The Equally Weighted Portfolio Still Remains a Challenging Benchmark

Matteo Gelmini[1], Pierpaolo Uberti[1,2]

Abstract

This research replicates the paper "Optimal Versus Naive Diversification: How Inefficient is the 1/N Portfolio Strategy?", DeMiguel et al. (2009b). Similar to the referring paper, working in the mean-variance context, we compare the out-of-sample performance of the same investment strategies on the basis of standard metrics (Sharpe ratio, certainty equivalent and turnover). We consider proportional transaction costs and estimation rolling windows of limited length. Our study updates the original paper for many interesting aspects. First, to exclude that the empirical evidence of DeMiguel et al. (2009b), whose data stopped in 2004, could depend on very specific market behavior, we use an updated version of the original databases that contains the returns of the last 20 years. Recent data are characterized by a few severe systemic events, the 2008 global financial crisis and the shock related to the pandemic, and a generally higher level of price volatility than the previous periods. In our opinion, this variation in the market's conditions makes the replication very interesting. Second, we introduce the Equally Risk Contribution (ERC) portfolio within the allocation strategies under comparison. This allocation rule is strictly related to the mean-variance approach when the variance is used as the referring risk measure and it constitutes a very interesting alternative investment benchmark. Moreover, using real data, we study whether a variation of the holding period or the length of the estimation window can modify the performance of all the strategies under comparison. Our findings confirm the results of DeMiguel et al. (2009b), i.e. that the equally weighted portfolio still remains a challenging benchmark

Email addresses: m.gelmini1@campus.unimib.it (Matteo Gelmini), pierpaolo.uberti@unimib.it (Pierpaolo Uberti)

[1]Department of Statistics and Quantitative Methods, University of Milano-Bicocca

[2]Corresponding author

to beat. Nevertheless, we find a few significant differences: the number of strategies that outperform naive diversification is larger due to the increased market volatility; limiting the impact of transaction costs by investing in a portfolio with a stable allocation as the ERC, or modifying the lengths of the estimation window and the holding period, is not sufficient to beat naive diversification systematically.

Keywords: Replication Study, Portfolio Choice, Investment Decisions, Naive Diversification, Out-of-sample Performance
JEL: G10, G11

---

1. Introduction

As the scientific literature grows, replication studies are necessary to confirm or critically discuss the results of previously published papers. This is especially the case for real data empirical analysis.

This article replicates "Optimal Versus Naive Diversification: How Inefficient is the 1/N Portfolio Strategy?", DeMiguel et al. (2009b). This paper was published in 2009 in the journal *Review on Financial Studies* and, at the moment we are writing, it counts approximately 4100 citations on Google Scholar. In well-defined experimental settings, the study shows that the equally weighted portfolio is a challenging benchmark to beat for standard mean-variance optimization-based investment strategies. The paper indirectly raised a deep and shocking question: what is the utility of complex theoretical optimization models that are difficult to formalize and to apply if, in practice, they cannot beat the basic equally weighted portfolio systematically? The publication of this paper opened a huge debate in the scientific community. On the one hand, some researchers focused their attention on the theoretical reasons that can explain the empirical evidence, see Malladi and Fabozzi (2017). Model uncertainty is considered the principal cause of the poor out-of-sample performance of the optimization-based approaches, see Pflug et al. (2012). One further stream of research identifies the numerical instability of the Markowitz model as the principal cause of its poor out-of-sample performance, see Hirschberger et al. (2010) and Best and Hlouskova (2008). On the opposite, many other researchers focused on the attempt to find non-trivial portfolio allocation strategies able to beat the equally weighted portfolio in the out-of-sample framework to justify the proposal and the use of complex decision-making optimization-based models, see Kritzman et al. (2010), or to beat a challenging benchmark, see among the others Kirby and Ostdiek (2012), Ackermann et al. (2017), Bessler et al. (2017), Fugazza et al. (2015), Hanke et al. (2019), Jiang et al. (2019), Yuan and Zhou (2022). The debate is still active and far to be concluded.

The well-known paper by DeMiguel et al. (2009b) compares the out-of-sample performance of optimization-based approaches with the so-called *naive diversification*, which allocates the same proportion of wealth in each asset. In that context, the equally weighted portfolio is the benchmark. The out-of-sample empirical results show the difficulty to beat the equally weighted portfolio in terms of Sharpe ratio, see Sharpe (1966), certainty equivalent and turnover, in the mean-variance context using optimization-based approaches, when estimation windows of limited length and proportional transaction costs are considered.

We first test the robustness of the findings of DeMiguel et al. (2009b) by expanding the original databases to the last 20 years. We compare the same strategies, maintaining all the settings of the original experiment, i.e. the same length for the rolling windows, the same values of the parameters, the same proportional structure and amount of the transaction costs and the same metrics to evaluate the performance of the investment strategies. This first experiment aims to verify if the findings of DeMiguel et al. (2009b) are still valid. More recent data significantly differ from the ones used in the original research for the presence of a few severe systemic crises, the 2008 global financial crisis and the shock related to the COVID-19 pandemic, and a general increased level of volatility. Moreover, we add the equally risk contribution (ERC) allocation strategy, see Maillard et al. (2010), to the investment strategies under comparison. The ERC belongs to the mean-variance framework when the variance is used as the referring risk measure. It has been proposed to overcome some of the principal critical issues of the Markowitz model: optimal long-only mean-variance portfolios concentrate the allocation in a few assets and show a numerically unstable composition for the uncertainty of the parameters. Moreover, the ERC identifies a portfolio with an in-sample variance that is bounded between the variance of the global minimum variance portfolio and the variance of the equally weighted portfolio. In a further experiment, we use estimation windows whose length grows with the length of the time series, considering all the available data to estimate the parameters. In DeMiguel et al. (2009b) this experiment was performed only on simulated data to test the impact of the length of the estimation window on the out-of-sample returns of the strategies. In the last experiment we increased the holding period from 1 month to 12 months, following the idea proposed by Chan et al. (1999) and Jagannathan and Ma (2003); this last exercise is totally new.

2. Empirical Experiment

In this section, we report our experiment. First we describe the experiment and introduce the notation. Then we enumerate the datasets, the allocation strategies and the metrics used for the comparison. We do not provide a detailed mathematical formalization of the single investment strategies and the metrics used for their evaluation; we refer to the original papers through bibliographic references.

2.1. Description of the experiment and notation

The referring experiment is performed in a standard rolling window framework. The length of the estimation window is called W while N is the number of assets. The first $W \times N$ entries of the matrix of returns are used to estimate the optimal allocation for one model. Then, given the length of the holding period H, the performance of the strategy is computed considering the returns from $W + 1$ to $W + H$. To maintain the length W of the estimation, the described calculation is repeated after shifting the estimation window H periods ahead, thus containing the returns from $H+1$ to $W+H$. The turnover (TO) of one strategy is computed as the absolute difference between the allocation at the end of one holding period and the optimal allocation at the beginning of the subsequent period. Let us underline that the allocation at the end of one holding period is different from the optimal allocation at the beginning of the same period due to the variation of the prices. This explains why the equally weighted portfolio has a small but positive turnover. The transaction costs are finally calculated proportionally to the TO; the coefficient is fifty basis points as in DeMiguel et al. (2009b). The effective returns of a strategy are then obtained subtracting the transaction costs for its implementation from the gross return. This procedure is repeated for each database and strategy under comparison, see section 2.2.

