Optimal versus naive diversification: a replication, and the same question under an index mandate

Carlo Hofer. November 2026. Code and notebooks: github.com/ochofer/optimal-vs-naive-diversification.

DeMiguel, Garlappi and Uppal (2009) find that no optimised portfolio they test is consistently better than naive 1/N out of sample. I first rerun their comparison on their own sample of French's data, and the replication matches the paper. I then extend the replication to August 2026, and in the years since their sample no optimising rule's Sharpe ratio exceeds 1/N's by a margin that is statistically significant. Finally, I put their question to a practical test, an index mandate on 49 US industries from 1979 to 2026, in which I set seven optimised portfolios, each built every month from a covariance estimate, beside naive portfolios that meet the same rules without one. With Coal, Oil and Utilities held at no more than half their benchmark weight, the optimised portfolios realise 15 to 26 basis points a year less tracking error than a naive tilt that spreads the removed weight in proportion to the benchmark, and each difference is statistically significant. With a budget of tracking error spent on four factors, their forecast exposure to those factors, per basis point of tracking error, is 4.5 to 5.3 times that of a naive tilted index, and each difference is statistically significant. Measured on their returns, no difference in this exposure is statistically significant, so the returns do not show that the optimised portfolios realised more of it than the tilted index.

Terms of this study

1/N: the rule that holds every asset at the same weight and rebalances to it each month.

Benchmark: the cap-weighted portfolio of the 49 industries that the fund tracks.

Tilt: holding Coal, Oil and Utilities at no more than half their benchmark weight.

Mandate: the rules the fund's portfolio obeys, set out in Section 2.

Optimiser: the routine that chooses the portfolio's weights each month from a covariance estimate, within the mandate's rules.

Optimised portfolio: the portfolio the optimiser holds month by month with one covariance estimator, or the robust portfolio built from five of them, each formed from data before its month. There are seven in each setting.

Tilt alone: the first setting, in which the mandate holds the tilt without the two rules the factor objective adds.

Bound: the rule that holds each industry outside the tilt within 2 percentage points of its benchmark weight.

Turnover limit: the rule that each optimised portfolio trades at most 2 per cent of the portfolio a month one way.

Naive tilt: the benchmark with Coal, Oil and Utilities held at half their benchmark weight and the removed weight spread over the other 46 industries in proportion to their benchmark weights, rebalanced to that target every month, with no risk model.

Tilted index: the benchmark under the tilt and the exclusions of the factor objective, with each of the other 44 industries' weights changed by a percentage proportional to the distance of its factor score from their weighted average score, as far as the bound of 2 percentage points on each active weight and a floor of zero allow, rebalanced every month, with no risk model.

The two naive portfolios: the naive tilt and the tilted index together.

Factor: a return series that moves many assets at once. The six here are the market, size, value, profitability, investment and momentum (Fama and French, 2015; Carhart, 1997).

Active exposure: the portfolio's exposure to a factor minus the benchmark's.

Factor score: the sum of an industry's exposures to value, profitability, investment and momentum, the four targeted factors.

Factor objective: the sum of the portfolio's active exposures to value, profitability, investment and momentum.

Budget: the largest forecast tracking error the optimiser may take while it makes the factor objective as large as it can.

Forecast exposure: the factor objective as the optimiser computes it: the sum over industries of the active weights times each industry's betas to the four factors, from the three years before the month.

Realised exposure: the exposure measured after the fact: the sum of the slopes on the four factors when the portfolio's monthly active returns are regressed on the six factors' monthly returns.

Exposure per basis point: the mean forecast exposure over the months, or the realised exposure, divided by the realised tracking error in basis points a year.

Attribution: the split of each month's active return into the contributions of six factors and a residual, the part the factors do not explain.

Vintage: the version of French's files, named by the CRSP database they were built from.

Standard terms

Excess return: a return minus the one-month Treasury bill rate.

Sharpe ratio: the mean monthly excess return divided by its standard deviation.

Turnover: the sum of the absolute weight changes at a rebalance. One-way turnover is half of it, the share of the portfolio sold.

Estimation window: the months of returns a rule sees when it forms its weights, 120 in the paper.

Jobson-Korkie test: the test of whether two Sharpe ratios differ, with Memmel's correction, as the paper uses it.

Active weight: the portfolio's weight in an industry minus the benchmark's.

Active return: the portfolio's return minus the benchmark's.

Tracking error: the standard deviation of the monthly active return, times the square root of 12. 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.

Information ratio: the active return per year divided by the realised tracking error.

Paired variance test: the test of Ledoit and Wolf (2011) of whether two portfolios, measured on the same months, have the same tracking error.

Block bootstrap: resampling the months in blocks of twelve consecutive months, so that each resample keeps the months of a block together.

Holm's correction: an adjustment of p-values for the number of comparisons read together (Holm, 1979).

1. Introduction

Whether an optimised portfolio does better than the naive 1/N rule is an old question in portfolio construction, and DeMiguel, Garlappi and Uppal (2009) gave the answer that has become the reference point. On seven datasets, none of the fourteen models they evaluate is consistently better than 1/N out of sample, and they attribute this to estimation error: with 120 months of returns, the error in the estimated means and covariances costs more than optimising gains (Michaud, 1989; Jagannathan and Ma, 2003). An index fund faces the same choice in a narrower form. Its mandate fixes most of the portfolio and asks for no return forecast, so what the fund decides is where to put the weight the mandate moves away from the benchmark. It can let an optimiser place that weight with a covariance estimate, or place it by a fixed rule with no estimate at all. The question for such a fund is therefore DeMiguel, Garlappi and Uppal's: whether optimising does better than a naive portfolio once the mandate is set.

