THE RIMINI CENTRE FOR ECONOMIC ANALYSIS

# WP 66_13

**Dimitrios D. Thomakos**
University of Peloponnese, Greece
Quantf Research, Greece
The Rimini Centre for Economic Analysis (RCEA), Italy

**Fotis Papailias**
Queen's University Belfast, UK
Quantf Research, Greece

# COVARIANCE AVERAGING FOR IMPROVED ESTIMATION AND PORTFOLIO ALLOCATION

# Covariance Averaging for Improved Estimation and Portfolio Allocation

Dimitrios D. Thomakos*

Department of Economics, University of Peloponnese, Greece

Quantf Research, `www.quantf.com`

Rimini Center for Economic Analysis, Italy

and

Fotis Papailias†

Queen's University Management School, Queen's University Belfast, UK

Quantf Research, `www.quantf.com`

December 4, 2013

**Abstract**

We propose a new method for estimating the covariance matrix of a multivariate

*Department of Economics, University of Peloponnese, Greece, and Rimini Center for Economic Analysis, Rimini, Italy. E-mail Addresses: `thomakos@uop.gr, thomakos@quantf.com`

†**Corresponding Author**. Queen's University Management School, Queen's University Belfast, Riddel Hall, 185 Stranmillis Road, BT9 5EE, Northern Ireland. Tel.: +442890974667, Fax.: +442890974201. E-mail Addresses: `f.papailias@qub.ac.uk, f.papailias@quantf.com`

time series of financial returns. The method is based on estimating sample covariances from overlapping windows of observations which are then appropriately weighted to obtain the final covariance estimate. We extend the idea of (model) covariance averaging offered in the covariance shrinkage approach by means of greater ease of use, flexibility and robustness in averaging information over different data segments. The suggested approach does not suffer from the curse of dimensionality and can be used without problems of either approximation or any demand for numerical optimization.

*Keywords*: averaging, covariance estimation, financial returns, multivariate time series, portfolio allocation, risk management, rolling window.

*JEL Codes*: C32, C58, G11.

# 1 Introduction

In this paper we offer an alternative approach to the estimation of covariance matrices to multivariate time series data, with particular focus on series of financial returns. The main motivation of our work is to extend the opportunity for model covariance averaging, which inherently exists in the covariance shrinkage approach, and to merge it with the literature of window averaging, which is part of the time series forecasting literature. In doing so we offer a straightforward way of robustly estimating, possibly time-varying, covariances of arbitrary large dimensions without getting into problems of dimensionality or heavy non-linear optimization. The method of implementing this idea is simply to combine sample covariance matrices which are estimated from different data windows, i.e. rolling window averaging. The appropriate weights can be chosen in a variety of ways: some of them are ad-hoc, while others can be based on an optimizing criterion function. We present theoretical and empirical results using both these approaches.

A reasonable question to ask is: what may be the possible advantages of using the estimation approach of this paper? Simplicity in estimation, applicability to problems of any dimension and robustness in various data generating processes are the principal ones. Contrary to other methods which attempt to deal with the estimation of covariances in large dimensions and need some kind of approximation, our method requires only the estimation of sample covariances - which can be done pair wise for any time series length greater than two. Furthermore, the averaging across different estimation windows preserves the positive-definiteness of the final covariance matrix.

The literature related to our paper falls into two different strands. The first strand is associated with the forecasting literature which deals with performance improvements via the averaging of rolling and recursive windows. On rolling window averaging work has been done by, among others, Pesaran et al. (2009), Clark and McCracken (2009), Bhattacharya and Thomakos (2011) and Rossi and Inoue (2012). The second strand relates to various methods of covariance matrix estimation. For the area of covariance shrinkage see the seminal works of Ledoit and Wolf (2003) and Ledoit and Wolf (2004) and reference therein, and also the recent paper by Ledoit and Wolf (2013) on spectrum estimation and Principal Components Analysis in estimating large scale covariances. Other papers which work with shrinkage include Wang (2005), Kwan (2008), Bajeux-Besnainou et al. (2012) and Kourtis et al. (2012). On large(r) scale covariance estimation see, among others, Chan et al. (1999), Engle (2002), Ledoit et al. (2003), Bauwens et al. (2006), Pelletier (2006), Fan et al. (2008), Silvennoinen and Tersvirta (2009) and Huo et al. (2012).

Note that we do not cover the literature which deals with multivariate parametric GARCH-type models or realized measures for covariance estimation; both these approaches are beyond the spirit and scope of the methods presented in this and the above papers (although we use parametric models later in the paper for simulations and performance evaluation based on realized covariance estimates).

Finally, the literature on portfolio optimization, where the input of a covariance matrix is essential, cannot be possibly reviewed in depth here. The papers of Kan and Zhou (2007), DeMiguel et al. (2009) and Martellini and Ziemann (2010) present portfolio optimization methods where the estimation and use of the covariance matrix plays a central role and our results might be of use.

In our analysis we offer a number of new results and insights. On the theoretical side, we first generalize the idea embedded in covariance shrinkage and extend it to the case of averaging across different segments of the data. Second, we present different approaches for selecting the weights which should be allocated to covariances estimated from these segments, taking our lead again from the shrinkage methodology and extending them to distance-based and optimization-based weights; for this particular area of the paper we show how a variety of methods can be employed, including single and multi-parameter specifications and adjustments for estimation bias. Third, we show that the weights which are obtained from the optimization scheme can be interpreted in terms of estimation risk, namely, that covariances which carry higher estimation risk should be down-weighted before forming the final covariance estimate. On the empirical, and practical side, we first perform an extensive simulation analysis with two types of data generating process. Our results from the simulations show that the proposed covariance estimator is highly competitive in fitting and forecasting many steps ahead, vis-a-vis the realized covariance estimator benchmark, and in most experiments it outperforms the sample and simple shrinkage estimator. These results are practically relevant for two reasons: simplicity of estimation and robustness in a time-varying environment, both making the suggested method a direct competitor of more complicated parametric methods, even when forecasts of covariances/correlations alone are required.

Finally, our results from the application on a GMV portfolio context suggest that, while our covariance estimates perform at least on a par and (most often) better in terms of the final wealth of the portfolio, they clearly outperform the competition in terms of risk measures, in particular maximum drawdown. This means that using the covariance averaging suggested in this paper can lead to possibly significant improvements in the risk-return trade-off for an investor operating in the particular portfolio environment considered here.

The rest of the paper is organized as follows: in Section 2 we present the problem under review, with an overview of covariance shrinkage and covariance averaging and we discuss in detail our proposed methodology; in Section 3 we conduct a simulation analysis; and in Section 4 we present the details of our application in the context of a GMV portfolio; Section 5 offers some concluding remarks and directions for future work.

# 2 Problem and Methodology

## 2.1 Preliminaries

Suppose that we have available $N$ assets whose returns at period $t$ are denoted by the $(N \times 1)$ vector $\boldsymbol{R}_t \stackrel{\text{def}}{=} [R_{t1}, \ldots, R_{tN}]^\top$. These returns have an unknown conditional distribution with mean $\boldsymbol{\mu}_t$ and covariance matrix $\boldsymbol{\Sigma}_t$ and we write:

$$\boldsymbol{R}_t|\Omega_t \sim \mathcal{D}\left(\boldsymbol{\mu}_t, \boldsymbol{\Sigma}_t\right), \tag{1}$$

where $\Omega_t$ is the available information set at period $t$. We need not make particular assumptions about the process of the returns, except that they have a conditional distribution but we provide some explicit results on selecting optimal weights for covariance averaging for the special case of i.i.d. returns with finite fourth moments, as in the covariance shrinkage literature (more details in subsequent sections; note that the i.i.d. assumption implies a constant covariance matrix $\boldsymbol{\Sigma}$). For the rest of our discussion we denote the suitably demeaned returns by $\boldsymbol{r}_t \stackrel{\text{def}}{=} \boldsymbol{R}_t - \widehat{\boldsymbol{\mu}}_t$, with $\widehat{\boldsymbol{\mu}}_t$ being a consistent estimator of the expected returns.

Given an increasing sample of $t$ observations, $\{\boldsymbol{r}_j\}_{j=1}^{t}$, we are interested in obtaining an accurate estimate $\widehat{\boldsymbol{\Sigma}}_t$ of the covariance matrix $\boldsymbol{\Sigma}_t$. Furthermore, we would like to do so without having to resort to a particular parametric model by using rolling window averaging, i.e., averaging across different segments of the data. This approach has been used successfully in forecasting applications, as discussed in the previous section, but it has not been used – to the best of our knowledge – in the current context of estimating a covariance matrix.

## 2.2 Covariance Averaging using Shrinkage

The idea of (model) averaging covariance estimates is implicitly embedded in the shrinkage-based methodology of Ledoit and Wolf (2003) and Ledoit and Wolf (2004), although they neither presented shrinkage as part of an averaging approach nor examined the performance of shrinkage estimation in a statistical context, their focus being exclusively on improving portfolio performance. Their approach is also different from what we present below, in that they consider a linear combination of an unstructured and a highly structured matrix, while in our context averaging takes place using the sample covariances computed over different observation segments. Still, covariance shrinkage is covariance averaging and hence we start from their seminal work on it.

The idea of covariance shrinkage is that a potentially improved estimator of the covariance matrix can be obtained by taking a linear combination of an estimator with no structure, e.g. the recursive sample covariance matrix $\widehat{\boldsymbol{\Sigma}}_t(t)$:

$$\widehat{\boldsymbol{\Sigma}}_t(t) \stackrel{\text{def}}{=} \frac{1}{t}\sum_{i=1}^{t} \boldsymbol{r}_i \boldsymbol{r}_i^{\top}, \tag{2}$$

and a highly structured estimator denoted here by $\boldsymbol{S}$. Regarding the choice of $\boldsymbol{S}$, we do not consider any highly structured estimator derived from a factor model but we use the covariance estimator of the constant correlation model. For more details on the choices of $\boldsymbol{S}$ we refer to Ledoit and Wolf (2004).

The linear combination, i.e., the averaged covariance, can be represented as:

$$\widehat{\boldsymbol{\Sigma}}_t^S(\delta) \stackrel{\text{def}}{=} \delta \boldsymbol{S} + (1-\delta)\widehat{\boldsymbol{\Sigma}}_t(t), \tag{3}$$

where $\delta \in (0,1)$ is the shrinkage (i.e., averaging) coefficient. The papers mentioned above suggest that the optimal choice for the shrinkage coefficient $\delta$ should minimize the distance between the shrinking covariance $\widehat{\boldsymbol{\Sigma}}_t^S$ estimate and the true covariance matrix $\boldsymbol{\Sigma}$, which is assumed constant. Note, furthermore, that Ledoit and Wolf (2003) and Ledoit and Wolf (2004) require i.i.d. returns with finite fourth moments in order to solve the minimization problem which follows. We use the same assumption, but it is not required in all of our subsequent methodological discussion. Formally, the optimal choice $\widehat{\delta}$ is obtained as a solution to the following Mean Squared Error type minimization problem:

$$\widehat{\delta} \stackrel{\text{def}}{=} \min_{\delta} \mathsf{E}\|\boldsymbol{\Sigma} - \widehat{\boldsymbol{\Sigma}}_t^S(\delta)\|_F^2, \tag{4}$$

where $\|\boldsymbol{A}\|_F^2 \stackrel{\text{def}}{=} \sum_{i=1}^{N}\sum_{j=1}^{N} a_{ij}^2$ is the Frobenius matrix norm. Letting $\sigma_{ij}$ and $\widehat{\sigma}_{ij}^S(t,\delta)$ denote the corresponding elements of the matrices in the above equation, we can easily see that:

$$\mathsf{E}\|\boldsymbol{\Sigma}-\widehat{\boldsymbol{\Sigma}}_t^S(\delta)\|_F^2 = \sum_{i=1}^{N}\sum_{j=1}^{N} \mathsf{E}\left[\sigma_{ij} - \widehat{\sigma}_{ij}^S(t,\delta)\right]^2 = \sum_{i=1}^{N}\sum_{j=1}^{N}\left\{\mathsf{Var}\left[\widehat{\sigma}_{ij}^S(t,\delta)\right] + \left(\mathsf{E}\left[\sigma_{ij} - \widehat{\sigma}_{ij}^S(t,\delta)\right]\right)^2\right\}. \tag{5}$$

Assuming that the averaged estimate $\widehat{\boldsymbol{\Sigma}}_t^S(\delta)$ is (almost) unbiased for the true covariance, and thus $\mathsf{E}\left[\sigma_{ij} - \widehat{\sigma}_{ij}^S(t,\delta)\right] = 0$, then the rightmost term with the square of the expectation drops out and we end up with:

$$\mathsf{E}\|\boldsymbol{\Sigma} - \widehat{\boldsymbol{\Sigma}}_t^S(\delta)\|_F^2 \approx \sum_{i=1}^{N}\sum_{j=1}^{N} \mathsf{Var}\left[\widehat{\sigma}_{ij}^S(t,\delta)\right], \tag{6}$$

i.e., the objective function to be minimized over the parameter $\delta$ is the sum of all (asymptotic) variances of the covariance estimate. For any given value of $\delta$ the above can be directly estimated under the assumptions (and in the notation) of Ledoit and Wolf (2003) and Ledoit and Wolf (2004) by:

$$\widehat{\pi}(\delta) \stackrel{\text{def}}{=} \sum_{i=1}^{N}\sum_{j=1}^{N} \widehat{\pi}_{ij}(\delta) where \quad \widehat{\pi}_{ij}(\delta) \stackrel{\text{def}}{=} \frac{1}{t}\sum_{h=1}^{t}\left[r_{ih}r_{jh} - \widehat{\sigma}_{ij}^S(t,\delta)\right]^2. \tag{7}$$

The above expressions are useful for our discussion of other forms of covariance averaging. Note that one can further elaborate the expression in Equation (4) on the basis of the form that $\boldsymbol{S}$ might take, and this is useful in finding different optimal $\delta$ estimates. For more details, see Ledoit and Wolf (2003) and Ledoit and Wolf (2004).

## 2.3 A General Framework for Covariance Averaging

There are both pros and cons in using the covariance averaging scheme based on shrinkage. The pros include the use of a structured matrix $\boldsymbol{S}$, which can be based on economic reasoning, such as factor models, the parametric parsimony of having only one parameter $\delta$ and the optimization based on a well-defined objective function. However, the way in which the shrinkage problem is formulated is actually amenable to further generalization; This is indeed its only con. We thus turn next to setting up a more general framework for covariance averaging based on rolling window covariance estimators.

The main idea of such covariance averaging is fairly intuitive: compute the sample covariance matrix using different segments of the data, either overlapping or non-overlapping, and average the resulting covariances. Let us consider the case of overlapping windows first, since this corresponds to rolling window averaging. Consider a sequence of overlapping windows $B \stackrel{\text{def}}{=} (m_1, m_2, \ldots, m_M)$ where $1 < m_1 < m_2 < \cdots < m_M < t$. The windows need not be equidistant and the lengths $m_j$ as well as the number of windows $M$ are assumed to be fixed in advance by the researcher. Using the last $m_s$ observations, the sample covariance is estimated in the standard fashion as:

$$\widehat{\boldsymbol{\Sigma}}_t(m_s) \stackrel{\text{def}}{=} \frac{1}{m_s} \sum_{i=t-m_s+1}^{t} \boldsymbol{r}_i \boldsymbol{r}^\top, \tag{8}$$

and once we have the $M$ covariance estimates from the different rolling windows we obtain the averaged covariance as:

$$\widehat{\boldsymbol{\Sigma}}_t^A \stackrel{\text{def}}{=} \sum_{s=1}^{M} w_s \widehat{\boldsymbol{\Sigma}}_t(m_s), \tag{9}$$

where $\{w_s\}_{s=1}^M$ are the averaging weights that obey:

$$w_s \in [0,1] with \ \ \sum_{s=1}^{M} w_s = 1. \tag{10}$$

Assigning different weights in averaging the covariances gives us a variety of different estimates and, therefore, the main problem is how to choose these weights - either heuristically or somehow optimally, as in the covariance shrinkage literature. We explore both approaches in what follows.

