95.8, because each portfolio's active return is positively correlated from one day to the next. A month therefore carries more variance than its days would if they were independent, by an amount that differs between risk models. The two structured models are the most affected, so their excess over the sample covariance on daily returns, +3.4 basis points for the principal-component model and +4.1 for the six-factor model, is about half their excess on monthly returns, +6.8 and +8.8. The monthly sample covariance's active returns are the least correlated from one day to the next: their variance on monthly returns is 1.21 times their variance on daily returns, against 1.29 to 1.41 for the other optimised portfolios (Table 2). A comparison on daily returns therefore understates the other portfolios' tracking error more than theirs, and they realise 12.7 basis points a year more than the sample covariance on daily returns against 11.1

on monthly returns. Shrinkage's difference from the sample covariance changes sign, +0.17 basis points a year on monthly returns and −0.15 on daily returns.

3.3 Why the structured models realise more

The two structured models assume that the part of each industry's return their factors leave unexplained moves independently of every other industry's. Under the mandate this assumption matters, because the six factors explain only 14 to 25 per cent of the active return's variance, so most of the tracking error sits in the residuals, whose co-movement the two models leave out. Two analyses bear on whether the excess of the two models is the cost of that assumption, set against the estimation error it saves: a model with fewer numbers to estimate is estimated more precisely.

In the first, I include in the two models a share of the residual covariance between industries, from none to all of it; with all of it, both become the sample covariance. On daily returns the six-factor model's tracking error falls from 78.6 basis points a year with none to 75.7 with a quarter, 74.4 with half and 74.0 with three quarters, against the sample covariance's 74.5. The principal-component model's falls from 77.9 to 75.5, 74.4 and 74.1. With half, neither differs from the sample covariance on daily returns by a margin that is statistically significant at the 5 per cent level. With three quarters both realise about 0.4 basis points a year less than the sample covariance on daily returns, statistically significant at the 5 per cent level after Holm's correction across ten comparisons, one for each model at each of the five shares from none to all. On monthly returns both fall at every step, to the sample covariance's 84.7. With half, the six-factor model realises 2.1 and the principal-component model 1.6 basis points a year more than the sample covariance on monthly returns, neither statistically significant at the 5 per cent level after Holm's correction across the ten comparisons on monthly returns.

In the second, I build the four estimators that use daily returns from the last 3, 6 and 12 months as well as the last three years. The window then holds about 2.5, 5 and 10 returns for each of the covariance matrix's 1,225 numbers, against 30 with three years. On daily returns, the six-factor model's excess over the sample covariance of the same window falls from 4.1 basis points a year with three years to 1.9 with one year and about zero with six months, and the principal-component model's from 3.4 to 1.5 and about zero. With three months the six-factor model realises 2.2 basis points a year less than the sample covariance and the principal-component model 2.3 less, each statistically significant at the 5 per cent level after Holm's correction. As the window shortens from three years to three months, shrinkage's reduction of the tracking error below the sample covariance's grows from 0.15 to 2.2 basis points a year, and its median shrinkage intensity from 0.012 to 0.097.

The two analyses point the same way. With three years of daily returns the sample covariance is estimated precisely, so the structured models' assumption plausibly costs more than the estimation error it saves, and the cost disappears once the models include part of the residual covariance. With less data the balance tips the other way, and the assumption saves more estimation error than it costs, in line with the trade-off Dom, Howard, Jansen and Lohre (2025) describe for the exact factor model.