The strict replication exercise is limited to the first two experiments performed for $W = 120$ and $H = 1$, see section 2.3 and, $W = 60$ and $H = 1$, see section 2.4. The parameter W plays a fundamental role in the application. On one hand, a short estimation window considers only the most recent data, which are usually the most important to obtain a good predictive performance of the model; it is then intuitively useless to estimate the parameters based on the old data that do not reflect actual market conditions. Conversely, in the mean-variance framework, the covariance must be estimated using adequate length time series. The length of the estimation windows primarily depends on the portfolio's size N; a sufficient length of the time series is needed to obtain a meaningful, in terms of statistical content, estimation of the entries of the covariance matrix. The minimal technical requirement is $W \geq N$, which ensures the covariance matrix is a full-rank invertible matrix. The literature on the topic is wide; we cite, among the others, Bickel and Levina (2008) and Ledoit and Wolf (2003). $W = 60$ is a relatively short window strongly restricting the size of the portfolios we can build. Nevertheless, with monthly observations, $W = 60$ is equivalent to considering the data of the last 5 years, thus using relatively old information for the estimation of the covariance matrix. One last consideration regarding parameter W is that it strongly affects the stability of the allocation, with direct consequences on the turnover. When W is short, one observation significantly impacts the estimation of the covariance matrix. Then the optimal allocation significantly changes from one investing period to the next one. In the original paper, this aspect was studied through an exercise based on simulated data. It showed that the optimization-based strategies become stable and very competitive if the estimation windows are long enough. In section 2.5, we will investigate this aspect in more detail with a dedicated real data experiment.

The last two experiments, see sections 2.5 and 2.6, are totally new. They show that the poor performance of the optimization-based active investment strategies does not depend exclusively on the on the impact of the transaction costs. Indeed, a longer estimation windows reduce the turnover and the costs of the active strategies but the improvement is not enough to identify one strategy that systematically outperforms the $1/N$.

2.2. Data, strategies, metrics

The datasets used to implement the present replication experiment are the same as those used in the original paper, except that they have been updated with recent data. In our opinion, one ex-post major critical aspect of the original research is that it was performed on data with a significantly lower level of volatility for the data of the last 20 years. The principal motivation of this research is to test the robustness of the findings of the original paper, as the most recent data are characterized by a significantly higher volatility. Table 1 provides the enumeration and a brief qualitative description of the databases used in the application.

Table 1: Datasets used in the application.

Name	N	Time Period	Frequency	Description
S&P Sectors “S&P”	11	07/2002 - 07/2023	Monthly	Ten sector portfolios of the S&P 500 and the US equity market portfolio
Industry “Ind”	11	07/1982 - 06/2023	Monthly	10 Ten industry portfolios and the US equity market portfolio
International “Int”	9	04/1992 - 04/2023	Monthly	Eight country indexes and the World Index
MKT/SMB/HML “MSH”	3	06/1982 - 06/2023	Monthly	SMB and HML portfolios and the US equity market portfolio
FF 1-factor “FF1”	21	06/1982 - 06/2023	Monthly	Twenty size- and book-to-market portfolios and the US equity MKT
FF 4-factor “FF4”	24	06/1982 - 06/2023	Monthly	Twenty size- and book-to-market portfolios and the MKT, SMB, HML and UMD portfolios

For a detailed description of the database’s assets see DeMiguel et al. (2009b). Except for the data in “S&P” and in “Int” that have been downloaded from FactSet, the other databases under analysis can be downloaded for free at the following website:
$http : //mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html$.

We have re-implemented the totality of the investing strategies that were compared in the original paper. We added the equally risk contribution (ERC) within the strategies under comparison, see Maillard et al. (2010). This strategy was not considered in the referring paper probably because it was proposed in the literature almost simultaneously. Even if it is not strictly an optimization-based strategy in the mean-variance framework, i.e. when the variance is used as the referring risk measure, is of great interest for the present experiment. It is characterized by a variance that is bounded between the ones of the minimum variance portfolio and of the equally weighted portfolio. We do not describe the allocation strategies in detail, but we simply enumerate, by referring to the original papers for the respective formalizations.

- “1/N”, with re-balancing (benchmark strategy).

- "BS", Bayes-Stein Model, see James and Stein (1992) and Stein (1956).
- "DM", Data-and-Model Approach[3], see Pástor (2000) and Pástor and Stambaugh (2000).
- "MIN", Minimum-variance (this is the global minimum variance portfolio), see Markowitz (1952).
- "VW", The value-weighted market portfolio[4] with respect to CAPM, see Sharpe (1964), Lintner (1965), Mossin (1966).
- "MP", Portfolio implied by a asset pricing models with missing (unobservable) factors, see Craig MacKinlay and Pástor (2000).
- "MV-C", mean-variance portfolio with restrictions on short selling, see Jagannathan and Ma (2003).
- "BS-C", Bayes-Stein portfolio with restrictions on short selling, see James and Stein (1992) and Stein (1956).
- "MIN-C", global minimum-variance portfolio with restrictions on short selling, see Markowitz (1952).
- "G-MIN-C", minimum-variance portfolio with generalized restrictions, see DeMiguel et al. (2009a).
- "ERC-C", equally risk contribution portfolio with restrictions on short selling, see Maillard et al. (2010) and Roncalli and Weisang (2016).
- "MV-MIN", three funds model, see Kan and Zhou (2007).
- "EW-MIN", mixture of minimum variance and 1/N, see Kan and Zhou (2007) and Garlappi et al. (2007).

[3] In this strategy the investment portfolio depends on one parameter (σ_α that we set equal to 1% as in DeMiguel et al. (2009b)).

[4] This strategy requires to buy the market portfolio at the beginning of the investment period and hold it. This strategy is strictly passive and it is characterized by a null turnover by construction.

As in DeMiguel et al. (2009b), we compare the out-of-sample performance of the investment strategies based on three main metrics: Sharpe ratio (SR), certainty equivalent (CE), turnover(TO). The SR, see Sharpe (1966), is the well-known risk adjusted performance measure. It is defined as the ratio between the excess return with respect to the risk-free rate and the standard deviation of the returns. The CE, see DeMiguel et al. (2009b), is a real-valued function of the returns and their variance that depends on one parameter[5]. Both the SR and CE are calculated on the net returns obtained subtracting the transaction costs from the gross returns as described in section 2.1.

2.3. The referring case: $W = 120$, $H = 1$

Tables 2, 3 and 4 in this section collect the Sharpe ratios, the certainty equivalent, and the turnover of the strategies under comparison when the estimation window is $W = 120$ and the holding period is $H = 1$. This experiment can be considered the referring one and will be used as the benchmark for the comparative comments when the parameters vary. Tables 2 and 3 also report the p-values (the numbers within brackets) of the statistical test, see Jobson and Korkie (1981), which permits the discussion of whether the performance measure relative to one strategy is statistically significantly different from the one of 1/N. To make the tables more immediate to interpret at first sight, we highlight in bold the cases in which one strategy outperforms the benchmark for a given database, denoting with *, ** and *** respectively, the significance levels 10%, 5% and 1%. We do not differentiate the remaining cases, i.e. when the benchmark and the single strategy do not show statistically significant differences concerning one metric and when the single strategy significantly under-performs the benchmark.