The closest existing evidence compares covariance estimators with one another. Dom, Howard, Jansen and Lohre (2025) compare them for the minimum-variance portfolio of the 500 largest US stocks, and find that most of the differences between them disappear once the portfolio is long-only with a cap on each stock's weight. A companion study (Hofer, 2026) makes the same comparison under a mandate that limits tracking error, on French's 49 industries, and finds that the choice of estimator changes the tracking error by about a tenth. Neither study sets the optimiser beside a portfolio that meets the same rules without any estimate. That is the comparison DeMiguel, Garlappi and Uppal's question needs, and it bears on whether a fund needs a risk model at all.

I do two things. First, I rerun DeMiguel, Garlappi and Uppal's comparison of optimised portfolios with 1/N on the four of their datasets that come from French's library, over their own sample, and then extend it to the 261 months since. Second, I ask their question under an index mandate on French's 49 industries, in two settings, each with a naive portfolio built for it. In the first setting, the tilt alone, the fund holds Coal, Oil and Utilities at no more than half their benchmark weight. There the optimiser holds the long-only portfolio with the smallest forecast tracking error, and the naive tilt halves the three and spreads the weight it removes over the other industries in proportion to their benchmark weights. In the second setting, under the factor objective, the mandate also excludes tobacco and weapons and holds the portfolio's beta to the benchmark at one. There the optimiser spends a budget of 100 basis points a year of forecast tracking error on exposure to value, profitability, investment and momentum, and the tilted index moves each industry's weight up or down with its exposure to these four factors, as far as the mandate's bound allows. The optimised portfolios are the companion study's, seven in each setting: one for each of six covariance estimators, and a robust portfolio built from five of them. Each naive portfolio is judged on what the optimiser is built to deliver in its setting: the realised tracking error under the tilt alone, and the exposure per basis point of realised tracking error under the factor objective, so that taking more risk does not count as obtaining more exposure.

French's 49 industries suit this question for two reasons. First, the data are public and every figure is computed on a recorded version of French's files, so every comparison can be rerun. Second, the tilt moves a large and changing amount of weight, between 2.3 and 15.9 per cent of the portfolio over the sample, so where that weight goes makes a measurable difference to the tracking error. The mandate is of a realistic size: the tracking error the optimised portfolios reach, 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 replication matches the paper. On its sample, 29 of the 30 Sharpe ratios I test are within 0.03 of the published values, and all 26 turnover figures of the portfolios that trade are within 25 per cent. In the 261 months since, no optimising rule's Sharpe ratio exceeds 1/N's by a margin that is statistically significant at the 5 per cent level on the Jobson-Korkie test. Under the mandate, the optimiser realises less tracking error than the naive tilt, and its forecast exposure per basis point is larger than the tilted index's. Under the tilt alone, the naive tilt realises 105.6 basis points a year of tracking error, 15 to 26 basis points more than each of the seven optimised portfolios. Under the factor objective, the optimised portfolios' forecast exposure per basis point, the exposure their own betas give, is 4.5 to 5.3 times the tilted index's. In each setting, each of the seven differences is statistically significant at the 5 per cent level on a paired test after Holm's correction. Measured on the returns, the optimised portfolios' realised exposure per basis point is 1.1 to 1.5 times the tilted index's, and no difference is statistically significant, with intervals from about 0.4 to about 2.7 times. Both naive portfolios earned more active return over the sample. On a paired test, no difference in information ratio between a naive portfolio and an optimised portfolio is statistically significant at the 5 per cent level after Holm's correction.

2. Data and design

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

I run the paper's nine rules on four of its datasets: Industry (ten industries and the market), MKT/SMB/HML (the three Fama-French factors), FF-1-factor (20 size and book-to-market portfolios and the market) and FF-4-factor (the same 20 portfolios with four factors). Each month a rule forms its weights from the previous 120 months and holds them for one month, over the paper's sample from July 1963 to November 2004, which gives 377 out-of-sample months. The rules are 1/N, sample mean-variance, mean-variance with the means shrunk towards a common value (Bayes-Stein, Jorion, 1986), minimum variance, the value-weighted market, the versions of mean-variance, Bayes-Stein and minimum variance that allow no negative weight, and minimum variance with every weight at least 1/(2N).

A replicated figure counts as a match when it is within a set margin of the published one: each Sharpe ratio within 0.03, each turnover figure within 25 per cent, and the top three rules of each dataset in the published order. I test a Sharpe ratio only where the paper's own turnover for the rule is below 100 times that of 1/N, because above that level ten basis points of noise added to every return move the rule's Sharpe ratio by 0.12 to 0.25. This leaves 30 tested Sharpe ratios and, among the portfolios that trade, 26 tested turnover figures. The margins were fixed before the code ran, except the limit on which Sharpe ratios are tested, set after the first replication results. Across all 36 out-of-sample Sharpe ratios, including the six the limit leaves untested, 31 are within 0.03.

I also rerun the paper's analytical result (its Proposition 1) and its simulation (its Table 6), and then run the nine rules on every month to August 2026. I compare a rule's Sharpe ratio with 1/N's using the Jobson-Korkie test with Memmel's correction (Jobson and Korkie, 1981; Memmel, 2003), two-sided. The paper's p-values match the one-sided form of the test, which Section 3.1 uses when it compares them.