3.4 The minimum-variance portfolios

Table 3 reports the same estimators judged as Dom, Howard, Jansen and Lohre (2025) judge theirs, by the volatility of the minimum-variance portfolio built from each. When weights may be negative, the principal-component model's portfolio realises a volatility 1.58 percentage points a year higher than the sample covariance's 11.93 per cent on monthly returns, and the six-factor model's 1.24 higher. The monthly sample covariance's portfolio realises 2.19 more. Each of the three differences is statistically significant at the 5 per cent level after Holm's correction. On daily returns the five differences have the same signs and all five are statistically significant: the principal-component model realises 1.09 and the six-factor model 0.95 percentage points more than the sample covariance's 11.16 per cent, the monthly sample covariance 4.32 more, RiskMetrics 0.28 less and shrinkage 0.06 less. Against the sample covariance on the same five years, as the paper compares them, RiskMetrics realises 0.58 percentage points a year less on daily returns when weights may be negative and 0.37 less long-only, each with a p-value of 0.0005, statistically significant at the 5 per cent level. On the paper's own comparison, then, RiskMetrics lowers the volatility of these portfolios when weights may be negative, where on the paper's stocks, unconstrained (GMV UNC), it raises it, 11.36 against 10.91. Long-only, no estimate's volatility differs from the sample covariance's 12.17 per cent on monthly returns by a margin that is statistically significant, the five lying between 0.09 percentage points below and 0.10 above it. On monthly returns the

data rule out a difference of more than about 0.4 percentage points a year for the four estimators built from daily returns and about 1.0 for the monthly sample covariance. On daily returns RiskMetrics realises 0.20 percentage points less than the sample covariance's 12.25 per cent and the monthly sample covariance 1.28 more, both statistically significant, and the other three lie within 0.03 points of it, above or below.

Table 3. Minimum-variance portfolios of six covariance estimates, 1979-07 to 2026-08: volatility on 566 months of excess returns and on 11,888 days of total returns, and the difference from the daily sample covariance's, percentage points a year

	On monthly returns			*On daily returns*		
Covariance estimate	*Volatility*	*Difference*	*Holm p-value*	*Volatility*	*Difference*	*Holm p-value*
Weights may be negative						
Sample covariance, 3 years daily	11.93			11.16		
RiskMetrics, 5 years daily	11.66	−0.28	0.103	10.89	−0.28	0.0025
Ledoit-Wolf shrinkage, 3 years daily	11.85	−0.08	0.313	11.10	−0.06	0.0025
Principal components, 3 years daily	13.51	+1.58	0.0025	12.25	+1.09	0.0025
Six-factor model, 3 years daily	13.17	+1.24	0.0025	12.11	+0.95	0.0025
Sample covariance, 120 months	14.12	+2.19	0.015	15.48	+4.32	0.0025
Long-only						
Sample covariance, 3 years daily	12.17			12.25		
RiskMetrics, 5 years daily	12.08	−0.09	1.000	12.05	−0.20	0.0025
Ledoit-Wolf shrinkage, 3 years daily	12.20	+0.03	1.000	12.26	+0.01	1.000
Principal components, 3 years daily	12.18	+0.01	1.000	12.27	+0.02	1.000
Six-factor model, 3 years daily	12.11	−0.06	1.000	12.27	+0.02	1.000
Sample covariance, 120 months	12.27	+0.10	1.000	13.53	+1.28	0.0025

Each month each estimate's minimum-variance portfolio is built from the data before the month and held for the month. Volatility is the standard deviation of the monthly excess return times the square root of 12, or of the daily total return times the square root of 252. The difference is the estimate's volatility minus the daily sample covariance's, tested with the paired variance test as in Table 2, and the Holm p-value is corrected for the five comparisons of each panel and frequency. RiskMetrics, built from five years of daily returns, is compared here with the sample covariance of three years, as every row is. Against the sample covariance of the same five years it realises 0.58 points less on daily returns when weights may be negative and 0.37 less long-only.

The paper measures volatility on daily portfolio returns (its Table 2's note), so the daily figures here are the ones to set beside its Table 3. My long-only minimum-variance portfolios correspond to the paper's long-only one (GMV LO). The paper's constrained one (GMV CON) also caps each stock's weight and penalises trading. Unconstrained (GMV UNC), its exact factor model realises 15.70 per cent against the sample covariance's 10.91, and RiskMetrics 11.36. Here too the structured models realise more than the sample covariance, while RiskMetrics realises 0.28 points less, the other way from the paper. Long-only (GMV LO), its exact factor model realises 14.50 against 12.24, and RiskMetrics 11.31. Here the structured models are within 0.06 points of the sample covariance, above or below, on monthly and on daily returns, unlike the paper's, which stays 2.26 points above it, and RiskMetrics is 0.20 below it on daily returns, as in the paper. The two settings differ in four ways: 500 US stocks against 49 industries; January 1995 to December 2021 against July 1979 to August 2026; five years of daily returns for every estimator against three years here, except RiskMetrics on five and the monthly sample covariance on 120 monthly returns; and an exact factor model with one factor, the market, against six factors here, and the benchmark with four components for the principal-component model.