Globally looking at tables 2, 3 and 4, there is no bold line in the tables, so there is no investment strategy that significantly outperforms the benchmark for each database and performance metric.

If we focus on the Sharpe ratio, see Table 2, the strategy that more often significantly beats the 1/N is the ERC. This fact is not surprising if we consider that the ERC is an intuitive generalization of the 1/N strategy, since it provides an equally weighted allocation for the risk undertaken. Moreover, we underline how the results of the 1/N and the ERC strategies are very similar; this could depend on the fact that, in this experiment, the asset

[5] As in DeMiguel et al. (2009b) we set the value of this parameter equal to 1.

Table 2: Sharpe ratios of the out-of-sample returns, $W = 120$, $H = 1$.

Strategy	S&P	Ind	Int	MSH	FF1	FF4
1/N	0.2186	0.2205	0.0924	0.1517	0.1893	0.1942
MV (in sample)	0.4857	0.5472	0.2728	0.1707	0.4788	0.6945
MV	0.1401	0.0647	0.0443	0.1244	0.2632	**0.4215*** **
	(0.179)	(0.017)	(0.290)	(0.357)	(0.141)	(0.001)
BS	0.1604	0.0663	0.0573	0.1351	**0.2931***	**0.4578*** **
	(0.223)	(0.018)	(0.336)	(0.412)	(0.059)	(0.000)
DM	0.2359	**0.3368*** **	0.0893	-0.0443	0.2106	-0.0093
	(0.159)	(0.007)	(0.481)	(0.003)	(0.327)	(0.002)
MIN	0.1837	0.2135	**0.1484***	0.0672	**0.2993*** **	-0.3344
	(0.296)	(0.434)	(0.079)	(0.001)	(0.009)	(0.000)
VW	**0.2426***	0.1619	**0.1139***	0.1619	0.1619	0.1619
	(0.069)	(0.000)	(0.064)	(0.386)	(0.111)	(0.089)
MP	0.2170	0.1875	0.0745	0.0258	0.1794	0.1780
	(0.438)	(0.083)	(0.266)	(0.009)	(0.002)	(0.000)
MV-C	0.2475	0.1379	**0.1341***	0.1628	**0.2189***	0.2165
	(0.282)	(0.017)	(0.085)	(0.397)	(0.088)	(0.152)
BS-C	0.2395	0.1446	0.1246	0.1671	**0.2186***	0.2138
	(0.334)	(0.024)	(0.130)	(0.354)	(0.086)	(0.191)
MIN-C	0.1909	0.2392	**0.1250****	0.0728	**0.2204***	0.1428
	(0.300)	(0.267)	(0.044)	(0.002)	(0.059)	(0.141)
G-MIN-C	0.2123	0.2391	0.1005	0.1109	**0.2085****	**0.2428****
	(0.423)	(0.180)	(0.233)	(0.023)	(0.048)	(0.049)
MV-MIN	0.1635	0.0655	0.0637	0.1378	**0.3012****	**0.4673*** **
	(0.228)	(0.018)	(0.354)	(0.426)	(0.043)	(0.000)
EW-MIN	0.1999	0.2291	**0.1485****	0.0832	**0.2863*** **	-0.3274
	(0.345)	(0.392)	(0.041)	(0.001)	(0.002)	(0.000)
ERC-C	0.2191	**0.2312****	0.0962	0.1254	**0.1963*** **	**0.2371****
	(0.483)	(0.027)	(0.086)	(0.039)	(0.003)	(0.048)

classes composing the investment portfolios are portfolios themselves, thus mitigating the differences in risk between the portfolio's constituents. Consequently, the allocation of the ERC portfolio is very similar to the one of the 1/N. The similarity of the allocation between the two strategies is also confirmed by the closeness of the values of the metrics for each experiment, see subsections 2.4, 2.5, and 2.6.

Another interesting fact is that, compared to the original study, the number of strategies that beat the 1/N is greater. Excluding the ERC to restrict to the same strategies, the number of scenarios in which one alternative

Table 3: Certainty equivalent of the out-of-sample returns, $W = 120$, $H = 1$.

Strategy	S&P	Ind	Int	MSH	FF1	FF4
1/N	0.0081	0.0084	0.0033	0.0030	0.0087	0.0080
MV (in sample)	0.0313	0.0481	0.0132	0.0047	0.0287	-0.0038
MV	0.0065	-1.4183	-0.0124	-0.6566	**0.0275*** **	0.0034
	(0.359)	(0.000)	(0.121)	(0.000)	(0.007)	(0.027)
BS	0.0064	-0.9526	0.0002	-0.1059	0.0252	0.0029
	(0.309)	(0.000)	(0.333)	(0.001)	(0.002)	(0.018)
DM	0.0088	**0.0153*** **	0.0036	-0.0206	0.0090	-0.9343
	(0.135)	(0.001)	(0.474)	(0.003)	(0.456)	(0.000)
MIN	0.0062	0.0068	0.0047	0.0011	0.0113	-0.0014
	(0.224)	(0.166)	(0.222)	(0.000)	(0.121)	(0.000)
VW	**0.0092** **	0.0062	0.0041	0.0062	0.0062	0.0062
	(0.042)	(0.000)	(0.123)	(0.023)	(0.016)	(0.045)
MP	0.0085	0.0076	0.0025	-0.0001	0.0084	0.0083
	(0.170)	(0.233)	(0.284)	(0.113)	(0.041)	(0.284)
MV-C	**0.0112***	0.0066	**0.0057***	**0.0053** **	**0.0103***	**0.0100** **
	(0.091)	(0.216)	(0.062)	(0.049)	(0.087)	(0.039)
BS-C	0.0107	0.0069	0.0052	**0.0053** **	**0.0102***	**0.0097***
	(0.124)	(0.237)	(0.101)	(0.041)	(0.087)	(0.068)
MIN-C	0.0061	0.0076	0.0043	0.0012	0.0088	0.0021
	(0.161)	(0.260)	(0.161)	(0.001)	(0.457)	(0.002)
G-MIN-C	0.0069	0.0080	0.0035	0.0019	0.0089	0.0055
	(0.185)	(0.324)	(0.395)	(0.007)	(0.399)	(0.049)
MV-MIN	0.0062	-0.9491	0.0019	-0.0889	**0.0239*** **	0.0028
	(0.270)	(0.000)	(0.392)	(0.002)	(0.001)	(0.016)
EW-MIN	0.0067	0.0072	0.0049	0.0014	0.0107	-0.0013
	(0.219)	(0.177)	(0.154)	(0.000)	(0.116)	(0.000)
ERC-C	0.0076	0.0083	0.0034	0.0022	**0.0089***	0.0056
	(0.181)	(0.374)	(0.275)	(0.009)	(0.066)	(0.045)

strategy significantly beats the benchmark for the Sharpe ratio in our experiment is 19 against the 5 times of the original paper. The same happens to the certainty equivalent: we observe 12 cases of statistically significant over-performance against the 2 cases of the referring paper. An explanation of this evidence was given by the authors themselves when they conjectured that the number of strategies outperforming the benchmark could depend on the volatility of the prices, "all else being equal, the performance of the sample-based mean-variance (and that of the optimizing policies in general) would improve relative to that of the 1/N policy if the idiosyncratic asset volatility was much higher than 20% (...). A second reason for the optimizing policies to perform relatively better is that with higher idiosyncratic volatility the covariance matrix of returns is less likely to be singular, and hence, easier to invert", see DeMiguel et al. (2009b). This confirms the common experience of traders and investors that it is easier to implement profitable active investment strategies in periods of high market turbulence.