The mandate's 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. Its tracking error against French's market return is 46.7 basis points a year over the 686 months from July 1969 (Hofer, 2026). Every portfolio under the mandate is long-only, holds each industry outside the tilt within 2 percentage points of its benchmark weight, which I call the bound, and holds Coal, Oil and Utilities at no more than half their benchmark weight. The optimised portfolios also trade at most 2 per cent of the portfolio a month one way, the turnover limit. The limit is relaxed only in the months in which the optimiser could not meet it together with the other rules, 2 or 3 of the 566 months under the tilt alone and 5 to 8 under the factor objective. Under the factor objective the mandate adds two rules: tobacco and weapons are excluded, and the portfolio's beta to the benchmark is held at one. The evaluation runs each month from July 1979 to August 2026, 566 months.

The optimised portfolios come from the companion study. I read them from its committed tables and copy the code that built them unchanged into this repository. In each setting, the notebook first rebuilds one of them from that code, the one built with the sample covariance of three years of daily returns, and checks that it matches the recorded one. The six

covariance estimators are each built only from data before the month: the sample covariance, Ledoit-Wolf shrinkage (Ledoit and Wolf, 2004), a six-factor model and a principal-component model, each from three years of daily returns; the sample covariance of 120 monthly returns; and RiskMetrics, an exponentially weighted covariance of five years of daily returns. Under the tilt alone, the optimiser chooses the weights with the smallest forecast tracking error. Under the factor objective, it chooses the weights with the largest factor objective whose forecast tracking error stays within the budget of 100 basis points a year. The seventh optimised portfolio is a robust portfolio that uses the first five estimators at once: under the tilt alone its weights make the largest of the five forecasts of tracking error as small as possible, and under the factor objective all five forecasts stay within the budget. An industry's betas to the six factors come from a regression of its daily excess returns on them over the three years before the month.

The two naive portfolios use no covariance estimate. The naive tilt takes each month's benchmark weights, halves Coal, Oil and Utilities, and scales the other 46 industries up by the same multiple so that the weights sum to one. The tilted index starts from the same tilt with tobacco and weapons excluded as well, and spreads the removed weight over the other 44 industries in proportion to their benchmark weights. It then multiplies each of the 44 weights by one plus k times the industry's centred score, the factor score minus the average score of the 44 weighted by those weights. The multiple k is one number for all 44, the largest at which every active weight stays within the bound and no weight falls below zero. An industry whose score is above the weighted average gains weight and one below it loses weight, and the weights still sum to one. Both naive portfolios start as the benchmark in July 1979, as the optimised portfolios do, and rebalance to their target every month. Their turnover is measured against their own weights as the previous month's returns have moved them. Neither is held to the turnover limit, and the tilted index is not held to a beta of one.

Under the tilt alone, I compare the naive tilt's realised tracking error with each optimised portfolio's using the paired variance test, in its bootstrap form with blocks of twelve months and 2,000 resamples, with Holm's correction across the seven comparisons. Under the factor objective, I compute the difference in forecast exposure per basis point on 2,000 block bootstrap resamples of the months, drawing the same months for the tilted index and the optimised portfolio. This gives the difference's 95 per cent interval and a two-sided p-value, and Holm's correction is applied across the seven. For the realised exposure, I regress each portfolio's 566 monthly active returns on the six factors' monthly returns, with an intercept, and add the slopes on the four targeted factors. Its difference per basis point is tested in the same way, with the regression run again on each resample. I give information ratios before costs and net of 50 and 100 basis points per unit of one-way turnover, each with a standard error from the same resamples. I compute the difference between a naive portfolio's ratio and an optimised portfolio's on the same resamples, which gives its paired standard error, its interval and its p-value, with Holm's correction across the seven at each cost. Finally, I split each month's active return into the contributions of the six factors and a residual.

3. Results

3.1 The replication

Table 1 reports the out-of-sample Sharpe ratios beside the paper's. Of the 30 Sharpe ratios tested, 29 are within 0.03 of the published values and 27 within 0.01. The Sharpe ratios of 1/N, which involves no estimation, are within 0.001 of the paper's on three of the four datasets. The one more than 0.03 away is minimum variance on FF-4-factor, 0.033 against the paper's −0.018. Noise in the returns does not account for it: adding ten basis points of noise to every return moves this Sharpe ratio over a range of 0.08, which does not reach the published value. The in-sample portfolio, built with hindsight from the whole sample, differs from the paper's by 0.06 and 0.04 on the two datasets built from the size and book-to-market portfolios, which is consistent with French having revised those portfolios since the paper. The ranking of the top three rules holds on two datasets. On the other two, the rules that swap places are 0.002 to 0.007 apart in the paper, well inside the margin of 0.03. The paper reports six differences between a constrained rule and 1/N as statistically significant at the 5 per cent level on its one-sided test, and all six are statistically significant here too.

Table 2 reports turnover, and all 26 turnover figures of the portfolios that trade are within 25 per cent of the published ones. The paper's analytical result and its simulation also reproduce: the estimation window at which mean-variance starts to do better than 1/N on average matches all 13 values the paper states, and 77 of the 81 Sharpe ratios of its simulation are within 0.03. In a simulated market with fat-tailed returns the estimated rules lose at most 0.002 in Sharpe ratio, and in one with volatility that clusters over time the comparison with 1/N does not change.