The simplest case is, naturally, to assign equal weights to all rolling estimates of the covariance matrices, which gives us:

$$w_s^E \stackrel{\text{def}}{=} \frac{1}{M}. \tag{11}$$

If one wants to assign greater weight to the most recent data then an exponentially weighted scheme can be used as in:

$$w_s^X(\alpha) \stackrel{\text{def}}{=} \frac{(1-\alpha)^{s-1}}{\sum_{s=1}^M (1-\alpha)^{s-1}}, \tag{12}$$

where $\alpha \in [0,1]$ is the smoothing parameter, whose optimal selection we discuss in the next section. For the time being we note that this can be pre-set to any desired value and we also experiment with two rules. The first rule sets $\alpha$ to the average of the standard deviations of the full sample of returns while the second rule is a scaled version of the first. We then have:

$$\widehat{\alpha}_1 \stackrel{\text{def}}{=} \frac{1}{N}\sum_{j=1}^N s_j and \qquad\qquad \widehat{\alpha}_2 \stackrel{\text{def}}{=} \frac{N+\widehat{\alpha}_1}{N+2}, \tag{13}$$

where $s_j$ is the estimate of the full-sample standard deviation of the $j^{th}$ asset[1]. According to the above rules we assign higher weight to the most recent segment of observations if the historical cross-asset volatility is relatively low and vice versa if it is high. This makes some intuitive sense: if over the period of observation we have periods of assets which show sudden changes in their returns (and thus higher volatility), then we should not be looking at

[1]Note that the $\widehat{\alpha}_1$ estimate of the smoothing parameter is, by construction, less than 1 and positive.

only the recent history in forming our covariance estimates, but should be combining longer segments to account for the higher past volatility. These two rules guide us quite well in practice.

The next approach that we consider is based on the shrinkage objective function adapted to the context of averaging. The idea is to assign weights based on expected distances from a target, the true covariance $\mathbf{\Sigma}$. This weighting approach is a variation on kernel smoothing and nearest neighbours combined. Consider thus the covariance estimate based on the $m_s$ window and write:

$$d_s \stackrel{\text{def}}{=} \mathsf{E}\|\mathbf{\Sigma} - \widehat{\mathbf{\Sigma}}_t(m_s)\|_F^2 = \sum_{i=1}^{N}\sum_{j=1}^{N} \mathsf{E}\left[\sigma_{ij} - \widehat{\sigma}_{ij}(m_s)\right]^2 \approx \sum_{i=1}^{N}\sum_{j=1}^{N} \mathsf{Var}\left[\widehat{\sigma}_{ij}(m_s)\right], \tag{14}$$

assuming the same condition of unbiasedness as in Equation (5) and Equation (6). Note that this does not depend on any parameter, such as the $\delta$ in shrinkage, and can be directly estimated as in Equation (6) by:

$$\widehat{\pi}(m_s) \stackrel{\text{def}}{=} \sum_{i=1}^{N}\sum_{j=1}^{N} \widehat{\pi}_{ij}(m_s) where \quad \widehat{\pi}_{ij}(m_s) \stackrel{\text{def}}{=} \frac{1}{m_s} \sum_{h=t-m_s+1}^{t} \left[r_{ih} r_{jh} - \widehat{\sigma}_{ij}(m_s)\right]^2. \tag{15}$$

These estimated distances are then used to construct weights for averaging which are inversely related to their magnitude: a higher distance being allocated a smaller weight and vice versa. For this we can use various heuristics such as the ones below:

$$\begin{aligned}
&(i)\lambda_s \stackrel{\text{def}}{=} d_s^{-1} and \quad w_s^D \stackrel{\text{def}}{=} \frac{\lambda_s}{\sum_{s=1}^{M} \lambda_s},\\
&(ii)\lambda_s \stackrel{\text{def}}{=} \frac{\sum_{j\neq s}^{M} d_j}{\sum_{j=1}^{M} d_j} and \quad w_s^D \stackrel{\text{def}}{=} \frac{\lambda_s}{\sum_{s=1}^{M} \lambda_s},\\
&(iii)\kappa_s \stackrel{\text{def}}{=} \exp\left[\frac{d_s}{\sum_{s=1}^{M} d_s}\right], \quad \lambda_s \stackrel{\text{def}}{=} \frac{\sum_{j\neq s}^{M} \kappa_j}{\sum_{j=1}^{M} \kappa_j} and \quad w_s^D \stackrel{\text{def}}{=} \frac{\lambda_s}{\sum_{s=1}^{M} \lambda_s}.
\end{aligned} \tag{16}$$

The above weighting schemes have an intuitive interpretation, although we treat them as completely heuristic rules – we see, however, that they work quite well in practice. The first weighting scheme in the above equation is the simplest one: assign greater weight to the covariance estimate which has the smallest distance from the true covariance $\boldsymbol{\Sigma}$. The second weighting scheme gathers together all the distances and assigns weights which relate to what the other distances are: if the relative sum of distances of windows other than $s$ is large, then the estimate based on this particular window should receive higher weight. The third weighting scheme is based on pre-processing the relative distances: first, those windows that have higher relative distances are 'pumped-up' (by the exponential function) and their magnitudes are exaggerated; then the weights are computed according to the second weighting scheme.

## 2.4 Selecting Optimal Weights

### 2.4.1 Solving the General Problem

While all the above approaches for averaging are intuitive and straightforward to implement one cannot but ask whether the weights $\{w_s\}_{s=1}^{M}$ can be optimized as the shrinkage parameter $\delta$ is optimized. The answer is yes, under the same assumptions as are used in shrinkage. We consider two approaches for this case of optimal weights: in one approach we let the weights simply obey the conditions of Equation (10) and in the other approach we parameterize them using the exponential smoothing weights of Equation (12). Letting $\boldsymbol{w} \stackrel{\text{def}}{=} [w_1, w_2, \ldots, w_M]^\top$ denote the $(M \times 1)$ vector of weights, we have the general set-up for the optimization problem being given by the same objective function used in the covariance shrinkage as in:

$$Q(\boldsymbol{w}) \stackrel{\text{def}}{=} \mathsf{E}\left\|\boldsymbol{\Sigma} - \widehat{\boldsymbol{\Sigma}}_t^A\right\|_F^2 = \mathsf{E}\left\|\boldsymbol{\Sigma} - \sum_{s=1}^{M} w_s \widehat{\boldsymbol{\Sigma}}_t(m_j)\right\|_F^2. \tag{17}$$

Expanding the above using the variance and mean-square decomposition and using similar notation to Equation (5) on shrinkage we get:

$$\begin{aligned} Q(\boldsymbol{w}) &= \textstyle\sum_{i=1}^N \sum_{j=1}^N \left\{\mathsf{Var}\left[\widehat{\sigma}_{ij}^A(t)\right] + \left(\mathsf{E}\left[\sigma_{ij} - \widehat{\sigma}_{ij}^A(t)\right]\right)^2\right\} \\ &= \textstyle\sum_{i=1}^N \sum_{j=1}^N \left\{\sum_{s=1}^M w_s^2 \mathsf{Var}\left[\widehat{\sigma}_{ij}(m_s)\right] + 2\sum_{k\neq s} w_k w_s \mathsf{Cov}\left[\widehat{\sigma}_{ij}(m_k), \widehat{\sigma}_{ij}(m_s)\right]\right\} \\ &\quad + \textstyle\sum_{i=1}^N \sum_{j=1}^N \left(\mathsf{E}\left[\sigma_{ij} - \sum_{s=1}^M w_s \widehat{\sigma}_{ij}(m_s)\right]\right)^2, \end{aligned} \tag{18}$$

which now has one extra term because of the presence of the covariances and another extra term (the last one) because of averaging the bias component. The latter term is easily dealt with: if we assume unbiasedness, then the term can be omitted and if not, then this term can be estimated; we do both in the simulations and application. The first term, however, is more of an issue. It creates a potential problem, since these covariances are non-zero because of the use of overlapping data, even when the data are assumed to be i.i.d. To avoid keeping track of the non-zero elements, and to minimize the computational burden, we convert the averaging scheme into one involving non-overlapping data segments at the (trivial) expense of re-expressing the weights. In this way we can eliminate the presence of the covariance terms and then proceed to estimation and optimization.

To see how the above works, consider the simple case of $M = 2$ and note that we have the following representations:

$$\begin{aligned} \widehat{\mathbf{\Sigma}}_t(m_1) &= m_1^{-1} \textstyle\sum_{i=t-m_1+1}^{t} \boldsymbol{r}_i \boldsymbol{r}_i^\top, \\ \widehat{\mathbf{\Sigma}}_t(m_2) &= m_2^{-1} \textstyle\sum_{i=t-m_2+1}^{t} \boldsymbol{r}_i \boldsymbol{r}_i^\top. \end{aligned} \tag{19}$$

Then the second covariance, which depends on more terms than the first, can be written as:

$$\begin{aligned} \widehat{\mathbf{\Sigma}}_t(m_2) &= m_2^{-1} \textstyle\sum_{i=t-m_2+1}^{t-m_1} \boldsymbol{r}_i \boldsymbol{r}_i^\top + m_2^{-1} \displaystyle\sum_{i=t-m_1+1}^{t} \boldsymbol{r}_i \boldsymbol{r}_i^\top \\ &= [(m_2 - m_1)/m_2]\, \widehat{\mathbf{\Sigma}}_t(m_2 - m_1) + (m_1/m_2)\widehat{\mathbf{\Sigma}}_t(m_1), \end{aligned} \tag{20}$$

and the second covariance is now composed of two covariances which are estimated by non-overlapping data, at the expense of different weights, since we can now write:

$$\begin{aligned} \widehat{\mathbf{\Sigma}}_t^A &= [w_1 + w_2(m_1/m_2)]\, \widehat{\mathbf{\Sigma}}_t(m_1) + w_2\, [(m_2 - m_1)/m_2]\, \widehat{\mathbf{\Sigma}}_t(m_2 - m_1) \\ &= a_1 \widehat{\mathbf{\Sigma}}_t(m_1) + a_2 \widehat{\mathbf{\Sigma}}_t(m_2 - m_1). \end{aligned} \tag{21}$$

Note that in the new weighting scheme we still satisfy the conditions of Equation (10), i.e., $a_s \in [0, 1]$ and $a_1 + a_2 = 1$. Now, however, when the above averaged covariance enters into the objective function of Equation (18), we will not have covariance terms such as $\mathsf{Cov}\,[\widehat{\sigma}_{ij}(m_1), \widehat{\sigma}_{ij}(m_2 - m_1)]$ since the individual covariances are now estimated from non-overlapping data segments.

We can easily generalize the above discussion when $M > 2$ since it involves only book-keeping on the way in which the sample covariances are converted and the new weights behave. Noticing that the new weights $a_s$ depend on some of the old weights $w_s$ we first define the new weights formally as:

$$
\begin{array}{lll}
a_1(\boldsymbol{w}_0) & \stackrel{\text{def}}{=} & \sum_{s=1}^{M} \frac{m_1}{m_s} w_s \text{ with } \boldsymbol{w}_0 = [w_1, \ldots, w_M]^\top, \\
a_2(\boldsymbol{w}_{-1}) & \stackrel{\text{def}}{=} & \sum_{s=2}^{M} \frac{m_2 - m_1}{m_s} w_s \text{ with } \boldsymbol{w}_{-1} = [w_2, \ldots, w_M]^\top, \\
a_3(\boldsymbol{w}_{-2}) & \stackrel{\text{def}}{=} & \sum_{s=3}^{M} \frac{m_3 - m_2}{m_s} w_s \text{ with } \boldsymbol{w}_{-2} = [w_3, \ldots, w_M]^\top, \\
\vdots & \vdots & \vdots \\
a_M(\boldsymbol{w}_{-M+1}) & \stackrel{\text{def}}{=} & \frac{m_M - m_{M-1}}{m_M} w_M \text{ with } \boldsymbol{w}_{-M+1} = w_M,
\end{array}
\tag{22}
$$

and then (re)define the non-overlapping sample covariances as:

$$
\begin{array}{lll}
\widehat{\boldsymbol{\Sigma}}_t(m_1) & \stackrel{\text{def}}{=} & m_1^{-1} \sum_{i=t-m_1+1}^{t} \boldsymbol{r}_i \boldsymbol{r}_i^\top, \\
\widehat{\boldsymbol{\Sigma}}_t(m_s - m_{s-1}) & \stackrel{\text{def}}{=} & (m_s - m_{s-1})^{-1} \sum_{i=t-m_{s-1}+1}^{t-m_s} \boldsymbol{r} \boldsymbol{r}_i^\top,
\end{array}
\tag{23}
$$

for $s = 2, \ldots, M$ with $m_0 \equiv 0$. With these we can re-write the objective function of Equation (18) as:

$$
\begin{aligned}
Q(\boldsymbol{w}) &= \mathsf{E} \left\| \boldsymbol{\Sigma} - \sum_{s=1}^{M} a_s(\boldsymbol{w}_{-s+1}) \widehat{\boldsymbol{\Sigma}}_t(m_s - m_{s-1}) \right\|_F^2 \\
&= \sum_{i=1}^{N} \sum_{j=1}^{N} \left\{ \sum_{s=1}^{M} a_s^2(\boldsymbol{w}_{-s+1}) \mathsf{Var}\left[\widehat{\sigma}_{ij}(m_s - m_{s-1})\right] \right\} \\
&\quad + \sum_{i=1}^{N} \sum_{j=1}^{N} \left( \mathsf{E}\left[ \sigma_{ij} - \sum_{s=1}^{M} a_s(\boldsymbol{w}_{-s+1}) \widehat{\sigma}_{ij}(m_s - m_{s-1}) \right] \right)^2,
\end{aligned}
\tag{24}
$$

which does not involve the covariance terms $\mathsf{Cov}\left[\widehat{\sigma}_{ij}(m_1), \widehat{\sigma}_{ij}(m_2 - m_1)\right]$. The optimization of the objective function is most easily done numerically: first we construct the estimates for the variance terms $\mathsf{Var}\left[\widehat{\sigma}_{ij}(m_s - m_{s-1})\right]$, the bias terms $\mathsf{E}\left[\sigma_{ij} - \sum_{s=1}^{M} a_s(\boldsymbol{w}_{-s+1}) \widehat{\sigma}_{ij}(m_s - m_{s-1})\right]$ and the composite weights $a_s(\boldsymbol{w}_{-s+1})$ (for a given value of the original weights $w_s$); then we impose the restrictions on the weights (either the original or the composite) of Equation (10) and optimize the objective function directly. As noted above, we do not necessarily have to impose unbiasedness during the computations, although under the i.i.d. assumption and the properties of the composite weights we can work without the bias terms.

We can take the above scheme a step further and show – as promised – how an optimal value for the smoothing parameter $\alpha$ can be obtained in this context. Both for reasons of parsimony, and to economize on the computations, we can easily adapt the above for a case where we parameterize the weights via the $\alpha$ parameter in the exponentially weighted scheme of Equation (12). The only thing that changes is the mapping from $w_s \mapsto a_s(\boldsymbol{w}_{-s+1})$ to one from $\alpha \mapsto w_s(\alpha) \mapsto a_s(\boldsymbol{w}_{-s+1}, \alpha)$. In this way the weights depend on a single parameter $\alpha$, which can be now optimized for use in applications. Note, however, that no explicit solution is available when the exponential weights of Equation (12) are used. In the case of the general optimization problem there is an explicit solution for the (new) weights $a_s(\boldsymbol{w}_{-s+1})$, to which we next turn.

### 2.4.2 Interpretation of Optimal Weights: Penalize Higher Estimation Risk

While the above solves the practical problem easily, and without further algebra, it would be useful to have an explicit formulation and interpretation of the weights which come from averaging. The transformation from $w_s \mapsto a(\boldsymbol{w}_{-s+1})$ allows us to have this and, in the process, obtain an explicit expression of these composite weights which is amenable to a favorable interpretation. Let us start with the first term in Equation (24), i.e. ignoring the bias terms, and pass the double-summation inside the curly brackets to obtain:

$$Q(\boldsymbol{a}) = \sum_{s=1}^{M} a_s^2(\boldsymbol{w}_{-s+1})\pi(m_s - m_{s-1}) = \boldsymbol{a}^\top \boldsymbol{\Pi} \boldsymbol{a}, \tag{25}$$

where we define $\boldsymbol{a} \stackrel{\text{def}}{=} [a_1(\boldsymbol{w}_0), a_2(\boldsymbol{w}_{-1}), \ldots, a_M(\boldsymbol{w}_{-M+1})]^\top$ as the vector of composite weights and $\mathbf{\Pi} \stackrel{\text{def}}{=} \mathsf{diag}\,[\pi(m_1), \pi(m_2 - m_1), \ldots, \pi(m_M - m_{M-1})]$ as the diagonal matrix of the sum of asymptotic variances. This is a quadratic form which is to be minimized with respect to the weights $\boldsymbol{a}$ in the following constrained problem:

$$\Lambda(\boldsymbol{a}) = \boldsymbol{a}^\top \mathbf{\Pi} \boldsymbol{a} + 2\lambda(1 - \boldsymbol{e}^\top \boldsymbol{a}), \tag{26}$$

where $\boldsymbol{e}$ is a vector of ones and we see that imposing positivity in the weights is not necessary. The above Lagrangian equation has a known, explicit solution which is similar to the GMV portfolio weights, namely:

$$\boldsymbol{a}^* \stackrel{\text{def}}{=} \operatorname{argmin} \Lambda(\boldsymbol{\alpha}) \equiv \frac{\mathbf{\Pi}^{-1}\boldsymbol{e}}{\boldsymbol{e}^\top \mathbf{\Pi}^{-1} \boldsymbol{e}}, \tag{27}$$

which implies, given the diagonal structure of $\mathbf{\Pi}$, that:

$$a_s^* \stackrel{\text{def}}{=} \frac{\pi^{-1}(m_s - m_{s-1})}{\sum_{s=1}^{M} \pi^{-1}(m_s - m_{s-1})}, \tag{28}$$

i.e., the weights assigned to the rolling window covariances for averaging are inversely proportional to the asymptotic variances (note that these weights are, by construction, always positive). This implies that higher estimation risk, vis--vis the true covariance, leads to a lower weight in constructing the averaged estimate. This is an intuitive and desirable characteristic, since it implies that greater weight is given to the more accurate estimate. Furthermore, this result justifies the heuristics presented in Equation (16), which work essentially on the same premise but using one estimate at a time.