Concerning the CE, see Table 3, the strategy that more often beats the 1/N is the MV-C. Again, there is no evidence of one systematically better strategy than the benchmark. Moreover, the fact that for different metrics, we identify different strategies as the best alternative to the equally weighted portfolio is a further and very strong argument that supports the evidence that it does not exist a strategy that systematically outperforms the benchmark for all the datasets, the metrics and the values of the parameters.

Table 4: Absolute turnovers of the out-of-sample strategies, $W = 120$, $H = 1$.

	S&P	Ind	Int	MSH	FF1	FF4
1/N	0.0219	0.0233	0.0175	0.0222	0.0170	0.0206
MV (in sample)	–	–	–	–	–	–
MV	2.8799	154.3392	787.4477	12.9836	4.9110	7.1502
BS	1.8709	120.8384	19.5333	7.4242	3.0461	6.4800
DM ($\sigma_\alpha = 1.0\%$)	0.2345	5.3886	4.0160	1.5111	2.7047	25.4996
MIN	0.4639	0.3644	0.4034	0.0261	0.7047	0.1269
VW	**0**	**0**	**0**	**0**	**0**	**0**
MP	0.0235	0.0412	0.0421	0.3859	0.0186	0.0206
MV-C	0.2241	0.1883	0.1046	0.0681	0.2330	0.2367
BS-C	0.2426	0.1895	0.0994	0.0685	0.2226	0.2347
MIN-C	0.0575	0.0576	0.0466	0.0249	0.0448	0.0311
G-MIN-C	0.0428	0.0403	0.0393	0.0244	0.0248	0.0335
MV-MIN	1.7764	123.0834	10.0334	9.0736	2.6325	6.2566
EW-MIN	0.3429	0.2752	0.3224	0.0261	0.5207	0.1264
ERC-C	0.0242	0.0243	0.0181	0.0229	0.0170	0.0330

Table 4 reports the absolute turnovers of the single strategies for the benchmark strategy. One specific preliminary comment is needed for the turnover. As expected, except for the VW strategy, which is a buy-and-hold strategy with null turnover, there is no less costly strategy with respect to the 1/N. The turnover, quantitatively measuring the level of re-balancing needed to implement a strategy, is a good approximation of the impact of the transaction costs on implementing the investment in practice. Nevertheless, we do not necessarily look for an investment strategy that is less costly than the 1/N since transaction costs are already considered in calculating the returns used to compute the out-of-sample SR and CE. The discussion on the level of turnover of different investment strategies hides a very important question when choosing which strategy to implement in practice: is it preferable to an active investment strategy characterized by high transaction costs or a passive investment strategy where re-balancing is so rare that the costs are not relevant? Considering the importance of the transactions costs in the monthly allocation problem of the present application, one can easily imagine their impact in the framework of high-frequency trading.

2.4. The second case: $W = 60$, $H = 1$

The second experiment, whose results are collected in tables 5, 6, and 7 of the present section, differs from the referring experiment only for the length of the estimation window.

Table 5: Sharpe ratios of the out-of-sample returns, $W = 60$, $H = 1$.

Strategy	S&P	Ind	Int	MSH	FF1	FF4
1/N	0.1404	0.2163	0.0937	0.1399	0.1832	0.1879
MV (in sample)	0.3868	0.5677	0.2729	0.1624	0.4852	0.7256
MV	0.1763	0.1931	0.0183	0.0456	-0.0226	0.1949
	(0.334)	(0.369)	(0.184)	(0.080)	(0.002)	(0.458)
BS	0.1896	0.1931	0.0070	0.0478	-0.0185	**0.2808***
	(0.256)	(0.369)	(0.152)	(0.083)	(0.002)	(0.082)
DM	-0.0684	0.2731	-0.0066	0.0512	0.1386	-0.0461
	(0.022)	(0.142)	(0.066)	(0.060)	(0.204)	(0.000)
MIN	0.1667	0.2403	0.1114	0.0713	0.2255	-0.4895
	(0.332)	(0.296)	(0.346)	(0.008)	(0.178)	(0.000)
VW	0.1486	0.1498	0.1104	0.1498	0.1498	0.1498
	(0.223)	(0.000)	(0.102)	(0.380)	(0.048)	(0.037)
MP	0.0298	0.1703	-0.0259	0.0399	0.1562	0.1547
	(0.035)	(0.040)	(0.008)	(0.035)	(0.083)	(0.043)
MV-C	0.1830	0.1739	**0.1343***	0.1272	0.1956	0.1509
	(0.172)	(0.119)	(0.068)	(0.376)	(0.262)	(0.138)
BS-C	0.1823	0.1833	**0.1323***	0.1395	0.2039	0.1511
	(0.177)	(0.170)	(0.070)	(0.496)	(0.134)	(0.142)
MIN-C	0.1587	0.2375	0.1139	0.0833	**0.2210****	0.1521
	(0.321)	(0.220)	(0.163)	(0.018)	(0.019)	(0.228)
G-MIN-C	0.1576	0.2304	0.1025	0.1101	**0.2075*****	**0.2259***
	(0.229)	(0.217)	(0.242)	(0.086)	(0.009)	(0.094)
MV-MIN	0.1844	0.1924	-0.0177	-0.0164	-0.0135	**0.3755*****
	(0.265)	(0.365)	(0.095)	(0.009)	(0.002)	(0.003)
EW-MIN	0.1562	0.2386	0.1093	0.0970	**0.2180***	-0.4680
	(0.325)	(0.201)	(0.301)	(0.017)	(0.067)	(0.000)
ERC-C	0.1443	**0.2270****	0.0970	0.1203	**0.1907*****	**0.2258***
	(0.310)	(0.019)	(0.117)	(0.111)	(0.002)	(0.084)

Concerning the referring experiment, a few interesting qualitative differences are clear. Regarding the out-of-sample Sharpe ratios, see Table 5, we note that the number of strategies that outperform the 1/N is smaller; the

Table 6: Certainty equivalent of the out-of-sample returns, $W = 60$, $H = 1$.