The monthly sample covariance's long-only portfolios realise a volatility of 13.53 per cent on daily returns against 12.27 on monthly returns, while the other estimates' portfolios realise about the same on both, so its long-only difference from the sample covariance is 1.28 percentage points on daily returns and 0.10 on monthly returns.

3.5 The forecasts

Over 1979 to 2026 every estimator's forecast of the tracking error is too low. The bias statistic is 1.22 to 1.39 for the four estimators built from three years of daily returns and RiskMetrics, 1.61 for the monthly sample covariance and 1.09

for the robust portfolio, whose forecast is the largest of five by construction. The direction is what optimisation predicts: an optimiser that minimises a forecast picks the weights for which the forecast is lowest relative to the truth (Michaud, 1989), and the same estimators' forecasts for the individual industries, where no optimiser chooses, carry a bias statistic of 1.10. For the five estimators built from daily returns the forecasts are too low mostly before 2000. Their bias statistics are 1.59 to 1.90 in 1979 to 1989, when the oil price doubled and then collapsed while the portfolio held Oil, 15 per cent of the benchmark in 1979, at half its benchmark weight, and over the two periods 1.44 to 1.68 before 2000 against 1.01 to 1.12 from it. Each difference between the two periods is statistically significant at the 5 per cent level on a block bootstrap. No value from 2000 differs from 1 by a margin that is statistically significant at that level, and the data rule out a bias statistic above 1.24 from 2000. The forecasts of the sample covariance of 120 monthly returns are too low in both periods, at 1.59 before 2000 and 1.63 from it. Scaling each month's forecast by the same optimised portfolio's ratio of realised to forecast tracking error over the previous 60 months brings the five estimators built from daily returns to 1.08 to 1.17 and the monthly sample covariance to 1.12.

3.6 The active return, and where it came from

The mandate's return is in the last columns of Table 1. The tilt alone is associated with an active return of +1.8 to +6.9 basis points a year across the seven optimised portfolios, against standard errors of 11.6 to 13.2, which is not statistically significant at the 5 per cent level. The data rule out a cost of the tilt of more than about 21 basis points a year. The full mandate is associated with −6.9 to −1.0 basis points a year, against standard errors of 12.3 to 13.9, which is not statistically significant either. The data rule out a cost of the mandate of more than about 32 basis points a year. Measured on the same months, the exclusions and the beta constraint cost 7.9 to 10.4 basis points a year against the tilt alone, with Newey-West standard errors of 3.4 to 4.7, a difference that is statistically significant at the 5 per cent level for six of the seven. The mandate's portfolios turn over 0.86 to 1.54 per cent a month one way against the benchmark's 0.39. At 50 basis points per unit of turnover they pay 5.1 to 9.2 basis points a year in trading against an index fund's 2.3.

Table 4 splits the active return into the contributions of the six factors. The beta constraint holds: the active exposure to the market is between −0.005 and +0.001 in every optimised portfolio. The tilt leaves five other exposures, with the same signs under every estimate. Under the mandate the portfolio's active exposure is +0.010 to +0.018 to size, −0.010 to −0.015 to value, −0.008 to −0.012 to profitability, −0.005 to −0.008 to investment and −0.004 to −0.005 to momentum. The signs follow from what the tilt removes, since Oil and Utilities behave like large, cheap, profitable firms that invest little (e.g., their exposures to value are +0.27 and +0.35, against the benchmark's 0.00). The exposures are small because the bound is small: 2 percentage points per industry, times factor exposures that differ by about 0.5 across industries, gives about 0.01.