Table 1. Out-of-sample Sharpe ratios, 1973-07 to 2004-11 (377 months), this study beside DeMiguel, Garlappi and Uppal (2009, Table 3)

	Industry		*MKT/SMB/HML*		*FF-1-factor*		*FF-4-factor*	
Rule	*this study*	*DGU*	*this study*	*DGU*	*this study*	*DGU*	*this study*	*DGU*
mean-variance, in sample	0.2034[b]	0.2124	0.2608[b]	0.2851	0.4509[b]	0.5098	0.4999[b]	0.5364
1/N	0.1353	0.1353	0.2347	0.2240	0.1613	0.1623	0.1745	0.1753
mean-variance	−0.0494[b]	0.0679	0.2135	0.2186	0.0513[b]	0.0128	0.1650[b]	0.1841
Bayes-Stein	−0.0383[b]	0.0719	0.2541	0.2536	0.0620[b]	0.0138	0.1672[b]	0.1791
minimum variance	0.1537	0.1554	0.2574	0.2493	0.2496	0.2778	0.0326[a]	−0.0183
value-weighted market	0.1163	0.1138	0.1163	0.1138	0.1163	0.1138	0.1163	0.1138
mean-variance, constrained	0.0627	0.0678	0.1093	0.1084	0.1963	0.1977	0.1978	0.2024
Bayes-Stein, constrained	0.0831	0.0819	0.1480	0.1514	0.1974	0.1955	0.2064	0.2062
minimum variance, constrained	0.1445	0.1425	0.2574	0.2493	0.1554	0.1546	0.3649	0.3580
minimum variance, generalised constraint	0.1454	0.1451	0.2552	0.2467	0.1625	0.1615	0.3080	0.3028

Each figure is the Sharpe ratio of the rule's monthly out-of-sample excess returns with a 120-month estimation window. Thirty Sharpe ratios are tested, and 29 are within 0.03 of the published value. Some tested figures repeat a portfolio: the value-weighted market appears in all four datasets, and on MKT/SMB/HML minimum variance is the same with or without the ban on short sales. Without these repeats, 25 of 26 Sharpe ratios are within 0.03, and 25 of 25 turnover figures of the portfolios that trade are within 25 per cent. [a] The one tested Sharpe ratio more than 0.03 from the published value. [b] Not tested: the paper's in-sample portfolio, built with hindsight from the whole sample, and sample mean-variance and Bayes-Stein on Industry, FF-1-factor and FF-4-factor, which trade more than 100 times as much as 1/N, so that ten basis points of noise in the returns move their Sharpe ratios by 0.12 to 0.25.

Table 2. Turnover, 1973-07 to 2004-11: 1/N in absolute terms (the fraction of the portfolio traded per month, counting purchases and sales), every other rule as a multiple of 1/N, this study beside DeMiguel, Garlappi and Uppal (2009, Table 5)

	Industry		*MKT/SMB/HML*		*FF-1-factor*		*FF-4-factor*	
Rule	*this study*	*DGU*	*this study*	*DGU*	*this study*	*DGU*	*this study*	*DGU*
1/N	0.0221	0.0216	0.0237	0.0237	0.0164	0.0162	0.0200	0.0198
mean-variance	2,347[b]	606,594	2.50	2.83	10,112[b]	10,466	678.03[b]	3,553
Bayes-Stein	1,952[b]	10,621	1.52	1.85	5,516[b]	11,796	795.93[b]	3,418
minimum variance	20.99	21.65	1.02	1.11	44.95	45.47	6.16	6.83
value-weighted market	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
mean-variance, constrained	7.70	7.17	3.86	4.12	19.07	17.53	16.46	13.82
Bayes-Stein, constrained	6.63	7.22	3.45	3.65	16.11	17.32	13.65	13.07
minimum variance, constrained	2.34	2.58	1.02	1.11	4.20	3.93	1.63	1.76
minimum variance, generalised constraint	1.49	1.52	1.00	1.09	1.95	1.78	1.66	1.70

The table holds 30 turnover figures. The 26 of the portfolios that trade are tested, and all 26 are within 25 per cent of the published value. The four of the value-weighted market are zero, since it trades nothing by construction. [b] Not tested: the paper's figure exceeds 100 times that of 1/N, and the figure moves by orders of magnitude when a single month's covariance matrix is close to singular, i.e., when its inverse multiplies small errors by thousands.

3.2 The months since the paper

Table 3 compares the paper's period with the 261 months from December 2004 to August 2026. In the new months no optimising rule's Sharpe ratio exceeds 1/N's by a margin that is statistically significant at the 5 per cent level on the two-sided Jobson-Korkie test, on any dataset. To be statistically significant, a margin would have to be about 0.02 to 0.18, depending on the dataset and the rule. One Sharpe ratio is above 1/N's at that level, the value-weighted market's on MKT/SMB/HML, which is not an optimising rule, and five are below it, all of them minimum-variance rules on the two factor datasets. Minimum variance, whose Sharpe ratio was above 1/N's on three of the four datasets in the paper's period, falls to −0.015 on MKT/SMB/HML and −0.034 on FF-4-factor, against 0.113 and 0.148 for 1/N. Its weights load on the size and value factors, whose returns were small or negative over these years, which is consistent with its fall.