Strategy	S&P	Ind	Int	MSH	FF1	FF4
1/N	0.0053	0.0084	0.0034	0.0027	0.0084	0.0078
MV (in sample)	0.0268	0.0525	0.0183	0.0045	0.0316	-0.0040
MV	0.0108	-1.2058	-4.7540	-2.2794	-2.0335	0.0032
	(0.161)	(0.000)	(0.000)	(0.000)	(0.000)	(0.024)
BS	0.0089	-1.0578	-1.6090	-0.1427	-0.6893	0.0029
	(0.171)	(0.000)	(0.000)	(0.000)	(0.000)	(0.016)
DM	-1.1734	**0.0178*** **	-0.0025	0.0013	0.0081	-0.4297
	(0.000)	(0.003)	(0.063)	(0.285)	(0.461)	(0.000)
MIN	0.0056	0.0082	0.0035	0.0011	0.0088	-0.0019
	(0.458)	(0.447)	(0.474)	(0.003)	(0.442)	(0.000)
VW	0.0058	0.0058	0.0040	**0.0058** **	0.0058	0.0058
	(0.157)	(0.000)	(0.167)	(0.023)	(0.006)	(0.019)
MP	0.0003	0.0070	-0.0026	0.0006	0.0072	0.0071
	(0.039)	(0.116)	(0.008)	(0.218)	(0.116)	(0.256)
MV-C	**0.0084***	0.0089	**0.0058** **	0.0042	0.0090	0.0066
	(0.086)	(0.409)	(0.042)	(0.144)	(0.304)	(0.247)
BS-C	**0.0083***	0.0091	**0.0057** **	**0.0045***	0.0093	0.0063
	(0.096)	(0.352)	(0.045)	(0.082)	(0.194)	(0.177)
MIN-C	0.0050	0.0077	0.0038	0.0014	0.0089	0.0021
	(0.434)	(0.254)	(0.318)	(0.006)	(0.308)	(0.002)
G-MIN-C	0.0053	0.0078	0.0036	0.0019	0.0089	0.0051
	(0.495)	(0.219)	(0.382)	(0.031)	(0.230)	(0.033)
MV-MIN	0.0076	-1.0746	-1.0242	-0.0664	-0.2738	0.0028
	(0.240)	(0.000)	(0.000)	(0.000)	(0.000)	(0.014)
EW-MIN	0.0052	0.0080	0.0035	0.0016	0.0085	-0.0019
	(0.472)	(0.373)	(0.456)	(0.005)	(0.480)	(0.000)
ERC-C	0.0052	0.0083	0.0035	0.0021	**0.0087***	0.0051
	(0.373)	(0.398)	(0.294)	(0.031)	(0.052)	(0.032)

best strategy within the others is again the ERC that significantly beats the benchmark in three cases.

If we focus on the certainty equivalent, see Table 6, the number of scenarios in which some alternative strategy beats the benchmark becomes very small. BS-C is the strategy that often beats the 1/N in this framework,.

Looking at the turnovers (see Table 7), we note that they are generally higher than the ones of the referring experiment; this result is not surprising since it is a direct consequence of the increased instability of the allocation due to the reduction of W.

Table 7: Absolute turnovers of the out-of-sample strategies, $W = 60$, $H = 1$.

	S&P	Ind	Int	MSH	FF1	FF4
1/N	0.0232	0.0229	0.0187	0.0222	0.0166	0.0202
MV (in sample)	–	–	–	–	–	–
MV	10.7800	127.3911	165.2349	13.1624	69.3177	8.5157
BS	6.2632	119.9690	85.0147	6.6851	48.5469	7.1016
DM ($\sigma_\alpha = 1.0\%$)	83.7853	17.8607	4.7192	1.2386	44.0289	15.7804
MIN	1.3284	1.0366	0.8827	0.0358	1.5229	0.1904
VW	**0**	**0**	**0**	**0**	**0**	**0**
MP	0.1623	0.0837	0.3142	0.8700	0.0381	0.0411
MV-C	0.2327	0.2547	0.1684	0.1080	0.3284	0.3451
BS-C	0.1919	0.2640	0.1666	0.1091	0.2845	0.3280
MIN-C	0.1025	0.1101	0.0851	0.0338	0.0903	0.0427
G-MIN-C	0.0669	0.0621	0.0610	0.0319	0.0394	0.0423
MV-MIN	5.7163	117.6299	50.5379	7.8344	23.4477	6.4694
EW-MIN	0.6786	0.5605	0.6014	0.0346	0.8173	0.1889
ERC-C	0.0294	0.0272	0.0222	0.0260	0.0183	0.0403

2.5. Experiment: increasing W and $H = 1$

Since the results of the application, as underlined by the authors of the original paper, can depend on the fact that the length of the rolling estimation window is short and fixed, we implement an experiment in which the estimation window grows by simply adding the newly available data to the old ones. Starting from $W = 120$, we calculate the optimal allocation for each strategy. At the end of the first holding period the length of the estimation window becomes $W + H = 121$. We keep adding new returns to the estimation window such that the last allocation is finally calculated using all the available data except for the last observation. We are conscious that this choice could have positive and negative consequences on the empirical results. On the one hand, increasing the length of W produces optimal portfolios with a more stable allocation in time. Then these portfolios are less affected by the impact of the transaction costs on their implementation. On the opposite, this approach gives the same weight to all the observations in the database. Consequently, very old data have the same importance as the more recent observations, negatively impacting on the capacity of the model's parameters to reflect the actual market conditions significantly. This is a standard trade-off between a stable long-run passive investment allocation and an aggressive short-term investment strategy that tries to make money efficiently by timing the market and needs to bear a large transaction cost.

Table 8: Sharpe ratios of the out-of-sample returns, increasing W, $H = 1$.

Strategy	S&P	Ind	Int	MSH	FF1	FF4
1/N	0.2186	0.2205	0.0924	0.1517	0.1893	0.1942
MV (in sample)	0.4857	0.5472	0.2728	0.1707	0.4788	0.6945
MV	0.1241	**0.3945*****	0.1491	0.1480	**0.3742*****	**0.4812*****
	(0.122)	(0.008)	(0.128)	(0.449)	(0.002)	(0.000)
BS	0.1665	**0.3960*****	**0.1322***	0.1289	**0.3958*****	**0.4845*****
	(0.228)	(0.007)	(0.083)	(0.166)	(0.000)	(0.000)
DM	0.2368	**0.4099*****	**0.1204****	0.1630	**0.2786*****	**0.2931***
	(0.122)	(0.000)	(0.040)	(0.348)	(0.000)	(0.067)
MIN	0.2028	0.2357	0.1168	0.1088	**0.2841****	-0.2816
	(0.404)	(0.349)	(0.165)	(0.032)	(0.021)	(0.000)
VW	**0.2426***	0.1619	**0.1139***	0.1619	0.1619	0.1619
	(0.069)	(0.000)	(0.064)	(0.386)	(0.111)	(0.089)
MP	0.2177	0.2132	0.0889	0.1250	0.1829	0.1819
	(0.451)	(0.057)	(0.075)	(0.261)	(0.003)	(0.001)
MV-C	0.2582	0.1851	0.1161	0.1601	0.1681	0.1688
	(0.226)	(0.144)	(0.182)	(0.406)	(0.167)	(0.120)
BS-C	0.2356	0.1884	**0.1326****	0.1449	0.1739	0.1731
	(0.346)	(0.165)	(0.026)	(0.423)	(0.223)	(0.149)
MIN-C	0.2032	0.2269	**0.1252****	0.1088	**0.2176***	0.1654
	(0.390)	(0.422)	(0.017)	(0.032)	(0.052)	(0.263)
G-MIN-C	0.2173	0.2404	**0.1111****	0.1122	**0.2090****	0.2184
	(0.482)	(0.204)	(0.011)	(0.040)	(0.043)	(0.191)
MV-MIN	0.1744	**0.3957*****	0.0445	0.1154	**0.4039*****	**0.4825*****
	(0.261)	(0.007)	(0.100)	(0.083)	(0.000)	(0.000)
EW-MIN	0.2162	0.2441	**0.1125***	0.1204	**0.2702*****	-0.2707
	(0.480)	(0.162)	(0.094)	(0.048)	(0.002)	(0.000)
ERC-C	0.2202	**0.2272***	**0.0954***	0.1301	**0.1946****	0.2230
	(0.427)	(0.086)	(0.067)	(0.096)	(0.005)	(0.134)

The increased numerical stability induced by the longer estimation window positively affects the strategies' out-of-sample performance under comparison. In this framework, with respect to the previous experiments, a larger number of strategies significantly outperform the 1/N, both in terms

Table 9: Certainty equivalent of the out-of-sample returns, increasing W, $H = 1$.