Extending the above result when including the bias term in Equation (24) leads to more complicated algebra but with the same essential intuitive result. To see this, change the Lagrangian to:

$$\Lambda(\boldsymbol{a}) = \boldsymbol{a}^\top \boldsymbol{\Pi} \boldsymbol{a} + \sum_{i=1}^{N} \sum_{j=1}^{N} \left[\sigma_{ij} - \boldsymbol{E}_{ij}^\top \boldsymbol{a}\right]^2 + 2\lambda(1 - \boldsymbol{e}^\top \boldsymbol{a}), \tag{29}$$

where we define the vectors $\boldsymbol{E}_{ij} \stackrel{\text{def}}{=} \mathsf{E}\left[\widehat{\sigma}_{ij}(m_1), \widehat{\sigma}_{ij}(m_2 - m_1), \ldots, \widehat{\sigma}_{ij}(m_M - m_{M-1})\right]^\top$. Solving for the first order conditions, we obtain:

$$\begin{aligned}
\tfrac{1}{2}\frac{\partial \Lambda(\boldsymbol{a})}{\partial \boldsymbol{a}} &= \boldsymbol{a}^\top \boldsymbol{\Pi} - \sum_{i=1}^{N} \sum_{j=1}^{N} \boldsymbol{E}_{ij}^\top \left[\sigma_{ij} - \boldsymbol{E}_{ij}^\top \boldsymbol{a}\right] - \lambda \boldsymbol{e}^\top \\
&= \left(\boldsymbol{\Pi} + \sum_{i=1}^{N} \sum_{j=1}^{N} \boldsymbol{E}_{ij} \boldsymbol{E}_{ij}^\top\right) \boldsymbol{a} - \sum_{i=1}^{N} \sum_{j=1}^{N} \boldsymbol{E}_{ij} \sigma_{ij} - \lambda \boldsymbol{e} \\
&= \boldsymbol{V} \boldsymbol{a} - \boldsymbol{b} - \lambda \boldsymbol{e},
\end{aligned} \tag{30}$$

where we define $\boldsymbol{V} \stackrel{\text{def}}{=} \left(\boldsymbol{\Pi} + \sum_{i=1}^{N} \sum_{j=1}^{N} \boldsymbol{E}_{ij} \boldsymbol{E}_{ij}^\top\right)$ and $\boldsymbol{b} \stackrel{\text{def}}{=} \sum_{i=1}^{N} \sum_{j=1}^{N} \boldsymbol{E}_{ij} \sigma_{ij}$. Using the first order conditions for the Lagrange multiplier we end up with the new solution:

$$\boldsymbol{a}^* = \operatorname{argmin} \Lambda(\boldsymbol{a}) = \boldsymbol{V}^{-1} \boldsymbol{b} + \left(1 - \boldsymbol{e}^\top \boldsymbol{V}^{-1} \boldsymbol{b}\right) \frac{\boldsymbol{V}^{-1} \boldsymbol{e}}{\boldsymbol{e}^\top \boldsymbol{V}^{-1} \boldsymbol{e}}. \tag{31}$$

Note that the new weights have three parts: first, there is a constant term $\boldsymbol{V}^{-1}\boldsymbol{b}$; second, there is a (scalar) slope term $\left(1 - \boldsymbol{e}^\top \boldsymbol{V}^{-1} \boldsymbol{b}\right)$; and, third, there is the main term whose structure resembles the structure of the weights in Equation (27). Note that when we do not take into account the bias term, the solution in the above equation collapses to that of Equation (27) and, therefore, that both equations have the same interpretation.

# 3 Simulation Analysis

As we present no analytical results concering the potential and efficacy of covariance averaging we start our empirical examination using a simulation experiment. We first consider the data generating process (DGP) of Patton and Sheppard (2009), which allows for time-varying covariances in the spirit of a multivariate GARCH-type model and also for DGP-consistent realized covariances to be computed. This latter property is important, since realized dispersion and correlation measures have been proven to be the state-of-the-art when the appropriate data are available. Then, we consider another DGP which conforms a little more closely to the idea of covariance averaging and is amenable to analysis where $N$ is large. In the sections below, we present the simulation models and their parameterizations and we then discuss our simulation results.

## 3.1 DGP #1: Bivariate GARCH-type Model

We take $N = 2$ so we have a bivariate system and now $\boldsymbol{r}_t$ denotes the $(2 \times 1)$ vector of (zero mean) returns. The structure of the DGP is then given as follows:

$$
\begin{aligned}
\boldsymbol{r}_t &\stackrel{\text{def}}{=} \boldsymbol{\Sigma}_t^{1/2}\boldsymbol{\epsilon}_t, \\
\boldsymbol{\epsilon}_t &\stackrel{\text{def}}{=} \textstyle\sum_{k=1}^{78} \boldsymbol{\xi}_{kt} \quad \text{with} \quad \boldsymbol{\xi}_{kt} \sim N(0, 78^{-1}), \\
\boldsymbol{\Sigma}_t &\stackrel{\text{def}}{=} 0.05\bar{\boldsymbol{\Sigma}} + 0.85\boldsymbol{\Sigma}_{t-1} + 0.10\boldsymbol{r}_{t-1}\boldsymbol{r}_{t-1}^{\top},
\end{aligned}
\tag{32}
$$

with $\bar{\boldsymbol{\Sigma}}$ as the unconditional covariance matrix (with unit diagonal). We consider five different values for the contemporaneous correlation between the two variables, i.e., the off-diagonal element of $\bar{\boldsymbol{\Sigma}}$, namely, -0.9, -0.5, 0.0, +0.5, +0.9. For each of these values, and for the other parameter values fixed as in the above equation, we proceed as follows. For each replication $R$:

1. Generate an initial sample of size $t^* = t_0+t+\tau$ and discard the pre-sample observations $t_0$.

2. Using only the $t$ in-sample observations estimate and/or forecast the various covariances, denoted generically by $\widehat{\mathbf{\Sigma}}_t^s(r)$, for method $s$. Then, compute the ratio of the distance of the covariance estimates vis--vis the true model covariance for all available values of $\tau$, i.e.,:

$$RD_R^s(h) \stackrel{\text{def}}{=} \frac{\sum_{r=1}^{R} \|\mathbf{\Sigma}_{t+h} - \widehat{\mathbf{\Sigma}}_t^s(r)\|_F^2}{\sum_{r=1}^{R} \|\mathbf{\Sigma}_{t+h} - \widehat{\mathbf{\Sigma}}_t(t)\|_F^2}, forh = 1, 2, ..\tau. \tag{33}$$

3. Repeat the above steps for a number of $R = 1000$ replications and then compute the average ratio, i.e.,:

$$\bar{D}_R^s(h) \stackrel{\text{def}}{=} \frac{1}{R}\sum_{r=1}^{R} RD_R^s(h). \tag{34}$$

This last statistic is what we report across different selections for $B$ vthe number and lengths of rolling windows and the different types of covariance estimate. A value less than one indicates that the corresponding estimator is better than the unconditional sample covariance one. We use the following combinations for $B$: (5,20,50); (5,20,50,100); (50,100,200,400); (5,20,50,100,200,400). The results from this model are summarized in Table 1 through Table 5.

### 3.2 DGP #2: Model of Weighted Past Returns

We next consider a simpler model, which has no realized covariance terms but which allows for an arbitrarily large number of assets to enter. It is based on a finite, exponential weighted scheme of past returns to generate the covariance. As in the previous model set-up, we also effect an unbiasedness correction with the unconditional covariance $\bar{\boldsymbol{\Sigma}}$ – although note that here this unconditional covariance is not the same as before and is determined by the pre-sample values used to initialize the recursion (details are available on request). The form of the model now is:

$$
\begin{aligned}
\boldsymbol{r}_t &\stackrel{\text{def}}{=} \boldsymbol{\Sigma}_t^{1/2}\boldsymbol{\epsilon}_t \quad \text{with} \quad \boldsymbol{\epsilon}_t \sim N(0, \boldsymbol{I}_N), \\
\boldsymbol{\Sigma}_t^* &\stackrel{\text{def}}{=} \textstyle\sum_{j=1}^{M} w_j^X(\alpha)\boldsymbol{r}_{t-j}\boldsymbol{r}_{t-j}^\top, \\
\boldsymbol{\Sigma}_t &\stackrel{\text{def}}{=} 0.1\bar{\boldsymbol{\Sigma}} + 0.9\boldsymbol{\Sigma}_t^*,
\end{aligned} \tag{35}
$$

where $w_j^X(\alpha)$ are the exponential weights of Equation (12), with $\alpha$ fixed at $\alpha = 0.9$. We follow a similar set-up as in the previous model for evaluating the performance of the various estimates and we consider two cases for $N$, $N = 5$ and $N = 50$, and the same combinations for $B$ as before. The results for this model are summarized in Table 6 and Table 7. Finally, in Tables 8 through 10 we present the 'winning' (best) methods from both DGP and the results from Table 1 through Table 7.

### 3.3 Estimators used in Simulations

Before embarking on a discussion of our simulation results, let us summarize the covariance estimators that we used in the order that they appear in the relevant tables.

The results for the full sample covariance estimator are not presented since this estimator serves as a numeraire, which is used to evaluate the performance of the rest. For the first simulation model, the bivariate GARCH-type model, we can compute the realized covariance estimator, as in Patton and Sheppard (2009). We then compute Ledoit and Wolf's shrinkage estimators based on the full and (the last) half-sample (LW Shrinkage, F and H) and the sample covariance estimator based on the (last) half-sample (Sample Covariance H). Our use of estimator using half the full observations per simulation is to examine whether the length of the sample (most of all in this time-varying context) affects the performance of the sample covariance and the shrinkage estimators. Then we have a sequence of estimators based on covariance averaging:

1. The estimator based on equal weights, from Equation (11) (Equal Weights).

2. The estimators based on exponential weights and the heuristics of Equation (12) and Equation (13) (EMA weights 1 and 2 respectively).

3. The estimators based on unrestricted optimal weights of Equation (18), with and without a bias correction (Optimized Weights 1 and 2 respectively).

4. The estimators based on the restricted optimal weights of Equation (18), parameterized via the exponential weighting scheme, with and without a bias correction (Optimized EMA weights 1 and 2 respectively).

5. The estimators based on the unrestricted weights of Equation (31), with bias correction (Optimized $a^*$).

6. The estimators based on the distance weights of Equation (14), Equation (15) and Equation (16) (Distance weights 1, 2 and 3 respectively).

### 3.4 Discussion of Simulation Results

It is important to stress beforehand that our simulation analysis focuses exclusively on the statistical performance of the covariance average estimators. We leave for the next section the economic evaluation of these estimators based on portfolio construction with historical data.

A prominent feature of all the tables, from Table 1 to Table 7, is that there is always at least one covariance average estimator which outperforms the corresponding benchmarks - save the realized covariance estimator at short periods ahead. The interesting part is, of course, to see whether there is one or a group of estimators that consistently delivers better performance than the competitors. We set out a yardstick at the equally weighted estimator as follows: if this estimator is within the Top-3 estimators (including +5 ppt from the top third performer) then we declare it a 'winner' on grounds of parsimony. We make this comparison by looking across the different combinations of $B$ window widths and $h$ forecast steps ahead. With this in mind, we can assess the results in the tables to see which estimator works consistently better.

The one generic and consistent result that we see is this: while for DGP #1 one might as well use the equally weighted covariance estimator, since it is almost always a Top-3 performer, the same is not true for DGP #2, where one of the other weighting schemes is always better. This is, of course, true if one does not have access to a realized covariance estimator for DGP #1, which has the best performance overall for $h = 0$ (but not for $h > 0$). In addition, the sample covariance and shrinkage estimators are always inferior to one or more of the proposed weighted estimators. These results suggest that covariance averaging is a potentially powerful way of computing robust covariance matrices, in larger dimensions in particular. We note that, in a realistic application where $N$ is large, the full parametric estimation of DGP #1 becomes increasingly difficult and possibly unattainable (i.e. when one specifies full parameter matrices and not merely the scalars used here).

With this generic result to hand, we need to probe further the performance of the estimators which we suggest. We are particularly interested in whether the optimization schemes for getting the weights work better or worse than the heuristics schemes, based on distances, and which of these two categories appears to perform better overall. To do this accurately, we discuss separately the results for DGP #1 and those for DGP #2. For the first simulation design we look at the type of averaging estimator that is either better than or close to the equally weighted estimator (should there be a tie between an optimizing and an heuristic estimator, we prefer the second type). The results, from Table 1 through Table 5, clearly support the use of the heuristics, distance-based weights. Among all instances of the DGP #1 we see that one of the optimizing estimators is better than the heuristic weighting schemes only 19% of the time. Among these heuristics, and for this simulation design, we find that those based on the second and third formulas of Equation (16) perform best. Note that although these heuristics are on average better, they may not differ substantially in performance from one of the optimizing schemes.

We next turn to the second simulation design, DGP #2. Here, as already noted, the equally weighted estimator is not the best performer but is outdone by one of the other weighting schemes. As in the case of DGP #1, we want to examine which type of estimator is better, heuristic or optimizing. In this design we have the more practically relevant case of the presence of a large number of variables. Here the results are more in favor of the optimizing weighting schemes. If we look at both Table 6 and Table 7, we find that in 62%

of the cases examined an optimizing weighting scheme is the top performer. Furthermore, if we look only at Table 6, there is a tie in our results, with the optimizing schemes being top performers for 50% of the time, and if we look only at Table 7 we find that the optimizing schemes are top performers 75% of the time. Therefore, on the basis of this set of results, one might have a preference for using an optimizing weighting scheme for problems involving a larger number of variables. Among the optimizing schemes the ones based on the use of exponential weights of Equation (12) and the ones based on the direct weights of Equation (28) and Equation (31) appear to be the best ones to use for this DGP #2.

Finally, Table 8 through Table 10 provide us with yet another summary on performance. In them we present the proportion of times that each method is a nominal winner across each DGP (Table 8 and Table 9) and across both DGPs (Table 10). What we discuss above shows clearly in these tables, namely, that for DGP #1 the heuristics work better and that for DGP #2 the optimizing schemes work better. Table 10 provides a complete picture on these success ratios.

# 4 Empirical Application: GMV Portfolio Allocation

While the fitting and forecasting performance of covariance averaging was found, in the simulations, to be high and robust, we still need to submit the proposed approach to a real-life/real-data test. In this section we turn to the evaluation of the proposed methodology using real data and one particular portfolio optimization approach, the Global Minimum Variance (GMV). The aim here is to make comparisons with other methods for estimating covariances and not on seeing on which portfolio allocation method produces the best results. So we continue with the simplest GMV approach on which we only need as input the covariance matrix of the returns and nothing else. In the next subsection we briefly present the exact approach that we use, including our methods of rebalancing. Then we discuss our evaluation approach across the many different portfolios that we consider and the data that we used. All our results are collectively discussed in the final subsection.

## 4.1 GMV Portfolio and Rebalancing Approaches

Let $\boldsymbol{x}_\tau$ denote the portfolio weights to be chosen at time $\tau$ and let $\widehat{\boldsymbol{\Sigma}}_\tau$ be the covariance estimate available in the same time period; in addition, let the weights obey lower and upper bounds $[b_L, b_U]$ respectively and sum up to one. The GMV portfolio allocation problem is then given as follows:

$$\begin{aligned} \text{Minimize:} \quad & \boldsymbol{x}_\tau^\top \widehat{\boldsymbol{\Sigma}}_\tau \boldsymbol{x}_\tau \\ \text{subject to:} \quad & \boldsymbol{e}^\top \boldsymbol{x}_\tau = 1 \\ & 0 \leq b_L \leq x_{\tau,i} \leq b_U \leq 1, \end{aligned} \tag{36}$$

where $\boldsymbol{x}_\tau$ changes at rebalancing/re-optimization times $\tau = t_{j_1}, t_{j_2}, \ldots$ where $j_1, j_2, \ldots$ are the points in time when rebalancing/re-optimization takes place. Note that we restrict our attention to long-only positions as $b_L \geq 0$. The solution to this problem does not have a closed form, so we employ numerical optimization every time we re-optimize the weights. After the weights are available we compute the portfolio return in the usual fashion and track its evolution until a new rebalancing or re-optimization point is reached.