Table 4. Factor attribution of the active return, 1979-07 to 2026-08: mean active exposures, contributions and residual

Optimised portfolio	*Mkt-RF*	*SMB*	*HML*	*RMW*	*CMA*	*Mom*	*Factor contributions, bp a year (s.e.)*	*Residual, bp a year (s.e.)*	*Factor share*
Under the mandate									
Sample covariance, 3 years daily	−0.001	+0.013	−0.014	−0.011	−0.005	−0.004	−13.8 (7.6)	+6.9 (10.4)	0.22
Ledoit-Wolf shrinkage, 3 years daily	−0.001	+0.014	−0.015	−0.011	−0.005	−0.004	−13.8 (7.7)	+7.5 (10.4)	0.23
Six-factor model, 3 years daily	+0.001	+0.012	−0.010	−0.008	−0.005	−0.004	−10.9 (7.5)	+9.9 (12.4)	0.14
Principal components, 3 years daily	0.000	+0.010	−0.015	−0.012	−0.008	−0.005	−15.1 (8.5)	+9.3 (11.7)	0.20
Sample covariance, 120 months	−0.005	+0.018	−0.015	−0.008	−0.006	−0.005	−16.7 (7.5)	+14.3 (11.4)	0.25
Robust (worst case of the five)	−0.001	+0.013	−0.014	−0.009	−0.005	−0.004	−14.1 (7.5)	+9.3 (10.7)	0.22
RiskMetrics (exponentially weighted), 5 years daily	−0.001	+0.013	−0.015	−0.011	−0.005	−0.004	−13.9 (7.8)	+8.6 (10.5)	0.21
Under the tilt alone									
Sample covariance, 3 years daily	0.000	+0.011	−0.014	−0.008	−0.003	−0.003	−10.5 (6.7)	+12.2 (9.8)	0.23
Ledoit-Wolf shrinkage, 3 years daily	0.000	+0.012	−0.015	−0.008	−0.004	−0.003	−10.4 (6.8)	+13.2 (9.8)	0.24
Six-factor model, 3 years daily	+0.001	+0.010	−0.010	−0.005	−0.003	−0.004	−9.3 (6.7)	+16.2 (12.0)	0.14
Principal components, 3 years daily	0.000	+0.008	−0.015	−0.009	−0.006	−0.005	−12.4 (7.5)	+16.3 (11.0)	0.22
Sample covariance, 120 months	−0.004	+0.017	−0.016	−0.004	−0.005	−0.004	−13.1 (7.1)	+19.7 (10.9)	0.28
Robust (worst case of the five)	−0.001	+0.012	−0.013	−0.007	−0.004	−0.004	−11.1 (6.6)	+16.8 (9.4)	0.27
RiskMetrics (exponentially weighted), 5 years daily	0.000	+0.012	−0.014	−0.008	−0.003	−0.003	−10.7 (6.7)	+13.4 (9.9)	0.22

The active exposure to a factor is the sum over industries of the active weight times the industry's exposure to the factor, from a regression of its daily excess returns on the six factors over the three years before the month. The table shows its mean over the 566 months. A contribution is the exposure times the factor's return in the month, and the residual is the active return minus the six contributions. The factor share is one minus the variance of the residual over the variance of the active return. Mkt-RF is the market factor, SMB size, HML value, RMW profitability, CMA investment and Mom momentum (Fama and French, 2015; Carhart, 1997). The standard errors in brackets are Newey-West standard errors with six lags.

The six factors explain 14 to 28 per cent of the variance of the active return across the fourteen optimised portfolios, the seven under the mandate and the seven under the tilt alone, and 14 to 25 per cent under the mandate, so most of the tracking error comes from the residual. The attribution splits the tilt's near-zero return into two parts of opposite sign, neither statistically significant at the 5 per cent level on a Newey-West standard error. Under the tilt alone the factor contributions sum to −9.3 to −13.1 basis points a year, because the portfolio was short of value, profitability, investment and momentum, which all had positive mean returns over these years. The residual adds +12.2 to +19.7: the industries bought in place of Oil and Utilities returned more than Oil and Utilities did, beyond what their factor exposures explain.