Strategy	S&P	Ind	Int	MSH	FF1	FF4
1/N	0.0081	0.0084	0.0033	0.0030	0.0087	0.0080
MV (in sample)	0.0313	0.0481	0.0132	0.0047	0.0287	-0.0038
MV	0.0053	**0.0509*****	0.0057	0.0033	**0.0276*****	0.0036
	(0.243)	(0.000)	(0.151)	(0.312)	(0.000)	(0.033)
BS	0.0060	**0.0484*****	0.0042	0.0024	**0.0237*****	0.0031
	(0.229)	(0.000)	(0.262)	(0.115)	(0.000)	(0.022)
DM	0.0088	**0.0184*****	**0.0044***	**0.0048****	0.0108	0.0111
	(0.125)	(0.000)	(0.095)	(0.028)	(0.055)	(0.143)
MIN	0.0066	0.0078	0.0036	0.0019	0.0103	-0.0013
	(0.277)	(0.347)	(0.401)	(0.012)	(0.229)	(0.000)
VW	**0.0092****	0.0062	0.0041	**0.0062****	0.0062	0.0062
	(0.042)	(0.000)	(0.123)	(0.023)	(0.016)	(0.045)
MP	0.0084	0.0085	0.0032	0.0054	0.0085	0.0085
	(0.150)	(0.309)	(0.273)	(0.142)	(0.063)	(0.165)
MV-C	**0.0121***	0.0064	0.0042	**0.0060****	0.0079	0.0080
	(0.056)	(0.068)	(0.242)	(0.027)	(0.264)	(0.482)
BS-C	0.0096	0.0065	**0.0049***	**0.0045***	0.0082	0.0082
	(0.201)	(0.076)	(0.053)	(0.097)	(0.329)	(0.449)
MIN-C	0.0063	0.0075	0.0043	0.0019	0.0089	0.0026
	(0.207)	(0.245)	(0.110)	(0.012)	(0.418)	(0.003)
G-MIN-C	0.0071	0.0081	**0.0040***	0.0020	0.0089	0.0051
	(0.205)	(0.390)	(0.078)	(0.015)	(0.362)	(0.025)
MV-MIN	0.0057	0.0485	0.0010	0.0020	0.0224	0.0030
	(0.185)	(0.000)	(0.092)	(0.041)	(0.000)	(0.020)
EW-MIN	0.0070	0.0081	0.0037	0.0021	0.0100	-0.0012
	(0.280)	(0.378)	(0.312)	(0.015)	(0.197)	(0.000)
ERC-C	0.0077	0.0083	0.0034	0.0023	**0.0089***	0.0052
	(0.190)	(0.315)	(0.264)	(0.033)	(0.085)	(0.025)

of Sharpe ratio, see Table 8, and certainty equivalent, see Table 9. Despite this evidence, finding at least one strategy that significantly outperforms 1/N for all the databases and the metrics is still impossible. By comparing this experiment with one of the original paper performed on simulated data, we confirm that the length of the estimation window is a key parameter to obtain good performance of the optimization-based models in the mean-variance framework.

One further interesting aspect to underline is how the performance of the ERC, that is very good with estimation windows of limited lengths, does not

Table 10: Absolute Turnovers of the out-of-sample strategies, increasing W, $H = 1$.

Strategy	S&P	Ind	Int	MSH	FF1	FF4
1/N	0.0219	0.0233	0.0175	0.0222	0.0170	0.0206
MV (in sample)	–	–	–	–	–	–
MV	1.5702	34.3216	0.5211	0.0297	1.4483	7.1640
BS	0.6810	13.9747	0.1724	0.0250	0.9417	6.6248
DM	0.1362	0.4944	0.0416	0.0210	0.2120	13.5535
MIN	0.1705	0.1157	0.0961	0.0209	0.3193	0.0989
VW	**0**	**0**	**0**	**0**	**0**	**0**
MP	**0.0217**	0.0234	0.0178	0.0518	0.0177	**0.0189**
MV-C	0.0709	0.0798	0.0561	**0.0081**	0.1385	0.1369
BS-C	0.1406	0.0747	0.0469	0.0292	0.1186	0.1221
MIN-C	**0.0199**	0.0286	0.0201	**0.0209**	0.0332	0.0252
G-MIN-C	**0.0215**	0.0238	0.0216	**0.0209**	0.0190	0.0291
MV-MIN	0.6322	15.7207	0.3726	0.0272	0.8182	6.4754
EW-MIN	0.1279	0.0738	0.0620	0.0214	0.1816	0.0981
ERC-C	0.0223	0.0237	**0.0174**	**0.0215**	**0.0165**	0.0289

get better while the results of the other strategies seem to generally improve.

We also underline that, not only the number of scenarios in which the strategies beat the benchmark is larger, but also the significance level of the statistical test gets better (roughly speaking, to note this fact it is sufficient to count the number of stars in table 8 with respect to the number of stars in tables 5 and 2).

As expected, in this experiment, the importance of W on the numerical stability of the allocation models is clear through the amount of the transaction costs represented by turnover. In general, we can observe a decrease in the transaction costs, noting how the turnover of the various strategies is closer to the referring one. Moreover, in this framework, few strategies are characterized by a lower turnover with respect to 1/N.

2.6. $W = 120$ *and* $H = 12$

In this experiment, we vary the length of the holding period of the investment strategies as suggested by Chan et al. (1999) and Jagannathan and Ma (2003). This permits to influence the turnover of the strategies acting on a different parameter. The results of this experiment are in Tables 11, 12 and 13. Again, as noted in the previous experiment, see subsection 2.5, the number of strategies that outperform 1/N is larger than what obtained in the first two experiments.

Table 11: Sharpe ratios of the out-of-sample returns, $W = 120$, $H = 12$.