In our analysis we try to go a step beyond the evaluation of different covariance estimates; we also examine the effects of different types of rebalancing, namely, time and threshold rebalancing. Time and threshold rebalancing are not usually used in the academic literature but they have attracted some attention from practitioners in the industry: see, for example, Smith Barney Consulting Group (2005), Sun et al. (2006), Jaconetti et al. (2010), Tunc and Kozat (2012) and Moallemi and Saglam (2013).

The first type is based on keeping the existing weights for a fixed time frame and then rebalancing but not necessarily re-optimizing them. For example, suppose that we re-optimize every month but we rebalance every ten trading days. The second type tracks the evolution of individual weights in the portfolio and rebalances if any of these exceeds a prespecified threshold. All in all we consider four different types of rebalancing, including re-optimization: optimization only (or no-rebalancing in the interim period); rebalancing based on a time threshold; rebalancing based on a weight threshold; and a combination of time-and-threshold rebalancing. Below we present in sequence the steps that we use in our GMV portfolio optimization and rebalancing approaches.

1. Start with an initial wealth of 1 million dollars and set the transaction cost of buying/selling one share at 0.005 dollars[2]. At the beginning of the algorithm we decide on the re-optimization period every $E$ trading days.

2. Using an in-sample rolling window of $n_{roll} = m_M + 1$ price observations, we calculate the returns and the covariance estimators and optimize the portfolio weights.

3. We find the number of shares that can be purchased, the positions we need to open/close, and then calculate the overall portfolio transaction costs.

4. In-between $E$ portfolio re-optimizations, the time and threshold rebalancing takes

[2]This is the average cost that one of the major online brokerage houses charges the individual investor.

place.

5. For every period and using the $n_{roll} + j$ out-of-sample prices and the corresponding returns we calculate the overall portfolio value, its return and the wealth that is carried over in the next round of re-optimization and rebalancing.

6. Using the new wealth we start again from step 2 of the algorithm.

This is the generic set-up employed. Below are some additional details on the time and threshold rebalancing which takes place in step 4 above.

R1. *Time Rebalancing.*

R1a. Set the number of days, $R_{TR}$, that we want to rebalance the portfolio weights in between the $E$ days of re-optimization. By definition, $R_{TR} < E$ and in the special case where $R_{TR} = E$ only re-optimization takes place.

R1b. We return the weights in their initial values every $R_{TR}$ days. Transaction costs are calculated and the investor's current wealth is again computed.

R1c. The above procedure is carried out $\lfloor E/R_{TR} \rfloor$ times within the $E$ trading days; where $\lfloor \bullet \rfloor$ denotes the integer part.

R2. *Threshold Rebalancing.*

R2a. Set the threshold parameter, $R_{ThR}$, in terms of a maximum allowable percentage change in any of the portfolio weights. We rebalance our portfolio every time that one (or more) of our daily percentage change of the portfolio weights exceeds the above threshold parameter.

R2b. Then all transaction costs are calculated and the current wealth is computed. All the procedures are repeated as above.

R2c. Note that threshold rebalancing may not take place in between $E$ trading days if the threshold parameter $R_{ThR}$ is not exceeded.

R3. *Time and Threshold Rebalancing.* We combine the two approaches, where rebalancing takes place either because the time period $R_{TR}$ is reached or the threshold parameter $R_{ThR}$ is exceeded.

## 4.2 Parameterizations and Portfolio Evaluation

All the above combinations analyzed in the previous sections result in a huge amount of results: we have different covariance estimators, different portfolio compositions (see the Portfolio Data section below), different rebalancing approaches, different re-optimization periods and different values for the bounds on the weights. We summarize these combinations below:

- The bounds on the weights $\boldsymbol{x}_\tau$ are set for three different intervals $[b_L, b_U]$: [0,0.1]; [0,0.25] and [0,1].
- The rebalancing parameters are set as follows:
  1. $E = 20$, $R_{TR} = 5$, $R_{ThR} = 10\%$,
  2. $E = 60$, $R_{TR} = 20$, $R_{ThR} = 10\%$,
  3. $E = 180$, $R_{TR} = 60$, $R_{ThR} = 10\%$,
  4. $E = 20$, $R_{TR} = 5$, $R_{ThR} = 5\%$,

5. $E = 60$, $R_{TR} = 20$, $R_{ThR} = 5\%$,

6. $E = 180$, $R_{TR} = 60$, $R_{ThR} = 5\%$.

- Therefore, the total number of runs for each combination of the above is $C_B \stackrel{\text{def}}{=} 36 \times N_B$, where $N_B$ is the number of rolling window combinations that we have.

Notice that because our data are collected daily, the above parameters are set weekly (5 observations), monthly (20 observations), quarterly (60 observations) and semi-annually (180 observations). All these $C_B$ combinations of output need to be summarized in some useful way for understanding whether, on average, the new method of covariance estimation that we propose works better than the benchmarks. We use a meta-data analysis approach based on aggregation across portfolio cases and methods, as follows.

1. For each portfolio case, say $i$, we compute the performance measures (cumulative return, Sharpe ratio, maximum drawdown etc.). Let a representative such measure be called $P_{ijk}$ when is based on the rebalancing method $j$ and estimation approach $k$.

2. We compute three types of averaged statistics (success rates), based on the way that we treat the performance measures (where we evaluate first) and we aggregate them:

   (a) For the first statistic (ALL) we pool the data for $P_{ijk}$ across all $i$ (all portfolio combinations considered) and across all $j$ for each $k$ (all rebalancing methods are jointly considered for each estimation method); then we compare performance across different $k$ (compare performance of our estimation method against the other estimation approaches). In this approach we have the largest sample size of our meta-data for computing statistics, since we pool portfolios and rebalancing methods together.

(b) For the second statistic (WITHIN) we take the data for each $i$ (each portfolio is considered separately) and within each portfolio we pool across all $j$ for each $k$ (all rebalancing methods are jointly considered for each estimation method); then we compare performance across all $k$ as before (compare performance of our estimation method against the other estimation approaches but within each portfolio).

(c) For the third statistic (BETWEEN) we first pool the data for each $j$ from all $i$ (i.e., we take the meta-data from all the portfolios that have the same rebalancing method); and then compare performance across all $k$ (compare our estimation method against the other estimation approaches across portfolios).

3. Once the data are available as described above, the comparison of our methods against the benchmark is made via a GMM-approach[3] which estimates the mean differences (of one method against the benchmarks). Formally, for the case of the ALL-based statistics we first compute:

$$\Delta P(k_1, k_2) \stackrel{\text{def}}{=} \frac{1}{n_{all}} \sum_{i,j} \left( P_{ij,k_1} - P_{ij,k_2} \right), \tag{37}$$

where $n_{all}$ is the total number of observations across all portfolios $i$ and the rebalancing methods $j$ and $(k_1, k_2)$ represent the two competing estimation methods. Similarly, for the WITHIN-based statistics we have:

$$\Delta P_i(k_1, k_2) \stackrel{\text{def}}{=} \frac{1}{n_{i,j}} \sum_{j} \left( P_{ij,k_1} - P_{ij,k_2} \right), \tag{38}$$

[3]We perform a standard estimation and test for zero mean differences using a GMM-based approach and standard errors. Details of the calculations and many other statistics beyond those presented are available on request.

where $n_{i,j}$ is the number of rebalancing methods for portfolio $i$. Finally, for the BETWEEN-based statistics we have:

$$\Delta P_j(k_1, k_2) \stackrel{\text{def}}{=} \frac{1}{n_{j,i}} \sum_i \left(P_{ij,k_1} - P_{ij,k_2}\right), \tag{39}$$

where $n_{j,i}$ is the number of available portfolios for the rebalancing method $j$.

Once these statistics are available we present, as the final output, *the percentage of times that the difference is in favour of the new methods.* If $k_1$ is the benchmark estimation method then we compute, say for the ALL-based statistics, the following success ratio against the benchmark:

$$SR(k_1) \stackrel{\text{def}}{=} \frac{1}{N_k} \sum_k I\left[\Delta P(k_1, k) < 0\right], \tag{40}$$

if the performance measure is the average return, the Sharpe ratio or the cumulative return. If the performance measure is the volatility or the maximum drawdown then the " $< 0$" is substituted with a " $> 0$" within the indicator function. Here $N_k$ indicate the number of covariance averaging methods that we have available to compete against the benchmark.

## 4.3 Portfolio Data

We investigate three portfolios: (i) one with a relatively small universe of eight securities, (ii) another with a medium universe of twenty securities and (iii) a large portfolio which consists of forty securities. The tickers of these securities are given in Table 14 and are all included in the S&P500. The data consist of daily closing prices from April 2, 1990 to

October 26, 2012. It is important to notice that even though the data are collected daily, optimization, re-optimization and rebalancing are performed on a weekly/monthly/quarterly basis, as described above. All data are downloaded from the Yahoo! Finance website.

## 4.4 Empirical results

Our results are presented in Table 11, Table 12 and Table 13 with each table corresponding to the ALL-, WITHIN- and BETWEEN-based success ratios. We present the results for five evaluation statistics: average return and volatility (both annualized); Sharpe ratio; and also cumulative return and maximum drawdown[4]. Each table has three panels, labelled A, B and C, corresponding to the three groupings of stocks from the S&P500, as described above (except for Table 13, see below).

The question of relevance in our empirical application is this: does covariance averaging make a difference? If so, then our success ratios will be relatively high, indicating that (most of the time on average) it is better to consider using one of the covariance averaging methods than the competitive benchmarks. With our aggregation approach we do not differentiate across different covariance averaging approaches since this is less important; it becomes relevant only after we can successfully show that covariance averaging in general works better that the benchmarks.

Table 11 presents the results for the ALL-based statistics. A casual look at all three panels of the table shows a result of immediate practical interest: the larger the portfolio size the better covariance averaging works - across all evaluation measures. This is important, since

[4]The maximum drawdown is defined as the largest peak-to-trough drop in the portfolio value during the underlying evaluation period.

one of our arguments in favor of covariance averaging is precisely that in larger portfolios it is progressively more difficult to use sophisticated methods - thus the use of shrinkage. Although we have considered only one possible form of the shrinkage estimator, our results in Table 11 indicate that it is a safer bet to use covariance averaging than covariance shrinkage. Furthermore, even for the smaller group of only 8 stocks we can see that covariance averaging results in lower portfolio volatility and lower drawdown. Surprisingly enough, covariance averaging is preferable to the sample covariance only on grounds of lower volatility in this group. As we move to panels B and C, the results become progressively more favorable to covariance averaging, in terms of all evaluation measures; and in panel C where we have 40 stocks it becomes clear that covariance averaging is generally a better route to take. Overall, covariance averaging methods are better[5] than the benchmarks 55% of the time in Table 11; the average return, volatility and cumulative return are better 58.3% of the time while the Sharpe ratio and maximum drawdown are better 50% of the time. A potential disadvantage of the ALL-based success ratios is that we mix different rebalancing approaches, something that the BETWEEN-based statistics correct.

Table 12 has the results for the WITHIN-based statistics. Here the comparisons are made with each portfolio run across the different estimation methods and by the pooling of all rebalancing approaches together. The results are even more strongly in favor of covariance averaging than before, in the sense that we have high success ratios across the three groupings of 8, 20 and 40 stocks. If we do the same counting as before, we find that, across the three panels of the table, for 60% of the time covariance averaging is better than the benchmarks - and the results when we look at individual performance measure counts are similar. The potential disadvantage of the WITHIN-based statistics is that they use fewer observations.

[5] Counting the number of times that a table entry is greater than 50%.

However, when we turn to the results in Table 13 we have neither of the disadvantage of the previous two types of success ratio. Here we have a good number of observations that are actually homogeneous, since they pertain to the same type of rebalancing. So, not only can we judge performance better but we can also discuss the relevant merit of each rebalancing approach. The three panels of the table now refer only to the larger portfolio of the 40 stocks (results for other portfolios are available on request) and we consider three measures: the cumulative return, the maximum drawdown and the Sharpe ratio. Let us start with the first approach, which involves rebalancing at the optimization time only. In terms of a better cumulative return the covariance averaging methods are better at least 70% of the time, with an average gain ranging from 1.16% to over 7%. In terms of the risk measures, the methods that we propose are again better in terms of nominal performance, as we can see that the success ratios are over 50% for all but two cases. However, in terms of mean differences we see that - on average - they do not perform as well as in terms of the cumulative return. When we apply time rebalancing we still get similar characteristics, with some improvement in terms of the risk characteristics of the portfolios - but not enough to fully justify (on this dataset) the potential of time rebalancing. However, when we consider the *second threshold* for rebalancing or the *time and threshold rebalancing* we get the best overall performance of covariance averaging against the benchmarks. If one looks at the relevant lines of the respective panels, one can see that the proposed method has the largest average gain in terms of cumulative return but also the largest gain in terms of smaller drawdown.

# 5 Concluding Remarks

In this paper we propose a new method for estimating a (possibly time varying) covariance with a focus on the estimator of covariances of financial returns. The method expands upon two strands of the literature which have been unrelated up to now: one strand is that of covariance shrinkage (which is essentially model averaging) and the other strand is that of rolling window averaging (which is the time-series averaging of overlapping observations). The seminal works of Ledoit and Wolf (2003) and Ledoit and Wolf (2004) solved the problem of covariance averaging via shrinkage methods and show how performance improvements can be obtained by covariance averaging. The advantages of covariance shrinkage include its use on essentially any number of assets, thus covering the practical needs among asset managers in dealing with large and very large portfolios. However, in covariance shrinkage one does have to assume a particular covariance structure to average with the sample covariance and this puts the averaging in the category of model averaging. In this paper we connect the essential idea of covariance averaging from shrinkage with the second strand of the literature, which has been used successfully in forecasting applications.

We consider a covariance averaging estimator which is based on combining sample covariances to be estimated using different segments of the data, thus maintaining the idea of averaging across different covariances but without having to impose any structure at all; instead of having model averaging we have rolling window averaging. The proposed method has an intuitive and practical appeal in financial applications: it combines data from periods which are characterized by different volatility and correlation structures (and can thus account for time variation in the second moments of returns) and it is easy to use in any dimension, however large. We propose a variety of different schemes for obtaining the appropriate weights for performing the average across rolling windows, including heuristics-based and optimized weights, and furthermore provide an intuitive explanation concerning the interpretation of such weights.

To evaluate the potential of the proposed method we conduct a simulation experiment using two different data generating processes and most of our proposed estimators. The results of the simulations are benchmarked against the sample covariance and covariance shrinkage estimators and clearly support the use of covariance averaging using different rolling windows, even if one considers the simplest case where equal weights are given to different segments of observations. Furthermore, we find that the use of covariance averaging can lead to better forecasting performance than the benchmarks when the data generating process has a time-varying covariance.

Finally, we provide an empirical application in the context of a GMV portfolio optimization with rebalancing. We choose the GMV context since the main input is the covariance of the underlying returns. Our results across a broad range of underlying practical scenarios show that the new covariance averaging estimators can lead to improvements in both the return and risk of the underlying portfolios – with particular success in improving the risk characteristics and providing lower maximum drawdown on average than the benchmarks do. Our results are thus of immediate practical significance in the context of risk management, of larger portfolios in particular.

A number of future research items can be explored as follow-ups to the current paper, including: (a) further analysis of the theoretical properties of the resulting covariance averaging estimator; (b) a systematic comparison of fit and forecasting performance of the covariance averaging estimator vis--vis full-blown, GARCH-type parametric models in small dimensions (e.g. BEKK, DCC and similar models) and also a comparison with parametric models for larger dimensions; and (c) further assessment of the practical value of the proposed methods across portfolios of different compositions in terms of assets used and objective functions beyond the GMV one.

# References


BAJEUX-BESNAINOU, I., BANDARA, W., BURA, E., 2012. A Krylov subspace approach to large portfolio optimization. *Journal of Economic Dynamics and Control*, **36-11**, 1688-1699.

BAUWENS, L., LAURENT, S., ROMBOUTS, J. V. K., 2006. Multivariate GARCH models: a survey. *Journal of Applied Econometrics*, **21-1**, 79-109.

BHATTACHARYA, P., THOMAKOS, D. D., 2011. Improving forecasting performance by window and model averaging. *Quantf Research, Working Paper Series*, Forthcoming.

CHAN, L. K. C., KARCESKI, J., LAKONISHOK, J., 1999. On portfolio optimization: forecasting covariances and choosing the risk model. *The Review of Financial Studies*, **12-5**, 937-974.

CLARK, T. E., McCRACKEN, M. W., 2009. Improving forecast accuracy by combining recursive and rolling forecasts. *International Economic Review*, **50-2**, 363-395.

DeMIGUEL, V., GARLAPPI, L., NOGALES, F. J., UPPAL, R., 2009. A generalized approach to portfolio optimization: improving performance by constraining portfolio norms. *Management Science*, **55-5**, 798-812.