3.7 The factor objective

Under the factor objective the optimiser spends tracking error on factor exposure: each month it maximises the sum of the portfolio's active exposures to value, profitability, investment and momentum within a budget on the forecast tracking error. Table 5 reports the optimised portfolios at a budget of 100 basis points a year. The combined exposure as the betas of the three years before each month forecast it, which is the quantity the optimiser maximises, is 0.095 to 0.140 in beta units, i.e., the portfolio's four factor betas together exceed the benchmark's by that much. Profitability and investment make up 74 to 85 per cent of it and momentum almost none, because the optimiser raises the exposures that cost the least forecast tracking error. Measured on the returns, by regressing each portfolio's 566 monthly active returns on the six factors' monthly returns and summing the four targeted slopes, the portfolios realise 0.40 to 0.56 of the exposure their betas forecast. The realised tracking error is 115.6 to 150.1 basis points a year, 16 to 50 per cent above the budget. Among the four estimators built from three years of daily returns, the sample covariance and shrinkage realise the least and the two structured models 13 to 18 basis points more, as in Section 3.1. The four targeted factors contributed +21.7 to +35.5 basis points a year, statistically significant at the 5 per cent level on a Newey-West standard error for five of the seven, all but the monthly sample covariance and the robust portfolio. The six factors together contributed +31 to +47. The residual, −28 to −37, is consistent with a forecast exposure above the realised one. No

optimised portfolio's active return, +2.3 to +15.5, is statistically significant at that level. At the realised tracking errors, an active return would have to be about 33 to 43 basis points a year to be statistically significant.

Table 5. The factor objective at a budget of 100 basis points a year, 1979-07 to 2026-08 (566 months): tracking error, forecast exposure and where the active return came from

Optimised portfolio	*Realised tracking error, bp a year*	*Realised over budget*	*Forecast combined exposure*	*Share in profitability and investment*	*Targeted factors, bp a year (s.e.)*	*Six factors, bp a year (s.e.)*	*Residual, bp a year (s.e.)*	*Active return, bp a year (s.e.)*
Sample covariance, 3 years daily	120.6	1.21	0.116	75%	+29.1 (13.2)	+41.6 (14.4)	−32.7 (15.9)	+8.9 (17.6)
Ledoit-Wolf shrinkage, 3 years daily	118.6	1.19	0.109	76%	+26.7 (13.1)	+39.2 (14.3)	−32.8 (15.5)	+6.5 (17.3)
Six-factor model, 3 years daily	133.6	1.34	0.123	74%	+33.3 (13.9)	+47.3 (15.1)	−31.8 (18.9)	+15.5 (19.5)
Principal components, 3 years daily	137.0	1.37	0.128	74%	+35.5 (14.7)	+47.3 (15.7)	−37.4 (18.4)	+9.9 (19.9)
Sample covariance, 120 months	150.1	1.50	0.140	85%	+34.8 (19.2)	+36.7 (19.2)	−30.7 (19.7)	+6.0 (21.9)
Robust (worst case of the five)	115.6	1.16	0.095	81%	+21.7 (13.4)	+30.7 (14.4)	−28.4 (15.0)	+2.3 (16.8)
RiskMetrics (exponentially weighted), 5 years daily	118.6	1.19	0.116	75%	+28.2 (12.8)	+38.1 (14.1)	−30.6 (15.9)	+7.5 (17.3)

The forecast combined exposure is the combined exposure as the betas of the three years before each month forecast it, which is the quantity the optimiser maximises: the mean over the 566 months of the sum of the active exposures to value, profitability, investment and momentum, in beta units. The share in profitability and investment is their share of it. A contribution is the forecast active exposure times the factor's return in the month, and the residual is the active return minus the six contributions. Their standard errors are Newey-West standard errors with six lags. The standard error of the active return is the realised tracking error over the square root of 47 years. The active return is gross of trading costs.