Table 3. Out-of-sample Sharpe ratios on the paper's period and on the months since, on the 202608 vintage of French's files

	1973-07 to 2004-11 (377 months)				*2004-12 to 2026-08 (261 months)*			
Rule	*Industry*	*MKT/SMB/HML*	*FF-1-factor*	*FF-4-factor*	*Industry*	*MKT/SMB/HML*	*FF-1-factor*	*FF-4-factor*
1/N	0.135	0.235	0.161	0.174	0.191	0.113	0.148	0.148
mean-variance	−0.049	0.213	0.051	0.165	0.042	0.176	0.176	0.102
Bayes-Stein	−0.038	0.254	0.062	0.167	0.053	0.158	0.208	0.101
minimum variance	0.154	0.257	0.250	0.033	0.168	−0.015	0.218	−0.034
value-weighted market	0.116	0.116	0.116	0.116	0.191	0.191	0.191	0.191
mean-variance, constrained	0.063	0.109	0.196	0.198	0.174	0.189	0.151	0.143
Bayes-Stein, constrained	0.083	0.148	0.197	0.206	0.190	0.188	0.159	0.146
minimum variance, constrained	0.144	0.257	0.155	0.365	0.210	−0.009	0.181	0.038
minimum variance, generalised constraint	0.145	0.255	0.163	0.308	0.210	0.042	0.165	0.149

Each figure is the out-of-sample Sharpe ratio with a 120-month estimation window. In the new period, on the two-sided Jobson-Korkie test with Memmel's correction against 1/N, one Sharpe ratio is above 1/N's at the 5 per cent level, the value-weighted market's on MKT/SMB/HML, and five are below it, all minimum-variance rules on the two factor datasets.

3.3 Under the tilt alone: the naive tilt and the optimised portfolios

Table 4 sets the naive tilt beside the seven optimised portfolios. The naive tilt realises 105.6 basis points a year of tracking error and the optimised portfolios 79.7 to 90.9, so the naive tilt realises 14.7 to 25.9 basis points a year more than each of them. Each of the seven differences is statistically significant at the 5 per cent level on the paired variance test after Holm's correction, with adjusted p-values of 0.0035 to 0.0115. The smallest difference is against the optimised portfolio built with the sample covariance of 120 monthly returns, 14.7 basis points with a 95 per cent interval of 6.9 to 22.4. The largest is against the one built with RiskMetrics, 25.9 with an interval of 13.0 to 38.9. The monthly active returns of the naive tilt and of each optimised portfolio are correlated at 0.76 to 0.86.

The attribution indicates where the optimised portfolios' lower tracking error comes from. Removing weight from Coal, Oil and Utilities and spreading it over the other 46 industries lowers the portfolio's exposures to value, profitability, investment and momentum. The naive tilt's four active exposures to these factors sum to −0.054, against −0.022 to −0.035 for the optimised portfolios. Their total closer to zero indicates that the optimiser offsets part of what the tilt removes, consistent with its holding more of the industries whose returns move with those of the three, which the naive tilt by construction does not do.

The naive tilt trades less. Its mean one-way turnover is 0.42 per cent a month against 0.81 to 1.50 for the optimised portfolios. It moves 11.7 per cent of the portfolio in its first month, when it leaves the benchmark. It exceeds the turnover limit of 2 per cent a month in 28 of the 566 months, while the optimised portfolios exceed it only in the 2 or 3 months in which no portfolio could meet the limit together with the other rules. The naive tilt's realised beta to the benchmark is 1.017, against 1.007 to 1.010 for the optimised portfolios.

Table 4. The naive tilt and the seven optimised portfolios under the tilt alone, July 1979 to August 2026 (566 months)

Covariance estimate	*Realised tracking error, bp a year*	*Naive tilt minus optimised portfolio, bp a year*	*95% interval*	*Holm p-value*	*Correlation*	*One-way turnover, % a month*	*Active return, bp a year*	*Net of 100 bp per unit of turnover, bp a year*
Sample covariance, 3 years daily	80.3	+25.3	+12.6 to +38.1	0.0035	0.81	1.26	+1.8	−13.4
Ledoit-Wolf shrinkage, 3 years daily	80.3	+25.3	+12.6 to +38.0	0.0035	0.82	1.18	+2.8	−11.4
Six-factor model, 3 years daily	88.3	+17.3	+5.5 to +29.1	0.0060	0.86	0.81	+6.9	−2.8
Principal components, 3 years daily	86.7	+18.9	+11.5 to +26.3	0.0035	0.84	0.89	+3.9	−6.8
Sample covariance, 120 months	90.9	+14.7	+6.9 to +22.4	0.0115	0.76	1.50	+6.6	−11.4
Robust (built on the five above)	82.3	+23.3	+17.4 to +29.2	0.0035	0.84	0.99	+5.6	−6.3
RiskMetrics (exponentially weighted), 5 years daily	79.7	+25.9	+13.0 to +38.9	0.0035	0.81	1.39	+2.7	−14.0
Naive tilt	105.6					0.42	+13.5	+8.5

Each optimised portfolio comes from the companion study under the tilt alone, built with one of six covariance estimators or as the robust portfolio that uses five of them. The difference is the naive tilt's realised tracking error minus the optimised portfolio's, with its 95 per cent interval from the paired variance test of Ledoit and Wolf (2011) in its bootstrap form, blocks of 12 months and 2,000 resamples. The Holm p-value is corrected for the seven comparisons. The correlation is that of the two monthly active returns. Active returns are simulated and annualised, gross and net of 100 basis points per unit of one-way turnover. Source: Kenneth R. French Data Library and author's calculations. Monthly, July 1979 to August 2026, 49 US industries, USD. results/202608/naive_tilt_paired_tests_202608.csv and naive_tilt_returns_and_costs_202608.csv, notebook 08.