ENGLE, R., 2002. Dynamic conditional correlation: a simple class of multivariate generalized autoregressive conditional heteroskedasticity models. *Journal of Business & Economic Statistics*, **20-3**, 339-350.

FAN, J., FAN, Y., LV, J., 2008. High dimensional covariance matrix estimation using a factor model. *Journal of Econometrics*, **147**, 186-197.

HUO, L., TAE-HWAN, K., KIM, Y., 2012. Robust estimation of covariance and its application to portfolio optimization. *Finance Research Letters*, **9-3**, 121-134.

JACONETTI, C. M., KINNIRY, F. M. Jr., ZILBERING, Y., 2010. Best practices for portfolio rebalancing. *Vanguard Research*, Vanguard Group.

KHAN, R., ZHOU, G., 2007. Optimal portfolio choice with parameter uncertainty. *Journal of Financial and Quantitative Analysis*, **42**, 621-656.

KOURTIS, A., DOTSIS, G., MARKELLOS, R. N., 2012. Parameter uncertainty in portfolio selection: Shrinking the inverse covariance matrix. *Journal of Banking & Finance*, **36-9**, 2522-2531.

KWAN, C. C. Y., 2008. Estimation error in the average correlation of security returns and shrinkage estimation of covariance and correlation matrices. *Finance Research Letters*, **5**, 236-244.

LEDOIT, O., SANTA-CLARA, P., WOLF, M., 2003. Flexible multivariate GARCH modeling with an application to international stock markets. *Review of Economics and Statistics*, **85**, 735-747.

LEDOIT, O., WOLF, M., 2003. Improved Estimation for the covariance matrix of stock returns with an application to portfolio selection. *Journal of Empirical Finance*, **10-5**, 603-621.

LEDOIT, O., WOLF, M., 2004. Honey, I shrunk the sample covariance matrix. *Journal of Portfolio Management*, **31-1**, 110-119.

LEDOIT, O., WOLF, M., 2012. Nonlinear shrinkage estimation of large-dimensional covariance matrics. *Annals of Statistics*, **40-2**, 1024-1060.

MARKOWITZ, H., 1952. Portfolio selection. *Journal of Finance*, **7**, 77-91.

MARTELLINI, L., ZIEMANN, V., 2010. Improved estimates of higher-order comoments and implications for portfolio selection. *The Review of Financial Studies*, **23-4**, 1467-1502.

MOALLEMI, C. C., SAGLAM, M., 2013. Dynamic portfolio choice with linear rebalancing rules. *Working Paper*.

PAPAILIAS, F., KAPETANIOS, G., 2013. Cross-Validation Based Covariance Shrinkage in Portfolio Selection. *Quantf Research, Working Paper Series*, Forthcoming.

PATTON, A. J., SHEPPARD, K., 2009. Evaluating volatility and correlation forecasts. In: *Handbook of Financial Time Series Analysis*, 801-838, Springer.

PELLETIER, D., 2006. Regime switching for dynamic correlations. *Journal of Econometrics*, **131-1,2**, 445-473.

PESARAN, M. H., SCHUERMANN, T., SMITH, L. V., 2009. Forecasting economic and financial variables with global VARs. *International Journal of Forecasting*, **25-4**, 642-675.

ROSSI, B., INOUE, A., 2012. Out-of-sample forecast tests robust to the choice of window size. *Journal of Business & Economic Statistics*, **30-3**, 432-453.

SHARPE, W. F., 1963. A simplified model for portfolio analysis. *Management Science*, **9-1**, 227-293.

SILVENNOINEN, A,. TERASVIRTA, T., 2009. Modeling multivariate autoregressive conditional heteroskedasticity with the double smooth transition conditional correlation GARCH model. *Journal of Financial Econometrics*, **7-4**, 373-411.

SMITH BARNEY CONSULTING GROUP, CITIGROUP, 2005. The art of rebalancing. *Executive Report*, Citigroup Global Markets Inc.

SUN, W., FAN, A., CHEN, L.-W., SCHOUWENAARS, T., ALBOTA, M. A., 2006. Using dynamic programming to optimally rebalance portfolios. *The Journal of Trading*, **1-2**, 16-27.

TUNC, S., KOZAT, S. S., 2012. Optimal investment under transaction costs: a threshold rebalanced portfolio approach. In: *Proceedings of CoRR*.

WANG, Z., 2005. A shrinkage approach to model uncertainty and asset allocation. *The Review of Financial Studies*, **18-2**, 673-705.

| Covariance Averaging using Bivariate Scalar Diagonal VECH with $\rho = -0.9$ | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | $B$=(5, 20, 50) | | | | $B$=(5, 20, 50, 100) | | | | $B$=(50, 100, 200, 400) | | | | $B$=(5, 20, 50, 100, 200, 400) | | | |
| Estimator | $t$ | $t+1$ | $t+5$ | $t+10$ | $t$ | $t+1$ | $t+5$ | $t+10$ | $t$ | $t+1$ | $t+5$ | $t+10$ | $t$ | $t+1$ | $t+5$ | $t+10$ |
| Realized Covariance | 0.56 | 0.62 | 0.76 | 0.84 | 0.50 | 0.55 | 0.78 | 0.94 | 0.46 | 0.53 | 0.74 | 0.96 | 0.46 | 0.53 | 0.79 | 0.99 |
| Sample Covariance (H) | 0.85 | 0.83 | 0.94 | 1.01 | 0.88 | 0.88 | 0.94 | 1.01 | 0.95 | 0.96 | 0.98 | 1.01 | 0.97 | 0.97 | 0.99 | 1.00 |
| LW Shrinkage (F) | 0.99 | 0.99 | 0.99 | 0.99 | 1.00 | 1.00 | 0.99 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |
| LW Shrinkage (H) | 0.83 | 0.83 | 0.93 | 0.99 | 0.87 | 0.88 | 0.93 | 1.00 | 0.95 | 0.96 | 0.97 | 1.00 | 0.97 | 0.97 | 0.99 | 1.00 |
| Equal Weights | 0.68 | 0.57 | 0.80 | 0.93 | 0.49 | 0.45 | 0.74 | 0.89 | 0.79 | 0.81 | 0.86 | 0.92 | 0.50 | 0.47 | 0.67 | 0.83 |
| EMA Weights 1 | 0.92 | 0.79 | 0.92 | 1.06 | 0.74 | 0.67 | 0.93 | 1.11 | 0.75 | 0.78 | 0.85 | 0.94 | 0.69 | 0.61 | 0.89 | 1.16 |
| EMA Weights 2 | 1.02 | 0.89 | 0.99 | 1.12 | 0.86 | 0.79 | 1.03 | 1.20 | 0.76 | 0.78 | 0.86 | 0.96 | 0.85 | 0.77 | 1.03 | 1.30 |
| Optimized Weights 1 | 0.80 | 0.77 | 0.84 | 0.88 | 0.73 | 0.70 | 0.79 | 0.88 | 0.83 | 0.83 | 0.85 | 0.87 | 0.74 | 0.71 | 0.79 | 0.85 |
| Optimized Weights 2 | 1.01 | 0.96 | 0.98 | 1.00 | 0.91 | 0.86 | 0.94 | 1.05 | 0.81 | 0.82 | 0.84 | 0.87 | 0.96 | 0.91 | 0.99 | 1.07 |
| Optimized EMA Weights 1 | 0.78 | 0.66 | 0.84 | 0.96 | 0.55 | 0.50 | 0.77 | 0.92 | 0.77 | 0.78 | 0.84 | 0.90 | 0.51 | 0.46 | 0.65 | 0.83 |
| Optimized EMA Weights 2 | 1.00 | 0.87 | 0.98 | 1.08 | 0.76 | 0.69 | 0.93 | 1.09 | 0.75 | 0.77 | 0.82 | 0.90 | 0.76 | 0.69 | 0.88 | 1.07 |
| Optimized $a^*$ | 0.84 | 0.81 | 0.87 | 0.89 | 0.79 | 0.75 | 0.83 | 0.90 | 0.88 | 0.88 | 0.89 | 0.90 | 0.79 | 0.76 | 0.82 | 0.88 |
| Distance Weights 1 | 0.99 | 0.91 | 0.96 | 1.06 | 0.84 | 0.78 | 0.94 | 1.09 | 0.78 | 0.78 | 0.87 | 1.02 | 0.74 | 0.75 | 0.81 | 0.92 |
| Distance Weights 2 | 0.71 | 0.64 | 0.79 | 0.90 | 0.51 | 0.47 | 0.71 | 0.86 | 0.77 | 0.79 | 0.84 | 0.89 | 0.51 | 0.48 | 0.66 | 0.81 |
| Distance Weights 3 | 0.67 | 0.57 | 0.78 | 0.91 | 0.48 | 0.44 | 0.72 | 0.88 | 0.79 | 0.80 | 0.85 | 0.91 | 0.50 | 0.47 | 0.67 | 0.83 |

Table 1: Simulation Design 1. Reporting the Average Mean Distance Related to the Unconditional Full Sample Covariance Estimate



**Notes:**

**Realized Covariance** denotes the realized covariance estimator.

**Sample Covariance (H)** denotes the sample covariance estimator using the last half of the sample.

**LW Shrinkage (F)** denotes the Ledoit and Wolf (2003) covariance shrinkage estimator using the full sample.

**LW Shrinkage (H)** denotes the Ledoit and Wolf (2003) covariance shrinkage estimator using the last half of the sample.

**Equal Weights** denotes the covariance averaging estimator using equal weights as in Equation (11).

**EMA Weights 1** denotes the covariance averaging estimator using data dependent weights as in Equation (12) and Equation (13) ($\widehat{\alpha}_1$).

**EMA Weights 2** denotes the covariance averaging estimator using data dependent weights as in Equation (12) and Equation (13) ($\widehat{\alpha}_1$).

**Optimized Weights 1** denotes the covariance averaging estimator using the standard objective function as in Equation (28) with bias correction.

**Optimized Weights 2** denotes the covariance averaging estimator using the standard objective function as in Equation (28) without bias correction.

**Optimized EMA Weights 1** denotes the covariance averaging estimator using the standard objective function as in Equation (28), weights that are parameterized via the exponential scheme of Equation (12) and with bias correction.

**Optimized EMA Weights 1** denotes the covariance averaging estimator using the standard objective function as in Equation (28), weights that are parameterized via the exponential scheme of Equation (12) and without bias correction.

**Optimized** $a^*$ denotes the covariance averaging estimator using the unrestricted weights of Equation (31) with bias correction.

**Distance Weights 1** denotes the covariance averaging estimator using the distance-based weights from Equation (14), Equation (15) and Equation (16) - first case.

**Distance Weights 2** denotes the covariance averaging estimator using the distance-based weights from Equation (14), Equation (15) and Equation (16) - second case.

**Distance Weights 3** denotes the covariance averaging estimator using the distance-based weights from Equation (14), Equation (15) and Equation (16) - third case.

| Covariance Averaging using Bivariate Scalar Diagonal VECH with $\rho = -0.5$ | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | $B$=(5, 20, 50) | | | | $B$=(5, 20, 50, 100) | | | | $B$=(50, 100, 200, 400) | | | | $B$=(5, 20, 50, 100, 200, 400) | | | |
| Estimator | $t$ | $t+1$ | $t+5$ | $t+10$ | $t$ | $t+1$ | $t+5$ | $t+10$ | $t$ | $t+1$ | $t+5$ | $t+10$ | $t$ | $t+1$ | $t+5$ | $t+10$ |
| Realized Covariance | 0.50 | 0.58 | 0.78 | 0.88 | 0.42 | 0.51 | 0.72 | 0.84 | 0.43 | 0.52 | 0.78 | 0.98 | 0.42 | 0.52 | 0.77 | 0.98 |
| Sample Covariance (H) | 0.79 | 0.77 | 0.94 | 1.02 | 0.89 | 0.90 | 0.95 | 1.01 | 1.00 | 1.00 | 1.02 | 1.04 | 1.00 | 1.00 | 1.03 | 1.05 |
| LW Shrinkage (F) | 0.99 | 0.99 | 0.99 | 0.99 | 1.00 | 1.00 | 0.99 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |
| LW Shrinkage (H) | 0.78 | 0.76 | 0.91 | 0.99 | 0.89 | 0.89 | 0.95 | 1.00 | 1.00 | 1.00 | 1.01 | 1.04 | 1.00 | 1.00 | 1.03 | 1.05 |
| Equal Weights | 0.64 | 0.56 | 0.83 | 0.96 | 0.52 | 0.48 | 0.72 | 0.86 | 0.84 | 0.84 | 0.91 | 0.97 | 0.52 | 0.49 | 0.72 | 0.89 |
| EMA Weights 1 | 0.91 | 0.78 | 1.00 | 1.11 | 0.71 | 0.62 | 0.85 | 0.99 | 0.81 | 0.82 | 0.92 | 1.01 | 0.66 | 0.57 | 0.90 | 1.16 |
| EMA Weights 2 | 1.02 | 0.89 | 1.08 | 1.19 | 0.83 | 0.74 | 0.94 | 1.08 | 0.81 | 0.82 | 0.94 | 1.03 | 0.82 | 0.72 | 1.01 | 1.28 |
| Optimized Weights 1 | 0.77 | 0.74 | 0.83 | 0.89 | 0.71 | 0.67 | 0.78 | 0.84 | 0.86 | 0.87 | 0.89 | 0.91 | 0.68 | 0.66 | 0.76 | 0.84 |
| Optimized Weights 2 | 0.90 | 0.84 | 0.92 | 0.97 | 0.80 | 0.75 | 0.85 | 0.91 | 0.86 | 0.86 | 0.89 | 0.91 | 0.77 | 0.75 | 0.86 | 0.96 |
| Optimized EMA Weights 1 | 0.73 | 0.63 | 0.87 | 0.99 | 0.57 | 0.50 | 0.73 | 0.86 | 0.82 | 0.83 | 0.89 | 0.95 | 0.51 | 0.47 | 0.70 | 0.88 |
| Optimized EMA Weights 2 | 0.89 | 0.77 | 0.97 | 1.07 | 0.71 | 0.62 | 0.83 | 0.96 | 0.81 | 0.82 | 0.89 | 0.95 | 0.66 | 0.61 | 0.84 | 1.02 |
| Optimized $a^*$ | 0.81 | 0.78 | 0.86 | 0.90 | 0.74 | 0.71 | 0.80 | 0.85 | 0.90 | 0.91 | 0.92 | 0.92 | 0.71 | 0.70 | 0.79 | 0.85 |
| Distance Weights 1 | 0.88 | 0.77 | 0.93 | 1.00 | 0.71 | 0.66 | 0.78 | 0.95 | 0.69 | 0.66 | 0.82 | 1.00 | 0.82 | 0.82 | 0.85 | 1.01 |
| Distance Weights 2 | 0.67 | 0.58 | 0.81 | 0.93 | 0.51 | 0.47 | 0.69 | 0.82 | 0.82 | 0.83 | 0.89 | 0.94 | 0.51 | 0.48 | 0.70 | 0.86 |
| Distance Weights 3 | 0.63 | 0.54 | 0.81 | 0.94 | 0.51 | 0.46 | 0.70 | 0.84 | 0.83 | 0.84 | 0.90 | 0.96 | 0.52 | 0.49 | 0.72 | 0.88 |

Table 2: Simulation Design 2. Reporting the Average Mean Distance Related to the Unconditional Full Sample Covariance Estimate



**Notes:**

**Realized Covariance** denotes the realized covariance estimator.

**Sample Covariance (H)** denotes the sample covariance estimator using the last half of the sample.

**LW Shrinkage (F)** denotes the Ledoit and Wolf (2003) covariance shrinkage estimator using the full sample.

**LW Shrinkage (H)** denotes the Ledoit and Wolf (2003) covariance shrinkage estimator using the last half of the sample.

**Equal Weights** denotes the covariance averaging estimator using equal weights as in Equation (11).

**EMA Weights 1** denotes the covariance averaging estimator using data dependent weights as in Equation (12) and Equation (13) ($\widehat{\alpha}_1$).

**EMA Weights 2** denotes the covariance averaging estimator using data dependent weights as in Equation (12) and Equation (13) ($\widehat{\alpha}_1$).

**Optimized Weights 1** denotes the covariance averaging estimator using the standard objective function as in Equation (28) with bias correction.

**Optimized Weights 2** denotes the covariance averaging estimator using the standard objective function as in Equation (28) without bias correction.

**Optimized EMA Weights 1** denotes the covariance averaging estimator using the standard objective function as in Equation (28), weights that are parameterized via the exponential scheme of Equation (12) and with bias correction.

**Optimized EMA Weights 1** denotes the covariance averaging estimator using the standard objective function as in Equation (28), weights that are parameterized via the exponential scheme of Equation (12) and without bias correction.

**Optimized** $a^*$ denotes the covariance averaging estimator using the unrestricted weights of Equation (31) with bias correction.

**Distance Weights 1** denotes the covariance averaging estimator using the distance-based weights from Equation (14), Equation (15) and Equation (16) - first case.

