Covariance Averaging

for Improved Estimation & Portfolio Allocation

Dimitrios Thomakos & Fotis Papailias

Financial Markets & Portfolio Management, 2015

UoP & QUB

Research Skills Seminar - Ph.D. Program at Econ-UoP

What's this all about?

- Recent literature in forecasting deals with performance improvements via *averaging of different data segments* (rolling and recursive windows). Why not use a similar approach in terms of *covariance estimation*?
- Large(r)-scale covariance/correlation & portfolio problems are plagued by difficulties of *over-parametrization* and *numerical optimization*.
- Improved covariance estimation is needed for *portfolio allocation* problems & the construction of *hedging* strategies in various asset classes.
- A simpler/more robust method of covariance/correlation estimation might be *competitive to existing parametric models* even in small(er) dimensions.
- Open questions that we address: set-up for covariance averaging; selection of weights; optimization of weights; performance assessment against the sample covariance and covariance shrinkage.

Stuff that has been said before...

- On covariance shrinkage see Ledoit and Wolf (2003, 2004) and reference therein, Wang (2005), Kwan (2008), Kourtis et al. (2012), Bajeux-Besnainou et al. (forthcoming).
- On large(r) scale covariance estimation see (among others) Chan et al. (1999), Engle (2002), Ledoit, Santa Clara and Wolf (2003), Andersen et al. (2006), Bauwens et al. (2006), Pelletier (2006), Fan et al. (2008), Palandri (2009), Silvennoinen and Terasvirta (2009), Huo et al. (2011).
- On portfolio allocation problems (where our results can be of use) see (among others) Kan and Zhou (2007), Miguel et al. (2007), Martellini and Ziemann (2009).
- On rolling window averaging see Pesaran, Schuermann and Smoth (2009), Clark and McCracken (2009), Rossin and Inoue (2011), Bhattacharya and Thomakos (2011).

The background

- We have N assets whose returns at t are denoted by $\boldsymbol{R}_t \stackrel{\text{def}}{=} [R_{t1}, \ldots, R_{tN}]^\top$. We assume that they have an unknown conditional distribution with mean $\boldsymbol{\mu}_t$ and covariance matrix $\boldsymbol{\Sigma}_t$:

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

- We need not make particular assumptions about the process of the returns beyond that they have a conditional distribution but we will 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 covariance shrinkage literature.
- For the rest of our discussion we denote by $\boldsymbol{r}_t \stackrel{\text{def}}{=} \boldsymbol{R}_t - \widehat{\boldsymbol{\mu}}_t$ the suitably demeaned 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 $\boldsymbol{\Sigma}_t$.

Shrink the covariance - Model averaging

- 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 $\widehat{\mathbf{\Sigma}}_t(t)$:

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

  and a highly structured estimator denoted here by $\boldsymbol{S}$. Note that this is *model averaging*!
- 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.
- The linear combination, i.e. the averaged covariance, can be represented as:

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

  where δ is the shrinkage (i.e. averaging) coefficient.

An objective function

- Formally, the optimal choice $\widehat{\delta}$ is obtained as a solution to the following MSE-type minimisation problem:

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

- Letting σ_{ij} and $\widehat{\sigma}_{ij}^S(t,\delta)$ denote the corresponding elements of the matrices the above is expressed as:

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

- For any given value of δ the above can be directly estimated (under the i.i.d. and finite fourth moment assumption) by:

$$\widehat{\pi}(\delta) \stackrel{\text{def}}{=} \sum_{i=1}^{N}\sum_{j=1}^{N}\widehat{\pi}_{ij}(\delta), \qquad \widehat{\pi}_{ij}(\delta) \stackrel{\text{def}}{=} \frac{1}{t}\sum_{h=1}^{t}\left[r_{ih}r_{jh} - \widehat{\sigma}_{ij}^S(t,\delta)\right]^2 \tag{6}$$

assuming unbiasedness.

Rolling window averaging

- We turn next to setting up a more general framework for covariance averaging based on rolling window covariance estimators.
- Compute the sample covariance matrix using different segments of the data, either overlapping or non-overlapping, and average the resulting covariances.
- Consider a sequence of overlapping windows $B \stackrel{\text{def}}{=} (m_1, m_2, \ldots, m_M)$ where $1 < m_1 < m_2 < \cdots < m_M < t$. Using the last m_s observations the sample covariance is estimated:

$$\widehat{\boldsymbol{\Sigma}}_t(m_s) \stackrel{\text{def}}{=} \frac{1}{m_s} \sum_{i=t-m_s+1}^{t} \boldsymbol{r}_i \boldsymbol{r}^\top \tag{7}$$

and we then average as follows:

$$\widehat{\boldsymbol{\Sigma}}_t^A \stackrel{\text{def}}{=} \sum_{s=1}^{M} w_s \widehat{\boldsymbol{\Sigma}}_t(m_s) \tag{8}$$

How to weight the rolling estimates? Heuristics first

- The main problem is how to choose the averaging weights - either heuristically or "optimally" as in covariance shrinkage.
- The simplest case is, naturally, to assign equal weights to all rolling estimates:

$$w_s^E \stackrel{\text{def}}{=} \frac{1}{M} \tag{9}$$

- Alternatively, if one wants to assign greater weight to the most recent data:

$$w_s^X(\alpha) \stackrel{\text{def}}{=} \frac{(1-\alpha)^{s-1}}{\sum_{s=1}^{M}(1-\alpha)^{s-1}} \tag{10}$$

  where $\alpha \in [0, 1]$ is the smoothing parameter, whose selection we discuss later.
- The next approach is based on the shrinkage objective function adapted to the context of averaging. The idea is one of assigning weights based on expected distances from a target, the true covariance $\mathbf{\Sigma}$. It is a variation of nearest neighbours!