3.4 Under the factor objective: the tilted index and the optimised portfolios

Table 5 sets the tilted index beside the seven optimised portfolios at the budget of 100 basis points. The tilted index has a mean forecast exposure of 0.022 at a realised tracking error of 119.3 basis points a year, a forecast exposure per basis point of 0.000185. The optimised portfolios have a mean forecast exposure of 0.095 to 0.140 at realised tracking errors of 115.6 to 150.1 basis points a year, so 0.000824 to 0.000979 per basis point, which is 4.5 to 5.3 times the tilted index's. Each of the seven differences is statistically significant at the 5 per cent level on the paired block bootstrap after Holm's correction, with an adjusted p-value of 0.007 for each. The tilted index's tracking error lies inside the optimised portfolios' range, so the comparison is made at a tracking error of the same size. At budgets of 75 and 150 basis points, which the companion study ran only with the five estimators other than RiskMetrics, the optimised portfolios' forecast exposure per basis point is 0.000478 to 0.000633 and 0.001015 to 0.001073, both above the tilted index's.

The forecast exposure is the quantity the optimiser makes as large as it can, and it is computed with the betas the optimiser itself estimates, so it carries their estimation error. Measured on the returns the portfolios realised, the optimised portfolios realise 0.40 to 0.56 of their forecast exposure and the tilted index 1.9 times its own. On realised exposure per basis point the optimised portfolios obtain 1.06 to 1.48 times the tilted index's 0.000351, and no difference is statistically significant at the 5 per cent level after Holm's correction, with adjusted p-values of 1.000. The 95 per cent intervals of the differences run from realised multiples of about 0.4 to 0.8 up to about 2.2 to 2.7, so they are too wide to rule out either a multiple below 1 or one above 2.

The tilted index's forecast exposure is small partly because of how it is built. One multiple k scales all 44 industries, and the change in an industry's active weight is its weight times k times its centred score, so the largest industries reach the bound first and set k for all the others. The bound sets k in 549 of the 566 months, most often at Banks (184 months) and Computer Software (138), and the floor at zero sets it in the other 17, all at Gold, from April 1981 to January 1984. The median of k over the months is 0.206. The smaller industries therefore use only part of the room the bound gives them.

The tilted index trades less than the optimised portfolios, 0.77 per cent of the portfolio a month against 1.78 to 2.00. It moves 13.9 per cent of the portfolio in its first month and exceeds the turnover limit in 38 of the 566 months. Its realised beta is 1.004, against 0.999 to 1.012 for the optimised portfolios.

Table 5. The tilted index and the seven optimised portfolios under the factor objective, budget 100 basis points a year, July 1979 to August 2026 (566 months)

Covariance estimate	*Realised tracking error, bp a year*	*Mean forecast exposure*	*Forecast exposure per bp*	*Tilted index minus optimised portfolio*	*95% interval*	*Holm p-value*	*Realised exposure per bp*	*Holm p-value, realised*	*One-way turnover, % a month*	*Active return, bp a year*	*Net of 100 bp per unit of turnover, bp a year*
Sample covariance, 3 years daily	120.6	0.116	0.000960	−0.000775	−0.001006 to −0.000570	0.007	0.000415	1.000	1.92	+8.9	−14.1
Ledoit-Wolf shrinkage, 3 years daily	118.6	0.109	0.000919	−0.000734	−0.000958 to −0.000529	0.007	0.000413	1.000	1.88	+6.5	−16.1
Six-factor model, 3 years daily	133.6	0.123	0.000919	−0.000734	−0.000957 to −0.000529	0.007	0.000417	1.000	1.78	+15.5	−5.9
Principal components, 3 years daily	137.0	0.128	0.000937	−0.000752	−0.000969 to −0.000564	0.007	0.000373	1.000	1.87	+9.9	−12.5
Sample covariance, 120 months	150.1	0.140	0.000933	−0.000748	−0.000946 to −0.000570	0.007	0.000521	1.000	2.00	+6.0	−18.0
Robust (built on the five above)	115.6	0.095	0.000824	−0.000639	−0.000891 to −0.000411	0.007	0.000382	1.000	1.86	+2.3	−20.0
RiskMetrics (exponentially weighted), 5 years daily	118.6	0.116	0.000979	−0.000794	−0.001028 to −0.000586	0.007	0.000400	1.000	1.96	+7.5	−16.1
Tilted index	119.3	0.022	0.000185				0.000351		0.77	+16.4	+7.2

Each optimised portfolio comes from the companion study under the factor objective, built with one of six covariance estimators or as the robust portfolio that uses five of them. The forecast exposure is the factor objective, the sum of the active exposures to value, profitability, investment and momentum computed with the betas of the three years before each month, and its mean is over the 566 months. The forecast exposure per basis point is the forecast exposure divided by the realised tracking error in basis points. The difference and its 95 per cent interval come from 2,000 circular block bootstrap resamples of the months in blocks of 12, the same months drawn for both portfolios, and the p-value is two-sided. The Holm p-value is corrected for the seven comparisons. The realised exposure is the sum of the slopes on the four factors in a regression of the 566 monthly active returns on the six factors' monthly returns, given per basis point of realised tracking error, with the Holm p-value of the tilted index minus the optimised portfolio from the same resamples. Active returns are simulated and annualised. Source: Kenneth R. French Data Library and author's calculations. Monthly, July 1979 to August 2026, 49 US industries, USD. results/202608/tilted_index_bootstrap_202608.csv, tilted_index_beside_paths_202608.csv and tilted_index_realised_exposure_202608.csv, notebook 09.