**Distance Weights 2** denotes the covariance averaging estimator using the distance-based weights from Equation (14), Equation (15) and Equation (16) - second case.

**Distance Weights 3** denotes the covariance averaging estimator using the distance-based weights from Equation (14), Equation (15) and Equation (16) - third case.

| Covariance Averaging using Bivariate Scalar Diagonal VECH with $\rho = 0$ | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | $B$=(5, 20, 50) | | | | $B$=(5, 20, 50, 100) | | | | $B$=(50, 100, 200, 400) | | | | $B$=(5, 20, 50, 100, 200, 400) | | | |
| Estimator | $t$ | $t+1$ | $t+5$ | $t+10$ | $t$ | $t+1$ | $t+5$ | $t+10$ | $t$ | $t+1$ | $t+5$ | $t+10$ | $t$ | $t+1$ | $t+5$ | $t+10$ |
| Realized Covariance | 0.47 | 0.57 | 0.77 | 0.89 | 0.40 | 0.50 | 0.75 | 0.87 | 0.39 | 0.47 | 0.77 | 0.94 | 0.40 | 0.51 | 0.77 | 0.96 |
| Sample Covariance (H) | 0.81 | 0.82 | 0.96 | 1.04 | 0.90 | 0.92 | 0.99 | 1.04 | 0.98 | 0.99 | 1.02 | 1.03 | 1.00 | 0.99 | 1.02 | 1.03 |
| LW Shrinkage (F) | 0.99 | 0.99 | 0.99 | 0.99 | 1.00 | 1.00 | 1.00 | 0.99 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |
| LW Shrinkage (H) | 0.79 | 0.80 | 0.94 | 1.02 | 0.90 | 0.91 | 0.98 | 1.03 | 0.98 | 0.98 | 1.02 | 1.03 | 0.99 | 0.99 | 1.02 | 1.02 |
| Equal Weights | 0.64 | 0.57 | 0.83 | 0.96 | 0.52 | 0.47 | 0.75 | 0.89 | 0.83 | 0.83 | 0.91 | 0.96 | 0.52 | 0.49 | 0.70 | 0.85 |
| EMA Weights 1 | 0.91 | 0.81 | 0.99 | 1.11 | 0.74 | 0.66 | 0.94 | 1.08 | 0.79 | 0.81 | 0.93 | 1.00 | 0.65 | 0.58 | 0.88 | 1.11 |
| EMA Weights 2 | 1.04 | 0.92 | 1.08 | 1.18 | 0.88 | 0.79 | 1.04 | 1.17 | 0.80 | 0.81 | 0.94 | 1.02 | 0.81 | 0.73 | 1.00 | 1.23 |
| Optimized Weights 1 | 0.75 | 0.70 | 0.81 | 0.89 | 0.70 | 0.68 | 0.78 | 0.86 | 0.86 | 0.86 | 0.89 | 0.92 | 0.70 | 0.67 | 0.77 | 0.84 |
| Optimized Weights 2 | 0.83 | 0.76 | 0.87 | 0.94 | 0.74 | 0.70 | 0.83 | 0.91 | 0.85 | 0.85 | 0.89 | 0.92 | 0.74 | 0.70 | 0.81 | 0.90 |
| Optimized EMA Weights 1 | 0.73 | 0.64 | 0.87 | 0.98 | 0.57 | 0.50 | 0.77 | 0.91 | 0.82 | 0.82 | 0.90 | 0.95 | 0.52 | 0.49 | 0.70 | 0.86 |
| Optimized EMA Weights 2 | 0.85 | 0.75 | 0.94 | 1.04 | 0.66 | 0.58 | 0.85 | 0.98 | 0.81 | 0.81 | 0.89 | 0.95 | 0.63 | 0.57 | 0.78 | 0.95 |
| Optimized $a^*$ | 0.78 | 0.73 | 0.83 | 0.89 | 0.74 | 0.71 | 0.80 | 0.87 | 0.89 | 0.89 | 0.91 | 0.93 | 0.72 | 0.70 | 0.79 | 0.84 |
| Distance Weights 1 | 0.84 | 0.75 | 0.90 | 0.95 | 0.70 | 0.63 | 0.81 | 0.92 | 0.66 | 0.61 | 0.76 | 0.86 | 0.84 | 0.83 | 0.86 | 0.95 |
| Distance Weights 2 | 0.65 | 0.56 | 0.80 | 0.92 | 0.51 | 0.46 | 0.72 | 0.86 | 0.81 | 0.82 | 0.89 | 0.94 | 0.51 | 0.48 | 0.68 | 0.83 |
| Distance Weights 3 | 0.62 | 0.55 | 0.81 | 0.94 | 0.51 | 0.46 | 0.74 | 0.88 | 0.82 | 0.83 | 0.90 | 0.95 | 0.51 | 0.49 | 0.69 | 0.85 |

Table 3: Simulation Design 3. Reporting the Average Mean Distance Related to the Unconditional Full Sample Covariance Estimate

**Notes:**

**Realized Covariance** denotes the realized covariance estimator.

**Sample Covariance (H)** denotes the sample covariance estimator using the last half of the sample.

**LW Shrinkage (F)** denotes the Ledoit and Wolf (2003) covariance shrinkage estimator using the full sample.

**LW Shrinkage (H)** denotes the Ledoit and Wolf (2003) covariance shrinkage estimator using the last half of the sample.

**Equal Weights** denotes the covariance averaging estimator using equal weights as in Equation (11).

**EMA Weights 1** denotes the covariance averaging estimator using data dependent weights as in Equation (12) and Equation (13) ($\widehat{\alpha}_1$).

**EMA Weights 2** denotes the covariance averaging estimator using data dependent weights as in Equation (12) and Equation (13) ($\widehat{\alpha}_1$).

**Optimized Weights 1** denotes the covariance averaging estimator using the standard objective function as in Equation (28) with bias correction.

**Optimized Weights 2** denotes the covariance averaging estimator using the standard objective function as in Equation (28) without bias correction.

**Optimized EMA Weights 1** denotes the covariance averaging estimator using the standard objective function as in Equation (28), weights that are parameterized via the exponential scheme of Equation (12) and with bias correction.

**Optimized EMA Weights 1** denotes the covariance averaging estimator using the standard objective function as in Equation (28), weights that are parameterized via the exponential scheme of Equation (12) and without bias correction.

**Optimized** $a^*$ denotes the covariance averaging estimator using the unrestricted weights of Equation (31) with bias correction.

**Distance Weights 1** denotes the covariance averaging estimator using the distance-based weights from Equation (14), Equation (15) and Equation (16) - first case.

**Distance Weights 2** denotes the covariance averaging estimator using the distance-based weights from Equation (14), Equation (15) and Equation (16) - second case.

**Distance Weights 3** denotes the covariance averaging estimator using the distance-based weights from Equation (14), Equation (15) and Equation (16) - third case.

| Covariance Averaging using Bivariate Scalar Diagonal VECH with $\rho = 0.5$ | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | $B$=(5, 20, 50) | | | | $B$=(5, 20, 50, 100) | | | | $B$=(50, 100, 200, 400) | | | | $B$=(5, 20, 50, 100, 200, 400) | | | |
| Estimator | $t$ | $t+1$ | $t+5$ | $t+10$ | $t$ | $t+1$ | $t+5$ | $t+10$ | $t$ | $t+1$ | $t+5$ | $t+10$ | $t$ | $t+1$ | $t+5$ | $t+10$ |
| Realized Covariance | 0.48 | 0.57 | 0.75 | 0.85 | 0.43 | 0.52 | 0.73 | 0.86 | 0.40 | 0.49 | 0.75 | 0.93 | 0.41 | 0.51 | 0.77 | 0.96 |
| Sample Covariance (H) | 0.78 | 0.79 | 0.91 | 1.00 | 0.88 | 0.87 | 0.96 | 1.01 | 1.02 | 1.02 | 1.04 | 1.05 | 0.98 | 0.98 | 1.00 | 1.02 |
| LW Shrinkage (F) | 0.99 | 0.99 | 0.99 | 0.99 | 1.00 | 1.00 | 1.00 | 0.99 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |
| LW Shrinkage (H) | 0.77 | 0.78 | 0.89 | 0.98 | 0.87 | 0.87 | 0.95 | 1.00 | 1.02 | 1.02 | 1.04 | 1.05 | 0.98 | 0.98 | 1.00 | 1.02 |
| Equal Weights | 0.65 | 0.59 | 0.81 | 0.94 | 0.52 | 0.46 | 0.72 | 0.87 | 0.85 | 0.86 | 0.90 | 0.94 | 0.51 | 0.47 | 0.69 | 0.85 |
| EMA Weights 1 | 0.93 | 0.83 | 0.97 | 1.08 | 0.72 | 0.61 | 0.87 | 1.02 | 0.79 | 0.81 | 0.89 | 0.96 | 0.66 | 0.56 | 0.87 | 1.09 |
| EMA Weights 2 | 1.04 | 0.94 | 1.05 | 1.15 | 0.85 | 0.73 | 0.97 | 1.11 | 0.79 | 0.81 | 0.90 | 0.98 | 0.82 | 0.71 | 0.99 | 1.21 |
| Optimized Weights 1 | 0.77 | 0.73 | 0.83 | 0.90 | 0.72 | 0.69 | 0.80 | 0.85 | 0.86 | 0.86 | 0.90 | 0.91 | 0.71 | 0.69 | 0.78 | 0.84 |
| Optimized Weights 2 | 0.89 | 0.83 | 0.91 | 0.96 | 0.82 | 0.77 | 0.88 | 0.94 | 0.85 | 0.85 | 0.89 | 0.91 | 0.77 | 0.74 | 0.86 | 0.92 |
| Optimized EMA Weights 1 | 0.74 | 0.67 | 0.85 | 0.97 | 0.59 | 0.51 | 0.76 | 0.88 | 0.83 | 0.83 | 0.89 | 0.93 | 0.51 | 0.46 | 0.68 | 0.85 |
| Optimized EMA Weights 2 | 0.88 | 0.80 | 0.94 | 1.04 | 0.73 | 0.64 | 0.87 | 0.98 | 0.81 | 0.82 | 0.88 | 0.92 | 0.64 | 0.57 | 0.80 | 0.97 |
| Optimized $a^*$ | 0.82 | 0.78 | 0.86 | 0.91 | 0.75 | 0.73 | 0.83 | 0.87 | 0.89 | 0.89 | 0.92 | 0.93 | 0.74 | 0.72 | 0.80 | 0.85 |
| Distance Weights 1 | 0.88 | 0.80 | 0.91 | 1.01 | 0.74 | 0.68 | 0.85 | 0.94 | 0.70 | 0.66 | 0.79 | 0.95 | 0.73 | 0.73 | 0.81 | 0.84 |
| Distance Weights 2 | 0.66 | 0.59 | 0.79 | 0.92 | 0.53 | 0.47 | 0.71 | 0.84 | 0.83 | 0.84 | 0.89 | 0.92 | 0.50 | 0.47 | 0.67 | 0.82 |
| Distance Weights 3 | 0.63 | 0.57 | 0.79 | 0.92 | 0.52 | 0.45 | 0.72 | 0.86 | 0.84 | 0.85 | 0.90 | 0.94 | 0.50 | 0.47 | 0.68 | 0.84 |

Table 4: Simulation Design 4. Reporting the Average Mean Distance Related to the Unconditional Full Sample Covariance Estimate

**Notes:**

**Realized Covariance** denotes the realized covariance estimator.

**Sample Covariance (H)** denotes the sample covariance estimator using the last half of the sample.

**LW Shrinkage (F)** denotes the Ledoit and Wolf (2003) covariance shrinkage estimator using the full sample.

**LW Shrinkage (H)** denotes the Ledoit and Wolf (2003) covariance shrinkage estimator using the last half of the sample.

**Equal Weights** denotes the covariance averaging estimator using equal weights as in Equation (11).

**EMA Weights 1** denotes the covariance averaging estimator using data dependent weights as in Equation (12) and Equation (13) ($\widehat{\alpha}_1$).

**EMA Weights 2** denotes the covariance averaging estimator using data dependent weights as in Equation (12) and Equation (13) ($\widehat{\alpha}_1$).

**Optimized Weights 1** denotes the covariance averaging estimator using the standard objective function as in Equation (28) with bias correction.

**Optimized Weights 2** denotes the covariance averaging estimator using the standard objective function as in Equation (28) without bias correction.

**Optimized EMA Weights 1** denotes the covariance averaging estimator using the standard objective function as in Equation (28), weights that are parameterized via the exponential scheme of Equation (12) and with bias correction.

**Optimized EMA Weights 1** denotes the covariance averaging estimator using the standard objective function as in Equation (28), weights that are parameterized via the exponential scheme of Equation (12) and without bias correction.

**Optimized** $a^*$ denotes the covariance averaging estimator using the unrestricted weights of Equation (31) with bias correction.

**Distance Weights 1** denotes the covariance averaging estimator using the distance-based weights from Equation (14), Equation (15) and Equation (16) - first case.

**Distance Weights 2** denotes the covariance averaging estimator using the distance-based weights from Equation (14), Equation (15) and Equation (16) - second case.

**Distance Weights 3** denotes the covariance averaging estimator using the distance-based weights from Equation (14), Equation (15) and Equation (16) - third case.

| Covariance Averaging using Bivariate Scalar Diagonal VECH with $\rho = 0.9$ | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | $B$=(5, 20, 50) | | | | $B$=(5, 20, 50, 100) | | | | $B$=(50, 100, 200, 400) | | | | $B$=(5, 20, 50, 100, 200, 400) | | | |
| Estimator | $t$ | $t+1$ | $t+5$ | $t+10$ | $t$ | $t+1$ | $t+5$ | $t+10$ | $t$ | $t+1$ | $t+5$ | $t+10$ | $t$ | $t+1$ | $t+5$ | $t+10$ |
| Realized Covariance | 0.56 | 0.63 | 0.79 | 0.89 | 0.53 | 0.62 | 0.85 | 0.98 | 0.48 | 0.56 | 0.79 | 0.98 | 0.50 | 0.58 | 0.79 | 0.94 |
| Sample Covariance (H) | 0.80 | 0.81 | 0.94 | 1.01 | 0.89 | 0.91 | 0.99 | 1.05 | 1.01 | 1.01 | 1.02 | 1.03 | 1.02 | 1.02 | 1.03 | 1.03 |
| LW Shrinkage (F) | 0.99 | 0.99 | 0.99 | 0.99 | 1.00 | 1.00 | 1.00 | 0.99 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |
| LW Shrinkage (H) | 0.79 | 0.79 | 0.92 | 0.99 | 0.89 | 0.90 | 0.98 | 1.03 | 1.01 | 1.01 | 1.02 | 1.03 | 1.01 | 1.01 | 1.02 | 1.03 |
| Equal Weights | 0.62 | 0.56 | 0.82 | 0.95 | 0.54 | 0.48 | 0.75 | 0.92 | 0.86 | 0.86 | 0.91 | 0.96 | 0.53 | 0.50 | 0.70 | 0.83 |
| EMA Weights 1 | 0.89 | 0.79 | 0.98 | 1.09 | 0.78 | 0.66 | 0.94 | 1.10 | 0.82 | 0.83 | 0.90 | 0.99 | 0.70 | 0.59 | 0.90 | 1.07 |
| EMA Weights 2 | 0.98 | 0.88 | 1.05 | 1.15 | 0.91 | 0.79 | 1.04 | 1.19 | 0.82 | 0.83 | 0.91 | 1.01 | 0.87 | 0.76 | 1.04 | 1.20 |
| Optimized Weights 1 | 0.80 | 0.76 | 0.86 | 0.92 | 0.75 | 0.74 | 0.83 | 0.89 | 0.87 | 0.87 | 0.90 | 0.91 | 0.72 | 0.70 | 0.77 | 0.82 |
| Optimized Weights 2 | 0.99 | 0.94 | 1.00 | 1.04 | 0.95 | 0.92 | 1.00 | 1.05 | 0.86 | 0.87 | 0.90 | 0.91 | 0.94 | 0.91 | 0.99 | 1.04 |
| Optimized EMA Weights 1 | 0.72 | 0.65 | 0.87 | 0.99 | 0.61 | 0.54 | 0.79 | 0.94 | 0.84 | 0.84 | 0.89 | 0.94 | 0.54 | 0.50 | 0.69 | 0.83 |
| Optimized EMA Weights 2 | 0.93 | 0.84 | 1.01 | 1.11 | 0.84 | 0.74 | 0.98 | 1.11 | 0.83 | 0.83 | 0.89 | 0.94 | 0.80 | 0.73 | 0.94 | 1.06 |
| Optimized $a^*$ | 0.85 | 0.82 | 0.89 | 0.93 | 0.81 | 0.80 | 0.86 | 0.90 | 0.91 | 0.91 | 0.93 | 0.94 | 0.76 | 0.75 | 0.80 | 0.84 |
| Distance Weights 1 | 0.98 | 0.91 | 1.02 | 1.06 | 0.95 | 0.89 | 0.99 | 1.04 | 0.76 | 0.73 | 0.80 | 0.94 | 0.85 | 0.87 | 0.90 | 0.98 |
| Distance Weights 2 | 0.69 | 0.63 | 0.81 | 0.92 | 0.55 | 0.50 | 0.74 | 0.89 | 0.84 | 0.85 | 0.89 | 0.94 | 0.53 | 0.49 | 0.68 | 0.81 |
| Distance Weights 3 | 0.62 | 0.55 | 0.80 | 0.93 | 0.53 | 0.47 | 0.74 | 0.91 | 0.85 | 0.86 | 0.90 | 0.95 | 0.53 | 0.50 | 0.69 | 0.83 |

Table 5: Simulation Design 5. Reporting the Average Mean Distance Related to the Unconditional Full Sample Covariance Estimate



**Notes:**

**Realized Covariance** denotes the realized covariance estimator.

**Sample Covariance (H)** denotes the sample covariance estimator using the last half of the sample.

**LW Shrinkage (F)** denotes the Ledoit and Wolf (2003) covariance shrinkage estimator using the full sample.

**LW Shrinkage (H)** denotes the Ledoit and Wolf (2003) covariance shrinkage estimator using the last half of the sample.