Strategy	S&P	Ind	Int	MSH	FF1	FF4
1/N	0.9140	0.7384	0.2691	0.4540	0.6228	0.6230
MV (in sample)	1.6825	1.8956	0.9448	0.5913	1.6586	2.4058
MV	0.6007	**0.9015****	0.2313	0.1051	**0.8054*****	**0.7049***
	(0.001)	(0.021)	(0.364)	(0.000)	(0.002)	(0.060)
BS	0.7217	**0.8800****	0.2670	0.0740	**0.8561*****	**0.7065***
	(0.019)	(0.036)	(0.492)	(0.000)	(0.000)	(0.060)
DM	0.9338	0.6007	0.1463	0.3217	0.6294	-0.1513
	(0.262)	(0.009)	(0.012)	(0.014)	(0.447)	(0.000)
MIN	0.7932	0.7182	**0.3232***	0.1982	**0.8614*****	-0.6124
	(0.125)	(0.346)	(0.062)	(0.000)	(0.000)	(0.000)
VW	0.8979	0.5586	**0.3244*****	**0.5586*****	0.5586	0.5586
	(0.007)	(0.000)	(0.002)	(0.000)	(0.000)	(0.000)
MP	0.8857	0.6892	0.0876	0.2525	0.6022	0.6004
	(0.022)	(0.000)	(0.000)	(0.001)	(0.000)	(0.000)
MV-C	0.6150	0.4099	**0.3555*****	0.5389	0.6365	0.6386
	(0.000)	(0.000)	(0.000)	(0.063)	(0.354)	(0.332)
BS-C	0.6082	0.4188	**0.3371*****	**0.5607***	0.6781	0.6676
	(0.000)	(0.000)	(0.000)	(0.023)	(0.068)	(0.109)
MIN-C	0.9539	**0.8438*****	**0.3090*****	0.2117	**0.7798*****	0.3475
	(0.348)	(0.001)	(0.004)	(0.000)	(0.000)	(0.000)
G-MIN-C	**1.0639*****	**0.8008*****	0.2663	**0.2991*****	**0.7008*****	0.6084
	(0.009)	(0.003)	(0.360)	(0.000)	(0.000)	(0.313)
MV-MIN	0.7568	**0.8740****	0.2167	0.0637	**0.8465*****	0.7000
	(0.043)	(0.042)	(0.220)	(0.000)	(0.000)	(0.079)
EW-MIN	0.9360	0.7776	**0.3484*****	**0.2412*****	**0.8419*****	-0.6085
	(0.389)	(0.163)	(0.003)	(0.000)	(0.000)	(0.000)
ERC-C	**0.9620*****	**0.7755*****	**0.2733****	**0.3428*****	0.6384	0.6234
	(0.001)	(0.000)	(0.040)	(0.000)	(0.000)	(0.495)

Table 12: Certainty equivalent of the out-of-sample returns, $W = 120$, $H = 12$.

Strategy	S&P	Ind	Int	MSH	FF1	FF4
1/N	0.1024	0.1032	0.0361	0.0358	0.1061	0.0974
MV (in sample)	0.0313	0.0481	0.0132	0.0047	0.0287	-0.0038
MV	0.0873	0.1230	-1.1027	-0.9703	**0.2979****	0.0478
	(0.355)	(0.482)	(0.029)	(0.005)	(0.027)	(0.032)
BS	0.0835	0.1838	-0.0647	-0.4629	**0.2816*****	0.0418
	(0.268)	(0.418)	(0.304)	(0.015)	(0.007)	(0.022)
DM	0.1136	0.1577	0.0063	0.0407	0.0925	-13.2557
	(0.135)	(0.187)	(0.257)	(0.431)	(0.316)	(0.000)
MIN	0.0704	0.0857	0.0475	0.0141	**0.1460***	-0.0222
	(0.149)	(0.206)	(0.347)	(0.001)	(0.063)	(0.001)
VW	0.1051	0.0738	0.0469	**0.0738****	0.0738	0.0738
	(0.396)	(0.000)	(0.235)	(0.036)	(0.058)	(0.109)
MP	0.1087	0.1033	-0.0125	0.0282	0.1027	0.1021
	(0.139)	(0.496)	(0.149)	(0.400)	(0.033)	(0.181)
MV-C	0.1164	0.0680	0.0600	**0.0618***	0.1127	0.1131
	(0.384)	(0.119)	(0.114)	(0.098)	(0.386)	(0.235)
BS-C	0.1104	0.0689	0.0547	**0.0617***	0.1194	0.1172
	(0.427)	(0.117)	(0.156)	(0.082)	(0.276)	(0.177)
MIN-C	0.0751	0.0956	0.0448	0.0153	0.1108	0.0257
	(0.167)	(0.297)	(0.253)	(0.001)	(0.376)	(0.002)
G-MIN-C	0.0855	0.0992	0.0352	0.0235	0.1095	0.0650
	(0.180)	(0.343)	(0.449)	(0.008)	(0.331)	(0.032)
MV-MIN	0.0817	0.1725	0.0235	-0.5624	**0.2652*****	0.0399
	(0.237)	(0.430)	(0.409)	(0.010)	(0.008)	(0.019)
EW-MIN	0.0807	0.0911	0.0533	0.0178	**0.1373***	-0.0219
	(0.165)	(0.234)	(0.234)	(0.001)	(0.059)	(0.001)
ERC-C	0.0972	0.1024	0.0371	0.0271	**0.1083***	0.0671
	(0.176)	(0.397)	(0.337)	(0.013)	(0.055)	(0.035)

ERC and G-MIN-C are strategies that more often beat the benchmark in terms of Sharpe ratio; if we look at the CE, a few strategies outperform in one single scenario. The fact that the strategies that are better in terms of CE are less numerous than the ones that over-perform with the SR is a behavior that can be found in all the experiments.

The present experiment confirms that, while the performance of the optimization-based strategies is strictly related to the estimation window's

Table 13: Absolute turnovers of the out-of-sample strategies, $W = 120$, $H = 12$.

Strategy	S&P	Ind	Int	MSH	FF1	FF4
1/N	0.0017	0.0019	0.0017	0.0017	0.0014	0.0017
MV (in sample)	–	–	–	–	–	–
MV	0.2115	6.1024	5.0181	0.7536	0.3263	0.7786
BS	0.1190	4.9655	2.3650	0.4481	0.2063	0.7229
DM ($\sigma_\alpha = 1.0\%$)	0.0127	0.2441	0.2538	0.0268	0.3108	6.6035
MIN	0.0275	0.0240	0.0548	0.0018	0.0598	0.0103
VW	**0**	**0**	**0**	**0**	**0**	**0**
MP	0.0018	0.0020	0.0174	0.0325	0.0015	**0.0016**
MV-C	0.0060	0.0105	0.0074	0.0069	0.0194	0.0196
BS-C	0.0172	0.0112	0.0065	0.0047	0.0193	0.0198
MIN-C	0.0039	0.0038	0.0050	0.0018	0.0020	0.0021
G-MIN-C	0.0026	0.0027	0.0037	0.0018	0.0015	0.0024
MV-MIN	0.1047	4.9390	0.6392	0.4873	0.1789	0.7037
EW-MIN	0.0202	0.0180	0.0456	0.0019	0.0444	0.0102
ERC-C	**0.0017**	**0.0019**	**0.0017**	**0.0017**	0.0015	0.0024

length W, the ERC strategy's results seem to be poorly influenced by this parameter. Since the alternative strategies under comparison in the paper use the covariance matrix as the fundamental input, it could be useful to study the differences in detail in future research. We can argue that different models can use the same input and, consequently, the same information, in very different ways.