3.5 The active returns

Both naive portfolios earned more active return over the sample than the optimised portfolios they are set beside, before and after costs. Before costs, the naive tilt earned +13.5 basis points a year against +1.8 to +6.9 for the optimised portfolios under the tilt alone, and the tilted index +16.4 against +2.3 to +15.5 under the factor objective. Net of 100 basis points per unit of one-way turnover, the naive tilt earned +8.5 against −14.0 to −2.8, and the tilted index +7.2 against −20.0 to −5.9. I compare the information ratios on the same resamples, with a paired test. Before costs no difference is statistically significant at the 5 per cent level: the differences are 0.05 to 0.11 under the tilt alone, with paired standard errors of 0.08 to 0.11, and 0.02 to 0.12 under the factor objective, with 0.12 to 0.15. Net of 100 basis points under the tilt alone, four of the seven differences are statistically significant when each is tested on its own, and none after Holm's correction. Under the factor objective none is, at any cost. These figures therefore do not show that a naive portfolio earns more than an optimised portfolio, or that it earns less. The gap in active return has two parts.

The first part is a property of the period. I write the naive tilt's active return, month by month, as the weight removed from Coal, Oil and Utilities times the spread between the cap-weighted return of the other 46 industries and that of the three. Over the 566 months the spread averaged +1.22 per cent a year, with a standard deviation of 15.28 per cent a year and a t-ratio of 0.55, taking the months as independent, so its mean is not statistically significant at the 5 per cent level. The optimised portfolios hedge part of that spread away, which is also why their tracking error is lower. In the

attribution under the tilt alone, the residual, the part of the active return the six factors do not explain, is +28.9 basis points a year for the naive tilt against +12.2 to +19.7 for the optimised portfolios.

The second part is a property of the two methods. After costs the gaps widen, because each naive portfolio trades less than every optimised portfolio it is set beside.

Under the factor objective the attribution shows where the optimised portfolios' active return came from. The six factors contributed +30.7 to +47.3 basis points a year, the four targeted factors +21.7 to +35.5 of it, and the residual −28.4 to −37.4, which cancelled most of the factors' contribution. For the tilted index the six factors contributed +17.9, the targeted factors +12.4, and the residual −1.4. The optimised portfolios' residual of −28.4 to −37.4 basis points a year, against the tilted index's −1.4, is consistent with a forecast exposure above the realised one in months when the targeted factors earned a premium.

Table 6. Information ratios and the attribution of the active return, July 1979 to August 2026 (566 months)

Portfolio	*Information ratio, gross (s.e.)*	*Information ratio net of 100 bp (s.e.)*	*Factor contributions, bp a year (s.e.)*	*Of which the four targeted factors, bp a year*	*Residual, bp a year (s.e.)*	*Active return, bp a year*
Under the tilt alone						
Naive tilt	0.128 (0.139)	0.081 (0.138)	−15.3 (9.7)	−19.4	+28.9 (12.4)	+13.5
Sample covariance, 3 years daily	0.022 (0.131)	−0.166 (0.130)	−10.5 (5.1)	−11.7	+12.2 (10.3)	+1.8
Ledoit-Wolf shrinkage, 3 years daily	0.034 (0.130)	−0.142 (0.129)	−10.4 (5.2)	−11.7	+13.2 (10.2)	+2.8
Six-factor model, 3 years daily	0.078 (0.140)	−0.032 (0.138)	−9.3 (5.0)	−10.2	+16.2 (12.0)	+6.9
Principal components, 3 years daily	0.045 (0.134)	−0.078 (0.130)	−12.4 (6.0)	−13.5	+16.3 (11.1)	+3.9
Sample covariance, 120 months	0.073 (0.121)	−0.125 (0.116)	−13.1 (6.3)	−11.6	+19.7 (11.2)	+6.6
Robust (built on the five above)	0.069 (0.125)	−0.076 (0.120)	−11.1 (5.6)	−11.3	+16.8 (10.2)	+5.6
RiskMetrics (exponentially weighted), 5 years daily	0.034 (0.131)	−0.175 (0.132)	−10.7 (5.1)	−11.6	+13.4 (10.3)	+2.7
Under the factor objective, budget 100 bp a year						
Tilted index	0.138 (0.146)	0.060 (0.145)	+17.9 (11.3)	+12.4	−1.4 (14.8)	+16.4
Sample covariance, 3 years daily	0.074 (0.159)	−0.117 (0.158)	+41.6 (10.8)	+29.1	−32.7 (15.9)	+8.9
Ledoit-Wolf shrinkage, 3 years daily	0.055 (0.155)	−0.136 (0.154)	+39.2 (10.7)	+26.7	−32.8 (15.6)	+6.5
Six-factor model, 3 years daily	0.116 (0.154)	−0.044 (0.155)	+47.3 (11.4)	+33.3	−31.8 (18.0)	+15.5
Principal components, 3 years daily	0.073 (0.155)	−0.091 (0.155)	+47.3 (12.1)	+35.5	−37.4 (18.0)	+9.9
Sample covariance, 120 months	0.040 (0.159)	−0.120 (0.156)	+36.7 (15.3)	+34.8	−30.7 (19.1)	+6.0
Robust (built on the five above)	0.020 (0.153)	−0.173 (0.150)	+30.7 (10.7)	+21.7	−28.4 (15.1)	+2.3
RiskMetrics (exponentially weighted), 5 years daily	0.063 (0.158)	−0.135 (0.158)	+38.1 (10.5)	+28.2	−30.6 (15.9)	+7.5