**Equal Weights** denotes the covariance averaging estimator using equal weights as in Equation (11).

**EMA Weights 1** denotes the covariance averaging estimator using data dependent weights as in Equation (12) and Equation (13) ($\widehat{\alpha}_1$).

**EMA Weights 2** denotes the covariance averaging estimator using data dependent weights as in Equation (12) and Equation (13) ($\widehat{\alpha}_1$).

**Optimized Weights 1** denotes the covariance averaging estimator using the standard objective function as in Equation (28) with bias correction.

**Optimized Weights 2** denotes the covariance averaging estimator using the standard objective function as in Equation (28) without bias correction.

**Optimized EMA Weights 1** denotes the covariance averaging estimator using the standard objective function as in Equation (28), weights that are parameterized via the exponential scheme of Equation (12) and with bias correction.

**Optimized EMA Weights 1** denotes the covariance averaging estimator using the standard objective function as in Equation (28), weights that are parameterized via the exponential scheme of Equation (12) and without bias correction.

**Optimized** $a^*$ denotes the covariance averaging estimator using the unrestricted weights of Equation (31) with bias correction.

**Distance Weights 1** denotes the covariance averaging estimator using the distance-based weights from Equation (14), Equation (15) and Equation (16) - first case.

**Distance Weights 2** denotes the covariance averaging estimator using the distance-based weights from Equation (14), Equation (15) and Equation (16) - second case.

**Distance Weights 3** denotes the covariance averaging estimator using the distance-based weights from Equation (14), Equation (15) and Equation (16) - third case.

| Covariance Averaging using Exponential Time Varying True Covariance with $N=5$ | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | $B$=(5, 20, 50) | | | | $B$=(5, 20, 50, 100) | | | | $B$=(50, 100, 200, 400) | | | | $B$=(5, 20, 50, 100, 200, 400) | | | |
| Estimator | $t$ | $t+1$ | $t+5$ | $t+10$ | $t$ | $t+1$ | $t+5$ | $t+10$ | $t$ | $t+1$ | $t+5$ | $t+10$ | $t$ | $t+1$ | $t+5$ | $t+10$ |
| Sample Covariance (H) | 0.98 | 1.02 | 1.09 | 1.07 | 0.93 | 0.95 | 0.96 | 0.98 | 1.09 | 1.10 | 1.18 | 1.17 | 1.04 | 1.01 | 1.04 | 1.06 |
| LW Shrinkage (F) | 0.95 | 0.96 | 0.96 | 0.94 | 0.95 | 0.95 | 0.96 | 0.94 | 0.96 | 0.96 | 0.95 | 0.95 | 0.98 | 0.99 | 0.96 | 0.96 |
| LW Shrinkage (H) | 0.93 | 0.96 | 1.00 | 1.02 | 0.88 | 0.90 | 0.91 | 0.92 | 1.06 | 1.06 | 1.10 | 1.10 | 1.00 | 1.00 | 1.00 | 1.02 |
| Equal Weights | 0.71 | 0.96 | 1.00 | 1.02 | 0.88 | 0.90 | 0.91 | 0.92 | 1.06 | 1.06 | 1.10 | 1.10 | 1.00 | 1.00 | 1.00 | 1.02 |
| EMA Weights 1 | 0.71 | 0.82 | 0.96 | 1.02 | 0.78 | 0.83 | 0.94 | 1.03 | 1.06 | 1.06 | 1.10 | 1.12 | 0.83 | 0.90 | 1.02 | 1.15 |
| EMA Weights 2 | 0.56 | 0.82 | 0.96 | 1.02 | 0.78 | 0.83 | 0.94 | 1.03 | 1.06 | 1.06 | 1.10 | 1.12 | 0.83 | 0.90 | 1.02 | 1.15 |
| Optimized Weights 1 | 0.71 | 0.73 | 0.98 | 1.07 | 0.55 | 0.64 | 0.99 | 1.08 | 1.04 | 1.04 | 1.13 | 1.12 | 0.64 | 0.78 | 1.26 | 1.48 |
| Optimized Weights 2 | 0.63 | 0.78 | 0.91 | 0.96 | 0.72 | 0.76 | 0.88 | 0.86 | 0.85 | 0.84 | 0.83 | 0.83 | 0.81 | 0.88 | 0.89 | 0.92 |
| Optimized EMA Weights 1 | 0.59 | 0.73 | 0.87 | 0.94 | 0.69 | 0.75 | 0.88 | 0.85 | 0.85 | 0.84 | 0.83 | 0.83 | 0.79 | 0.88 | 0.89 | 0.94 |
| Optimized EMA Weights 2 | 0.54 | 0.71 | 0.87 | 0.94 | 0.55 | 0.63 | 0.82 | 0.86 | 0.87 | 0.84 | 0.85 | 0.88 | 0.62 | 0.78 | 0.87 | 1.00 |
| Optimized $a^*$ | 0.71 | 0.64 | 0.82 | 0.93 | 0.50 | 0.59 | 0.81 | 0.85 | 0.85 | 0.84 | 0.85 | 0.86 | 0.60 | 0.76 | 0.87 | 1.00 |
| Distance Weights 1 | 0.72 | 0.94 | 0.90 | 0.85 | 0.45 | 0.50 | 0.54 | 0.69 | 0.70 | 0.80 | 1.01 | 0.99 | 0.64 | 0.39 | 0.91 | 1.14 |
| Distance Weights 2 | 0.63 | 0.71 | 0.82 | 0.89 | 0.66 | 0.71 | 0.81 | 0.85 | 0.91 | 0.90 | 0.93 | 0.93 | 0.77 | 0.86 | 0.87 | 0.96 |
| Distance Weights 3 | 0.68 | 0.78 | 0.91 | 0.96 | 0.72 | 0.80 | 0.88 | 0.97 | 1.02 | 1.02 | 1.05 | 1.05 | 0.81 | 0.90 | 1.00 | 1.10 |

Table 6: Simulation Design 6. Reporting the Average Mean Distance Related to the Unconditional Full Sample Covariance Estimate



**Notes:**

**Realized Covariance** denotes the realized covariance estimator.

**Sample Covariance (H)** denotes the sample covariance estimator using the last half of the sample.

**LW Shrinkage (F)** denotes the Ledoit and Wolf (2003) covariance shrinkage estimator using the full sample.

**LW Shrinkage (H)** denotes the Ledoit and Wolf (2003) covariance shrinkage estimator using the last half of the sample.

**Equal Weights** denotes the covariance averaging estimator using equal weights as in Equation (11).

**EMA Weights 1** denotes the covariance averaging estimator using data dependent weights as in Equation (12) and Equation (13) ($\widehat{\alpha}_1$).

**EMA Weights 2** denotes the covariance averaging estimator using data dependent weights as in Equation (12) and Equation (13) ($\widehat{\alpha}_1$).

**Optimized Weights 1** denotes the covariance averaging estimator using the standard objective function as in Equation (28) with bias correction.

**Optimized Weights 2** denotes the covariance averaging estimator using the standard objective function as in Equation (28) without bias correction.

**Optimized EMA Weights 1** denotes the covariance averaging estimator using the standard objective function as in Equation (28), weights that are parameterized via the exponential scheme of Equation (12) and with bias correction.

**Optimized EMA Weights 1** denotes the covariance averaging estimator using the standard objective function as in Equation (28), weights that are parameterized via the exponential scheme of Equation (12) and without bias correction.

**Optimized** $a^*$ denotes the covariance averaging estimator using the unrestricted weights of Equation (31) with bias correction.

**Distance Weights 1** denotes the covariance averaging estimator using the distance-based weights from Equation (14), Equation (15) and Equation (16) - first case.

**Distance Weights 2** denotes the covariance averaging estimator using the distance-based weights from Equation (14), Equation (15) and Equation (16) - second case.

**Distance Weights 3** denotes the covariance averaging estimator using the distance-based weights from Equation (14), Equation (15) and Equation (16) - third case.

| Covariance Averaging using Exponential Time Varying True Covariance with $N = 50$ | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | $B$=(5, 20, 50) | | | | $B$=(5, 20, 50, 100) | | | | $B$=(50, 100, 200, 400) | | | | $B$=(5, 20, 50, 100, 200, 400) | | | |
| Estimator | $t$ | $t+1$ | $t+5$ | $t+10$ | $t$ | $t+1$ | $t+5$ | $t+10$ | $t$ | $t+1$ | $t+5$ | $t+10$ | $t$ | $t+1$ | $t+5$ | $t+10$ |
| Sample Covariance (H) | 0.99 | 1.01 | 1.04 | 1.10 | 0.95 | 0.95 | 0.94 | 1.17 | 1.01 | 1.01 | 1.03 | 1.03 | 0.98 | 0.97 | 0.97 | 1.00 |
| LW Shrinkage (F) | 0.92 | 0.92 | 0.93 | 0.90 | 0.99 | 0.99 | 0.99 | 0.89 | 0.95 | 0.95 | 0.93 | 0.92 | 0.93 | 0.90 | 0.90 | 0.91 |
| LW Shrinkage (H) | 0.91 | 0.93 | 0.96 | 0.98 | 0.95 | 0.95 | 0.93 | 1.01 | 0.95 | 0.96 | 0.94 | 0.94 | 0.91 | 0.88 | 0.88 | 0.91 |
| Equal Weights | 0.75 | 0.93 | 0.96 | 0.98 | 0.95 | 0.95 | 0.93 | 1.01 | 0.95 | 0.96 | 0.94 | 0.94 | 0.91 | 0.88 | 0.88 | 0.91 |
| EMA Weights 1 | 0.75 | 0.81 | 0.95 | 1.07 | 0.65 | 0.67 | 0.55 | 1.97 | 1.04 | 1.04 | 1.09 | 1.09 | 0.77 | 0.85 | 0.96 | 1.02 |
| EMA Weights 2 | 0.60 | 0.81 | 0.95 | 1.07 | 0.65 | 0.66 | 0.54 | 1.98 | 1.04 | 1.04 | 1.09 | 1.09 | 0.77 | 0.85 | 0.96 | 1.02 |
| Optimized Weights 1 | 0.70 | 0.63 | 0.91 | 1.18 | 0.28 | 0.31 | 1.49 | 4.08 | 1.14 | 1.14 | 1.27 | 1.31 | 0.39 | 0.79 | 1.21 | 1.27 |
| Optimized Weights 2 | 0.64 | 0.75 | 0.84 | 0.82 | 0.92 | 0.92 | 0.89 | 0.86 | 0.84 | 0.85 | 0.79 | 0.80 | 0.69 | 0.61 | 0.66 | 0.70 |
| Optimized EMA Weights 1 | 0.59 | 0.69 | 0.81 | 0.79 | 0.91 | 0.91 | 0.89 | 0.85 | 0.84 | 0.85 | 0.79 | 0.80 | 0.68 | 0.60 | 0.66 | 0.71 |
| Optimized EMA Weights 2 | 0.52 | 0.67 | 0.85 | 0.95 | 0.59 | 0.61 | 0.47 | 1.92 | 0.95 | 0.95 | 0.97 | 0.98 | 0.57 | 0.60 | 0.76 | 0.85 |
| Optimized $a^*$ | 0.71 | 0.61 | 0.82 | 0.91 | 0.58 | 0.60 | 0.46 | 1.91 | 0.95 | 0.95 | 0.96 | 0.98 | 0.55 | 0.59 | 0.77 | 0.86 |
| Distance Weights 1 | 0.63 | 0.67 | 0.55 | 0.79 | 0.24 | 0.19 | 0.48 | 0.90 | 1.03 | 0.78 | 1.32 | 1.53 | 0.65 | 0.90 | 0.86 | 0.75 |
| Distance Weights 2 | 0.66 | 0.72 | 0.85 | 0.86 | 0.84 | 0.85 | 0.82 | 1.15 | 0.93 | 0.94 | 0.94 | 0.94 | 0.70 | 0.71 | 0.80 | 0.87 |
| Distance Weights 3 | 0.71 | 0.77 | 0.91 | 0.99 | 0.71 | 0.72 | 0.62 | 1.73 | 1.01 | 1.01 | 1.04 | 1.04 | 0.76 | 0.82 | 0.92 | 0.99 |

Table 7: Simulation Design 7. Reporting the Average Mean Distance Related to the Unconditional Full Sample Covariance Estimate

**Notes:**

**Realized Covariance** denotes the realized covariance estimator.

**Sample Covariance (H)** denotes the sample covariance estimator using the last half of the sample.

**LW Shrinkage (F)** denotes the Ledoit and Wolf (2003) covariance shrinkage estimator using the full sample.

**LW Shrinkage (H)** denotes the Ledoit and Wolf (2003) covariance shrinkage estimator using the last half of the sample.

**Equal Weights** denotes the covariance averaging estimator using equal weights as in Equation (11).

**EMA Weights 1** denotes the covariance averaging estimator using data dependent weights as in Equation (12) and Equation (13) ($\widehat{\alpha}_1$).

**EMA Weights 2** denotes the covariance averaging estimator using data dependent weights as in Equation (12) and Equation (13) ($\widehat{\alpha}_1$).

**Optimized Weights 1** denotes the covariance averaging estimator using the standard objective function as in Equation (28) with bias correction.

**Optimized Weights 2** denotes the covariance averaging estimator using the standard objective function as in Equation (28) without bias correction.

**Optimized EMA Weights 1** denotes the covariance averaging estimator using the standard objective function as in Equation (28), weights that are parameterized via the exponential scheme of Equation (12) and with bias correction.

**Optimized EMA Weights 1** denotes the covariance averaging estimator using the standard objective function as in Equation (28), weights that are parameterized via the exponential scheme of Equation (12) and without bias correction.

**Optimized** $a^*$ denotes the covariance averaging estimator using the unrestricted weights of Equation (31) with bias correction.

**Distance Weights 1** denotes the covariance averaging estimator using the distance-based weights from Equation (14), Equation (15) and Equation (16) - first case.

**Distance Weights 2** denotes the covariance averaging estimator using the distance-based weights from Equation (14), Equation (15) and Equation (16) - second case.

**Distance Weights 3** denotes the covariance averaging estimator using the distance-based weights from Equation (14), Equation (15) and Equation (16) - third case.

Covariance Averaging Performance using Bivariate Scalar Diagonal VECH across all $\rho$

| | $B$=(5, 20, 50) | | | | $B$=(5, 20, 50, 100) | | | | $B$=(50, 100, 200, 400) | | | | $B$=(5, 20, 50, 100, 200, 400) | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Estimator | $t$ | $t+1$ | $t+5$ | $t+10$ | $t$ | $t+1$ | $t+5$ | $t+10$ | $t$ | $t+1$ | $t+5$ | $t+10$ | $t$ | $t+1$ | $t+5$ | $t+10$ |
| Sample Covariance (H) | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| LW Shrinkage (F) | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| LW Shrinkage (H) | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| Equal Weights | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| EMA Weights 1 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| EMA Weights 2 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| Optimized Weights 1 | 0.0 | 0.0 | 0.0 | 1.0 | 0.0 | 0.0 | 0.0 | 0.2 | 0.0 | 0.0 | 0.0 | 0.2 | 0.0 | 0.0 | 0.0 | 0.2 |
| Optimized Weights 2 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.6 | 0.0 | 0.0 | 0.0 | 0.0 |
| Optimized EMA Weights 1 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.2 | 0.0 |
| Optimized EMA Weights 2 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.2 | 0.2 | 0.2 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| Optimized $a^*$ | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| Distance Weights 1 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.8 | 0.8 | 0.8 | 0.2 | 0.0 | 0.0 | 0.0 | 0.0 |
| Distance Weights 2 | 0.0 | 0.0 | 0.4 | 0.0 | 0.0 | 0.0 | 1.0 | 0.8 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.8 | 0.8 |
| Distance Weights 3 | 1.0 | 1.0 | 0.6 | 0.0 | 1.0 | 1.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |

Table 8: Reporting the Average Performance Across Simulation Designs 1-5

**Notes:**

**Realized Covariance** denotes the realized covariance estimator.

**Sample Covariance (H)** denotes the sample covariance estimator using the last half of the sample.

**LW Shrinkage (F)** denotes the Ledoit and Wolf (2003) covariance shrinkage estimator using the full sample.

**LW Shrinkage (H)** denotes the Ledoit and Wolf (2003) covariance shrinkage estimator using the last half of the sample.