4. Discussion

The natural comparison is with Dom, Howard, Jansen and Lohre (2025), and the results indicate that their finding carries to a tracking-error mandate in part. Their finding that constraints shrink the differences between estimators holds here, although the mandate does not remove them: the two structured models still realise 7 to 9 basis points a year more than the sample covariance and shrinkage. On minimum-variance portfolios of the same industries the difference shrinks under the long-only constraint: the structured models realise 0.95 to 1.58 percentage points more than the sample covariance when weights may be negative and within 0.06 points of it, above or below, when they may not. On the paper's stocks the exact factor model stays 2.26 points above the sample covariance long-only (GMV LO). Their finding that favouring recent returns lowers risk is not established here under the mandate. On monthly returns the test has too little power to detect a reduction of their size. On daily returns, against the sample covariance of the same five years, RiskMetrics realises 0.37 basis points a year less, which is not statistically significant, and the data rule out a reduction as large as theirs. The structured models' excess is plausibly the cost of setting the covariance between the industries' residuals to zero. It disappears on daily returns when the models include half of that covariance, and it turns into an advantage when the four estimators built from daily returns use only the last three months of them.

For the holder of such a mandate, the estimates therefore suggest three things. First, among estimators built from three years of daily returns, the choice changes the tracking error by 7 to 9 basis points a year, up to about 10 per cent of it, and the simplest of them, the sample covariance and shrinkage, realise the lowest tracking error. Second, over 1979 to 2026 the forecast of each of them is too low: the realised active return is 22 to 39 per cent more volatile than forecast, against 61 per cent for the sample covariance of 120 monthly returns. Their shortfall is statistically significant at the 5 per cent level before 2000 and not from 2000, and scaling the forecast by the optimised portfolio's own recent history removes most of it. Third, the tilt's cost in return is not statistically significant at the 5 per cent level, the data ruling out a cost of more than about 21 basis points a year, but the tilt leaves factor exposures that no rule of the mandate sets and for which its objective has no term.

At the same time, the design has limits. First, the assets are industry portfolios, so the findings are about 49 groups of firms. A stock-level mandate holds hundreds of positions under a much tighter bound, and whether the choice of estimator changes the tracking error by as much there is a question these data cannot answer. Second, the daily and the monthly measurements of an optimised portfolio's tracking error differ by an amount that depends on how its active

return is correlated from one day to the next, so the size of a difference on one frequency does not carry to the other. Third, the design holds no view on expected returns, so what a return forecast would add to the mandate is outside it. Finally, French revises his files, so a rerun on a later version would reproduce the comparisons and not every digit.

Against this evidence, the claim is a narrow one. On French's 49 industries, under a mandate that holds each industry outside the tilt within 2 percentage points of its benchmark weight, trades at most 2 per cent a month and halves three industries, the choice of covariance estimator still changes the realised tracking error, by up to about 10 per cent, and the two structured models plausibly realise more than the sample covariance because they set the covariance between the industries' residuals to zero. Over 1979 to 2026 every estimator's forecast of the tracking error is too low, by a margin statistically significant at the 5 per cent level before 2000 for every estimator, and from 2000 only for the sample covariance of 120 monthly returns.

References

Carhart, M. M. (1997). On persistence in mutual fund performance. *Journal of Finance*, 52(1), 57-82. https://doi.org/10.1111/j.1540-6261.1997.tb03808.x

Dom, M. S., Howard, C., Jansen, M. and Lohre, H. (2025). Beyond GMV: the relevance of covariance matrix estimation for risk-based portfolio construction. *Quantitative Finance*, 25(3), 403-419. https://doi.org/10.1080/14697688.2025.2468268

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

French, K. R. (2026). Data library. mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html, files built from the CRSP database of August 2026.

Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics*, 6(2), 65-70.

Ledoit, O. and Wolf, M. (2004). A well-conditioned estimator for large-dimensional covariance matrices. *Journal of Multivariate Analysis*, 88(2), 365-411. https://doi.org/10.1016/s0047-259x(03)00096-4

Ledoit, O. and Wolf, M. (2011). Robust performance hypothesis testing with the variance. *Wilmott*, 2011(55), 86-89. https://doi.org/10.1002/wilm.10036

Michaud, R. O. (1989). The Markowitz optimization enigma: is 'optimized' optimal? *Financial Analysts Journal*, 45(1), 31-42. https://doi.org/10.2469/faj.v45.n1.31

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

Robeco (2024). The smarter alternative: Enhanced Indexing. Robeco Institutional Asset Management.