3. Conclusion

The main finding of the present research qualitatively confirms the general results obtained in DeMiguel et al. (2009b). In the same settings of the original experiment (estimation windows of limited lengths, $W = 60$ or 120, monthly returns, same datasets, proportional transaction costs of 50 basis points), comparing the same allocation strategies to the equally weighted portfolio through the use of the same metrics (Sharpe ratio, certainty equivalent, turnover), none of the considered strategies can significantly and systematically outperform the 1/N. A more accurate observation of the results suggests that, compared to the original paper, the number of strategies that beat the benchmark is larger. This evidence probably depends on the increased volatility of the more recent data. The same authors underlined how the results could be somehow influenced by the level of volatility in the market. It is well known that active portfolio strategies can over-perform in periods characterized by financial turbulence. We also underline that the equally risk contribution strategy, not considered in the original paper, is very competitive in some scenarios. Even if this fact strongly supports the validity of the ERC in practice, limited to the experiment performed in the paper, we cannot conclude that it is statistically significantly better than the $1/N$ for all the metrics and for all the databases. This fact reinforces the main result.

It is also straightforward to notice that changing the settings for the referring experiment does not substantially modify the results. With a shorter estimation window $W = 60$, the number of strategies performing better than the benchmark decreases. One specific comment is needed for the case of a growing length of W: such an experiment was performed only on simulated data in the original paper with the result that optimization-based techniques could outperform the 1/N for very large values of W with respect to N. Our experiment is entirely conducted on real data, and we do not observe an analogous behavior. In our settings, even when W is large with respect to N, it is impossible to find a value of W/N such that the optimization-based strategies start to systematically outperform the 1/N. The last comment concerns the fact that an increased length of the holding period, even if positively impacting on the transaction costs of the active strategies, cannot significantly change the core evidence that *the 1/N strategy still remains a challenging benchmark*.

Declarations

Conflict of interest. The authors have no conflicts of interest to declare that are relevant to the content of this article.
Funding. No funding was received for conducting this study.
Data availability. The datasets analyzed in the current study are available from the corresponding author on reasonable request.
Authors' contributions The authors equally contributed to the design and implementation of the research, the analysis of the results and the writing of the manuscript.

References

Ackermann, F., Pohl, W., Schmedders, K., 2017. Optimal and naive diversification in currency markets. Management Science 63, 3347–3360.

Bessler, W., Opfer, H., Wolff, D., 2017. Multi-asset portfolio optimization and out-of-sample performance: an evaluation of black–litterman, mean-variance, and naïve diversification approaches. The European Journal of Finance 23, 1–30.

Best, M.J., Hlouskova, J., 2008. Quadratic programming with transaction costs. Computers & Operations Research 35, 18–33.

Bickel, P.J., Levina, E., 2008. Regularized estimation of large covariance matrices. The Annals of Statistics 36, 199–227.

Chan, L.K., Karceski, J., Lakonishok, J., 1999. On portfolio optimization: Forecasting covariances and choosing the risk model. The Review of Financial Studies 12, 937–974.

Craig MacKinlay, A., Pástor, L., 2000. Asset pricing models: Implications for expected returns and portfolio selection. The Review of Financial Studies 13, 883–916.

DeMiguel, V., Garlappi, L., Nogales, F.J., Uppal, R., 2009a. A generalized approach to portfolio optimization: Improving performance by constraining portfolio norms. Management Science 55, 798–812.

DeMiguel, V., Garlappi, L., Uppal, R., 2009b. Optimal versus naive diversification: How inefficient is the 1/n portfolio strategy? The Review of Financial Studies 22, 1915–1953.

Fugazza, C., Guidolin, M., Nicodano, G., 2015. Equally weighted vs. long-run optimal portfolios. European Financial Management 21, 742–789.

Garlappi, L., Uppal, R., Wang, T., 2007. Portfolio selection with parameter and model uncertainty: A multi-prior approach. The Review of Financial Studies 20, 41–81.

Hanke, B., Keswani, A., Quigley, G., Stolin, D., Zagonov, M., 2019. The equal-weight tilt in managed portfolios. Economics Letters 182, 59–63.

Hirschberger, M., Qi, Y., Steuer, R.E., 2010. Large-scale mv efficient frontier computation via a procedure of parametric quadratic programming. European Journal of Operational Research 204, 581–588.

Jagannathan, R., Ma, T., 2003. Risk reduction in large portfolios: Why imposing the wrong constraints helps. The Journal of Finance 58, 1651–1683.

James, W., Stein, C., 1992. Estimation with quadratic loss. Breakthroughs in Statistics: Foundations and Basic Theory , 443–460.

Jiang, C., Du, J., An, Y., 2019. Combining the minimum-variance and equally-weighted portfolios: Can portfolio performance be improved? Economic Modelling 80, 260–274.

Jobson, J.D., Korkie, B.M., 1981. Performance hypothesis testing with the sharpe and treynor measures. The Journal of Finance , 889–908.

Kan, R., Zhou, G., 2007. Optimal portfolio choice with parameter uncertainty. Journal of Financial and Quantitative Analysis 42, 621–656.

Kirby, C., Ostdiek, B., 2012. It's all in the timing: simple active portfolio strategies that outperform naive diversification. Journal of Financial and Quantitative Analysis 47, 437–467.

Kritzman, M., Page, S., Turkington, D., 2010. In defense of optimization: the fallacy of 1/n. Financial Analysts Journal 66, 31–39.

Ledoit, O., Wolf, M., 2003. Improved estimation of the covariance matrix of stock returns with an application to portfolio selection. Journal of Empirical Finance 10, 603–621.

Lintner, J., 1965. Security prices, risk, and maximal gains from diversification. The Journal of Finance 20, 587–615.

Maillard, S., Roncalli, T., Teïletche, J., 2010. The properties of equally weighted risk contribution portfolios. The Journal of Portfolio Management 36, 60–70.

Malladi, R., Fabozzi, F.J., 2017. Equal-weighted strategy: Why it outperforms value-weighted strategies? theory and evidence. Journal of Asset Management 18, 188–208.

Markowitz, H., 1952. Portfolio selection. The Journal of Finance 7, 77–91.

Mossin, J., 1966. Equilibrium in a capital asset market. Econometrica , 768–783.

Pástor, L., 2000. Portfolio selection and asset pricing models. The Journal of Finance 55, 179–223.

Pástor, L., Stambaugh, R.F., 2000. Comparing asset pricing models: an investment perspective. Journal of Financial Economics 56, 335–381.

Pflug, G.C., Pichler, A., Wozabal, D., 2012. The 1/n investment strategy is optimal under high model ambiguity. Journal of Banking & Finance 36, 410–417.

Roncalli, T., Weisang, G., 2016. Risk parity portfolios with risk factors. Quantitative Finance 16, 377–388.

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

Sharpe, W.F., 1966. Mutual fund performance. The Journal of Business 39, 119–138.

Stein, C., 1956. Inadmissibility of the usual estimator for the mean of a multivariate normal distribution. Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Contributions to the Theory of Statistics 3, 197–207.

Yuan, M., Zhou, G., 2022. Why naive diversification is not so naive, and how to beat it? Journal of Financial and Quantitative Analysis , 1–32.