The information ratio is the active return per year over the realised tracking error, before costs and net of 100 basis points per unit of one-way turnover, with its standard error from 2,000 block bootstrap resamples of the months, the cost held at its full-sample figure. A contribution is the active exposure to a factor times the factor's return in the month, and the residual is the active return minus the six contributions. The standard errors of the contributions and the residual take the months as independent. Under the tilt alone, the targeted factors' contribution of each optimised portfolio is the sum of its four contributions. Source: Kenneth R. French Data Library and author's calculations. Monthly, July 1979 to August 2026, 49 US industries, USD. results/202608/naive_tilt_information_ratios_202608.csv, naive_tilt_attribution_summary_202608.csv, tilted_index_information_ratios_202608.csv and tilted_index_attribution_summary_202608.csv, notebooks 08 and 09; for the optimised portfolios, results/202608/source/attribution_summary_202608.csv and factor_objective_attribution_202608.csv.

4. Discussion

The natural comparison is with DeMiguel, Garlappi and Uppal (2009). On their own datasets the replication indicates that their tables reproduce on a 2026 version of French's data, with one Sharpe ratio more than 0.03 away, on data French appears to have revised since. Their conclusion also holds on the 261 months since: no optimising rule's Sharpe ratio exceeds 1/N's by a margin that is statistically significant at the 5 per cent level, and minimum variance, whose

Sharpe ratio was above 1/N's on three of the four datasets in their period, fell as the size and value factors it loaded on earned little or lost. Under an index mandate, the optimiser realises less tracking error than the naive tilt, and its forecast exposure per basis point is larger than the tilted index's, while the difference in the exposure its returns realised is not statistically significant. Most of their optimisers needed estimated means as well as covariances, and all of them had the whole portfolio to place. Under the tilt alone the optimiser needs only a covariance estimate, and only to place the share of the portfolio the mandate moves away from the benchmark, and its lower tracking error plausibly comes from that narrower task. Under the factor objective it also maximises estimated betas, and the shortfall of its realised exposure is plausibly the cost of their estimation error, the kind of cost DeMiguel, Garlappi and Uppal found.

For the holder of such a mandate, the estimates therefore suggest two things. First, under the tilt alone the optimiser with a risk model realises 15 to 26 basis points a year less tracking error than the naive tilt, 14 to 25 per cent of the naive tilt's 105.6. Under the factor objective, its forecast exposure per basis point is 4.5 to 5.3 times the tilted index's. Its realised exposure per basis point is 1.1 to 1.5 times the tilted index's, and no difference in it is statistically significant, with intervals from about 0.4 to about 2.7 times. Second, the naive tilt's higher active return over this sample rests, before costs, on a spread whose mean is small against its variation, and both naive portfolios' advantage in active return widens after costs because they trade less. On a paired test, no difference in information ratio is statistically significant at the 5 per cent level after Holm's correction.

At the same time, the design has limits. First, each naive portfolio is one construction. The tilted index's multiple k is set by its largest industries, so its forecast exposure depends on that construction, and the comparison under the factor objective is with this tilted index and not with every naive portfolio that could be built. Second, 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 a risk model is worth as much there is a question these data cannot answer. Third, the comparison of returns rests on one sample, since the spread behind the naive tilt's return has a t-ratio of 0.55 over 47 years. 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 from 1979 to 2026, an optimiser with a covariance estimate realises 15 to 26 basis points a year less tracking error than the naive tilt under the tilt alone, and has 4.5 to 5.3 times the forecast factor exposure per basis point of the tilted index under the factor objective, each difference statistically significant at the 5 per cent level after Holm's correction. On the realised exposure, the multiple is 1.1 to 1.5 and no difference is statistically significant, with intervals from about 0.4 to about 2.7 times. No difference in information ratio between a naive portfolio and an optimised portfolio is statistically significant at the 5 per cent level on a paired test after Holm's correction.

References

Carhart, M. M. (1997). On persistence in mutual fund performance. *Journal of Finance*, 52(1), 57-82.

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), 1915-1953.

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.

Fama, E. F. and French, K. R. (2015). A five-factor asset pricing model. *Journal of Financial Economics*, 116(1), 1-22.

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.

Hofer, C. (2026). Which risk model for an index mandate? Covariance estimators and tracking error on 49 US industries. Code and notebooks: github.com/ochofer/risk-models-for-index-tracking.

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

Jagannathan, R. and Ma, T. (2003). Risk reduction in large portfolios: why imposing the wrong constraints helps. *Journal of Finance*, 58(4), 1651-1683.

Jobson, J. D. and Korkie, B. M. (1981). Performance hypothesis testing with the Sharpe and Treynor measures. *Journal of Finance*, 36(4), 889-908.

Jorion, P. (1986). Bayes-Stein estimation for portfolio analysis. *Journal of Financial and Quantitative Analysis*, 21(3), 279-292.

Ledoit, O. and Wolf, M. (2004). A well-conditioned estimator for large-dimensional covariance matrices. *Journal of Multivariate Analysis*, 88(2), 365-411.

Ledoit, O. and Wolf, M. (2011). Robust performance hypothesis testing with the variance. *Wilmott*, 2011(55), 86-89.

Memmel, C. (2003). Performance hypothesis testing with the Sharpe ratio. *Finance Letters*, 1(1), 21-23.

Michaud, R. O. (1989). The Markowitz optimization enigma: is 'optimized' optimal? *Financial Analysts Journal*, 45(1), 31-42.

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