**Equal Weights** denotes the covariance averaging estimator using equal weights as in Equation (11).

**EMA Weights 1** denotes the covariance averaging estimator using data dependent weights as in Equation (12) and Equation (13) ($\widehat{\alpha}_1$).

**EMA Weights 2** denotes the covariance averaging estimator using data dependent weights as in Equation (12) and Equation (13) ($\widehat{\alpha}_1$).

**Optimized Weights 1** denotes the covariance averaging estimator using the standard objective function as in Equation (28) with bias correction.

**Optimized Weights 2** denotes the covariance averaging estimator using the standard objective function as in Equation (28) without bias correction.

**Optimized EMA Weights 1** denotes the covariance averaging estimator using the standard objective function as in Equation (28), weights that are parameterized via the exponential scheme of Equation (12) and with bias correction.

**Optimized EMA Weights 1** denotes the covariance averaging estimator using the standard objective function as in Equation (28), weights that are parameterized via the exponential scheme of Equation (12) and without bias correction.

**Optimized $a^*$** denotes the covariance averaging estimator using the unrestricted weights of Equation (31) with bias correction.

**Distance Weights 1** denotes the covariance averaging estimator using the distance-based weights from Equation (14), Equation (15) and Equation (16) - first case.

**Distance Weights 2** denotes the covariance averaging estimator using the distance-based weights from Equation (14), Equation (15) and Equation (16) - second case.

**Distance Weights 3** denotes the covariance averaging estimator using the distance-based weights from Equation (14), Equation (15) and Equation (16) - third case.

| Covariance Averaging Performance using Exponential Time Varying True Covariance across $N$ | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | $B$=(5, 20, 50) | | | | $B$=(5, 20, 50, 100) | | | | $B$=(50, 100, 200, 400) | | | | $B$=(5, 20, 50, 100, 200, 400) | | | |
| Estimator | $t$ | $t+1$ | $t+5$ | $t+10$ | $t$ | $t+1$ | $t+5$ | $t+10$ | $t$ | $t+1$ | $t+5$ | $t+10$ | $t$ | $t+1$ | $t+5$ | $t+10$ |
| Sample Covariance (H) | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| LW Shrinkage (F) | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| LW Shrinkage (H) | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| Equal Weights | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| EMA Weights 1 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| EMA Weights 2 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| Optimized Weights 1 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.5 | 0.0 | 0.0 | 0.0 |
| Optimized Weights 2 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.5 | 0.5 | 0.0 | 0.0 | 0.5 | 1.0 |
| Optimized EMA Weights 1 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.5 | 0.5 | 0.0 | 0.5 | 0.5 | 0.0 | 0.0 | 0.0 | 0.0 |
| Optimized EMA Weights 2 | 1.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| Optimized $a^*$ | 0.0 | 1.0 | 0.5 | 0.0 | 0.0 | 0.0 | 0.5 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.5 | 0.5 | 0.5 | 0.0 |
| Distance Weights 1 | 0.0 | 0.0 | 0.5 | 0.5 | 0.5 | 1.0 | 0.5 | 0.5 | 0.5 | 1.0 | 0.0 | 0.0 | 0.0 | 0.5 | 0.0 | 0.0 |
| Distance Weights 2 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| Distance Weights 3 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |

Table 9: Reporting the Average Performance Across Simulation Designs 6-7



**Notes:**

**Realized Covariance** denotes the realized covariance estimator.

**Sample Covariance (H)** denotes the sample covariance estimator using the last half of the sample.

**LW Shrinkage (F)** denotes the Ledoit and Wolf (2003) covariance shrinkage estimator using the full sample.

**LW Shrinkage (H)** denotes the Ledoit and Wolf (2003) covariance shrinkage estimator using the last half of the sample.

**Equal Weights** denotes the covariance averaging estimator using equal weights as in Equation (11).

**EMA Weights 1** denotes the covariance averaging estimator using data dependent weights as in Equation (12) and Equation (13) ($\widehat{\alpha}_1$).

**EMA Weights 2** denotes the covariance averaging estimator using data dependent weights as in Equation (12) and Equation (13) ($\widehat{\alpha}_1$).

**Optimized Weights 1** denotes the covariance averaging estimator using the standard objective function as in Equation (28) with bias correction.

**Optimized Weights 2** denotes the covariance averaging estimator using the standard objective function as in Equation (28) without bias correction.

**Optimized EMA Weights 1** denotes the covariance averaging estimator using the standard objective function as in Equation (28), weights that are parameterized via the exponential scheme of Equation (12) and with bias correction.

**Optimized EMA Weights 1** denotes the covariance averaging estimator using the standard objective function as in Equation (28), weights that are parameterized via the exponential scheme of Equation (12) and without bias correction.

**Optimized $a^*$** denotes the covariance averaging estimator using the unrestricted weights of Equation (31) with bias correction.

**Distance Weights 1** denotes the covariance averaging estimator using the distance-based weights from Equation (14), Equation (15) and Equation (16) - first case.

**Distance Weights 2** denotes the covariance averaging estimator using the distance-based weights from Equation (14), Equation (15) and Equation (16) - second case.

**Distance Weights 3** denotes the covariance averaging estimator using the distance-based weights from Equation (14), Equation (15) and Equation (16) - third case.

Covariance Averaging Performance Across all Simulation Designs

| | $B$=(5, 20, 50) | | | | $B$=(5, 20, 50, 100) | | | | $B$=(50, 100, 200, 400) | | | | $B$=(5, 20, 50, 100, 200, 400) | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Estimator | $t$ | $t+1$ | $t+5$ | $t+10$ | $t$ | $t+1$ | $t+5$ | $t+10$ | $t$ | $t+1$ | $t+5$ | $t+10$ | $t$ | $t+1$ | $t+5$ | $t+10$ |
| Sample Covariance (H) | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| LW Shrinkage (F) | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| LW Shrinkage (H) | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| Equal Weights | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| EMA Weights 1 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| EMA Weights 2 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| Optimized Weights 1 | 0.00 | 0.00 | 0.00 | 0.71 | 0.00 | 0.00 | 0.00 | 0.14 | 0.00 | 0.00 | 0.00 | 0.14 | 0.14 | 0.00 | 0.00 | 0.14 |
| Optimized Weights 2 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.14 | 0.57 | 0.00 | 0.00 | 0.14 | 0.29 |
| Optimized EMA Weights 1 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.14 | 0.14 | 0.00 | 0.14 | 0.14 | 0.00 | 0.00 | 0.14 | 0.00 |
| Optimized EMA Weights 2 | 0.29 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.14 | 0.14 | 0.14 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| Optimized $a^*$ | 0.00 | 0.29 | 0.14 | 0.00 | 0.00 | 0.00 | 0.14 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.14 | 0.14 | 0.14 | 0.00 |
| Distance Weights 1 | 0.00 | 0.00 | 0.14 | 0.14 | 0.14 | 0.29 | 0.14 | 0.14 | 0.71 | 0.86 | 0.57 | 0.14 | 0.00 | 0.14 | 0.00 | 0.00 |
| Distance Weights 2 | 0.00 | 0.00 | 0.29 | 0.00 | 0.00 | 0.00 | 0.71 | 0.57 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.57 | 0.57 |
| Distance Weights 3 | 0.71 | 0.71 | 0.43 | 0.00 | 0.71 | 0.71 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |

Table 10: Reporting the Average Performance Across all Simulation Designs

**Notes:**

**Realized Covariance** denotes the realized covariance estimator.

**Sample Covariance (H)** denotes the sample covariance estimator using the last half of the sample.

**LW Shrinkage (F)** denotes the Ledoit and Wolf (2003) covariance shrinkage estimator using the full sample.

**LW Shrinkage (H)** denotes the Ledoit and Wolf (2003) covariance shrinkage estimator using the last half of the sample.

**Equal Weights** denotes the covariance averaging estimator using equal weights as in Equation (11).

**EMA Weights 1** denotes the covariance averaging estimator using data dependent weights as in Equation (12) and Equation (13) ($\widehat{\alpha}_1$).

**EMA Weights 2** denotes the covariance averaging estimator using data dependent weights as in Equation (12) and Equation (13) ($\widehat{\alpha}_1$).

**Optimized Weights 1** denotes the covariance averaging estimator using the standard objective function as in Equation (28) with bias correction.

**Optimized Weights 2** denotes the covariance averaging estimator using the standard objective function as in Equation (28) without bias correction.

**Optimized EMA Weights 1** denotes the covariance averaging estimator using the standard objective function as in Equation (28), weights that are parameterized via the exponential scheme of Equation (12) and with bias correction.

**Optimized EMA Weights 1** denotes the covariance averaging estimator using the standard objective function as in Equation (28), weights that are parameterized via the exponential scheme of Equation (12) and without bias correction.

**Optimized $a^*$** denotes the covariance averaging estimator using the unrestricted weights of Equation (31) with bias correction.

**Distance Weights 1** denotes the covariance averaging estimator using the distance-based weights from Equation (14), Equation (15) and Equation (16) - first case.

**Distance Weights 2** denotes the covariance averaging estimator using the distance-based weights from Equation (14), Equation (15) and Equation (16) - second case.

**Distance Weights 3** denotes the covariance averaging estimator using the distance-based weights from Equation (14), Equation (15) and Equation (16) - third case.

| | Sample-Full | Sample-Half | LW-Full | LW-Half |
|---|---|---|---|---|
| | A. Success rates across all runs - SP08 | | | |
| Average | 0.00% | 7.69% | 7.69% | 84.62% |
| Volatility | 76.92% | 69.23% | 92.31% | 92.31% |
| Sharpe | 7.69% | 7.69% | 7.69% | 84.62% |
| Cumulative | 0.00% | 7.69% | 7.69% | 84.62% |
| Drawdown | 7.69% | 7.69% | 76.92% | 76.92% |
| | B. Success rates across all runs - SP20 | | | |
| Average | 23.08% | 15.38% | 100.00% | 92.31% |
| Volatility | 30.77% | 30.77% | 23.08% | 0.00% |
| Sharpe | 0.00% | 0.00% | 100.00% | 69.23% |
| Cumulative | 30.77% | 7.69% | 100.00% | 69.23% |
| Drawdown | 0.00% | 0.00% | 0.00% | 0.00% |
| | C. Success rates across all runs - SP40 | | | |
| Average | 92.31% | 84.62% | 92.31% | 84.62% |
| Volatility | 53.85% | 38.46% | 53.85% | 53.85% |
| Sharpe | 76.92% | 46.15% | 76.92% | 69.23% |
| Cumulative | 84.62% | 76.92% | 84.62% | 84.62% |
| Drawdown | 84.62% | 69.23% | 84.62% | 61.54% |

Table 11: Application in Portfolios. Reporting the Success Rates Across All Runs.

| | Sample-Full | Sample-Half | LW-Full | LW-Half |
|---|---|---|---|---|
| | A. Success rates within methods - SP08 | | | |
| Average | 26.71% | 34.19% | 49.15% | 66.03% |
| Volatility | 63.46% | 51.50% | 73.50% | 67.31% |
| Sharpe | 32.26% | 33.55% | 51.71% | 69.23% |
| Cumulative | 28.21% | 33.97% | 48.29% | 66.03% |
| Drawdown | 38.46% | 32.91% | 51.92% | 48.29% |
| | B. Success rates within methods - SP20 | | | |
| Average | 41.60% | 44.73% | 69.37% | 64.10% |
| Volatility | 50.57% | 51.14% | 48.01% | 38.89% |
| Sharpe | 38.89% | 43.45% | 65.81% | 59.40% |
| Cumulative | 40.74% | 43.02% | 67.24% | 63.11% |
| Drawdown | 42.31% | 45.01% | 45.87% | 41.45% |
| | C. Success rates within methods - SP40 | | | |
| Average | 51.71% | 51.57% | 52.71% | 52.71% |
| Volatility | 55.40% | 47.14% | 56.11% | 50.41% |
| Sharpe | 55.13% | 53.13% | 55.70% | 54.27% |
| Cumulative | 51.14% | 51.85% | 53.13% | 53.42% |
| Drawdown | 68.74% | 58.44% | 69.17% | 59.16% |

Table 12: Application in Portfolios. Reporting the Success Rates Within Methods.

| | A. Cumulative return differences in between methods - SP40 | | | | B. Drawdown differences in between methods - SP40 | | | | C. Sharpe ratio differences in between methods - SP40 | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Sample-Full | Sample-Half | LW-Full | LW-Half | Sample-Full | Sample-Half | LW-Full | LW-Half | Sample-Full | Sample-Half | LW-Full | LW-Half |
| | No Rebalance | | | | No Rebalance | | | | No Rebalance | | | |
| Proportion | 92.31% | 69.23% | 92.31% | 76.92% | 84.62% | 53.85% | 76.92% | 53.85% | 92.31% | 23.08% | 92.31% | 30.77% |
| Mean Difference | -7.73% | -1.16% | -7.31% | -2.35% | 1.18% | -0.02% | 1.09% | 0.02% | -0.74% | 0.26% | -0.72% | 0.12% |
| | Time only | | | | Time only | | | | Time only | | | |
| Proportion | 92.31% | 76.92% | 84.62% | 76.92% | 0.00% | 46.15% | 0.00% | 53.85% | 76.92% | 69.23% | 76.92% | 69.23% |
| Mean Difference | -6.05% | -1.41% | -4.86% | -1.77% | 1.34% | 0.24% | 1.09% | -0.06% | -1.37% | -0.50% | -1.18% | -0.55% |
| | Thresh. #1 only | | | | Thresh. #1 only | | | | Thresh. #1 only | | | |
| Proportion | 15.38% | 38.46% | 38.46% | 69.23% | 7.69% | 38.46% | 7.69% | 38.46% | 23.08% | 23.08% | 38.46% | 53.85% |
| Mean Difference | 7.35% | 1.26% | 1.67% | -3.61% | 1.55% | -0.02% | 1.60% | 0.45% | 0.36% | 0.44% | -0.08% | -0.19% |
| | Thresh. #2 only | | | | Thresh. #2 only | | | | Thresh. #2 only | | | |
| Proportion | 84.62% | 84.62% | 84.62% | 84.62% | 23.08% | 38.46% | 15.38% | 38.46% | 76.92% | 69.23% | 76.92% | 76.92% |
| Mean Difference | -7.75% | -3.77% | -8.43% | -7.40% | 1.50% | 0.30% | 1.60% | 0.09% | -0.69% | -0.16% | -0.75% | -0.45% |
| | T & T #1 | | | | T & T #1 | | | | T & T #1 | | | |
| Proportion | 92.31% | 76.92% | 92.31% | 76.92% | 7.69% | 30.77% | 7.69% | 30.77% | 92.31% | 53.85% | 92.31% | 53.85% |
| Mean Difference | -8.25% | -3.84% | -8.83% | -3.98% | 2.16% | 0.75% | 2.21% | 0.71% | -0.82% | -0.09% | -0.86% | -0.18% |
| | T &T #2 | | | | T &T #2 | | | | T &T #2 | | | |
| Proportion | 84.62% | 84.62% | 84.62% | 84.62% | 23.08% | 38.46% | 15.38% | 46.15% | 76.92% | 61.54% | 76.92% | 76.92% |
| Mean Difference | -2.76% | -4.83% | -4.19% | -6.39% | 1.46% | 0.23% | 1.59% | -0.04% | -0.45% | -0.25% | -0.55% | -0.43% |

Table 13: Application in Portfolios. Reporting the Success Rates In Between Methods.

| SP08 | SP20 | | SP40 | | | |
|---|---|---|---|---|---|---|
| CMCSA | AAPL | MSFT | AAPL | CVS | LLY | PG |
| DD | CMCSA | PFE | ABT | CVX | MCD | SLB |
| DUK | CVX | PG | AMGN | DIS | MMM | T |
| GE | DD | SO | AXP | GE | MO | UNP |
| JNJ | DOW | T | BAC | HD | MRK | USB |
| T | DUK | UTX | BMY | IBM | MSFT | UTX |
| WMT | GE | VZ | C | INTC | ORCL | VZ |
| XOM | HD | WFC | CMCSA | JNJ | OXY | WFC |
| | JNJ | WMT | COP | JPM | PEP | WMT |
| | JPM | XOM | CSCO | KO | PFE | XOM |

Table 14: Application in Portfolios. Tickers of securities used in the three portfolios.