- Consider thus the covariance estimate based on the m_s window and write:

$$d_s \stackrel{\text{def}}{=} \mathsf{E}\|\boldsymbol{\Sigma} - \widehat{\boldsymbol{\Sigma}}_t(m_s)\|_F^2 = \sum_{i=1}^{N}\sum_{j=1}^{N} \mathrm{Var}\left[\widehat{\sigma}_{ij}(m_s)\right] \tag{11}$$

Note that this does not depend on any parameters, as the δ in shrinkage, and can be directly estimated:

$$\widehat{\pi}(m_s) \stackrel{\text{def}}{=} \sum_{i=1}^{N}\sum_{j=1}^{N} \widehat{\pi}_{ij}(m_s), \qquad \widehat{\pi}_{ij}(m_s) \stackrel{\text{def}}{=} \frac{1}{m_s} \sum_{h=t-m_s+1}^{t} \left[r_{ih}r_{jh} - \widehat{\sigma}_{ij}(m_s)\right]^2 \tag{12}$$

These estimated distances are then used to construct weights for averaging which are inversely related to their magnitude: a higher distance gets lets weight and vice versa.

- For example, we can have such heuristics as:

$$
\begin{aligned}
&\lambda_s \stackrel{\text{def}}{=} d_s^{-1}, \qquad w_s^D \stackrel{\text{def}}{=} \frac{\lambda_s}{\sum_{s=1}^M \lambda_s} \\
&\lambda_s \stackrel{\text{def}}{=} \frac{\sum_{j\neq s}^M d_j}{\sum_{j=1}^M d_j}, \qquad w_s^D \stackrel{\text{def}}{=} \frac{\lambda_s}{\sum_{s=1}^M \lambda_s} \\
&\kappa_s \stackrel{\text{def}}{=} \exp\left[\frac{d_s}{\sum_{s=1}^M d_s}\right], \qquad \lambda_s \stackrel{\text{def}}{=} \frac{\sum_{j\neq s}^M \kappa_j}{\sum_{j=1}^M \kappa_j}, \qquad w_s^D \stackrel{\text{def}}{=} \frac{\lambda_s}{\sum_{s=1}^M \lambda_s}
\end{aligned}
\tag{13}
$$

We still need some 'optimal' weights!

- We can get them, under the same assumptions that are used in shrinkage.
- Letting $\boldsymbol{w} \stackrel{\text{def}}{=} [w_1, w_2, \ldots, w_M]^\top$ be the vector of weights, we have that the general set-up for the optimization problem is given by the same objective function used in covariance shrinkage as in:

$$Q(\boldsymbol{w}) \stackrel{\text{def}}{=} \mathsf{E}\left\|\boldsymbol{\Sigma} - \widehat{\boldsymbol{\Sigma}}_t^A\right\|_F^2 = \mathsf{E}\left\|\boldsymbol{\Sigma} - \sum_{s=1}^{M} w_s \widehat{\boldsymbol{\Sigma}}_t(m_j)\right\|_F^2 \tag{14}$$

Expanding the above we get:

$$\begin{aligned} Q(\boldsymbol{w}) \quad = \quad & \textstyle\sum_{i=1}^N \sum_{j=1}^N \left\{\sum_{s=1}^M w_s^2 \mathsf{Var}\left[\widehat{\sigma}_{ij}(m_s)\right]\right\} \\ & + \textstyle\sum_{i=1}^N \sum_{j=1}^N \left\{2\sum_{k\neq s} w_k w_s \mathsf{Cov}\left[\widehat{\sigma}_{ij}(m_k), \widehat{\sigma}_{ij}(m_s)\right]\right\} \\ & + \textstyle\sum_{i=1}^N \sum_{j=1}^N \left(\mathsf{E}\left[\sigma_{ij} - \sum_{s=1}^M w_s \widehat{\sigma}_{ij}(m_s)\right]\right)^2 \end{aligned} \tag{15}$$

which now has an extra term!

- This creates a potential problem since these covariance will be 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.
- To see how the above works consider the simple case of $M = 2$ and note that we have the following representations:

$$\begin{aligned} \widehat{\boldsymbol{\Sigma}}_t(m_1) &= m_1^{-1} \sum_{i=t-m_1+1}^{t} \boldsymbol{r}_i \boldsymbol{r}_i^\top \\ \widehat{\boldsymbol{\Sigma}}_t(m_2) &= m_2^{-1} \sum_{i=t-m_2+1}^{t} \boldsymbol{r}_i \boldsymbol{r}_i^\top \end{aligned} \tag{16}$$

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

$$\begin{aligned}\widehat{\boldsymbol{\Sigma}}_t(m_2) &= m_2^{-1}\textstyle\sum_{i=t-m_2+1}^{t-m_1}\boldsymbol{r}_i\boldsymbol{r}_i^\top + m_2^{-1}\displaystyle\sum_{i=t-m_1+1}^{t}\boldsymbol{r}_i\boldsymbol{r}_i^\top \\ &= [(m_2-m_1)/m_2]\,\widehat{\boldsymbol{\Sigma}}_t(m_2-m_1) + (m_1/m_2)\widehat{\boldsymbol{\Sigma}}_t(m_1)\end{aligned} \tag{17}$$

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

$$\begin{aligned}\widehat{\boldsymbol{\Sigma}}_t^A &= [w_1 + w_2(m_1/m_2)]\,\widehat{\boldsymbol{\Sigma}}_t(m_1) + w_2\,[(m_2-m_1)/m_2]\,\widehat{\boldsymbol{\Sigma}}_t(m_2-m_1) \\ &= a_1\widehat{\boldsymbol{\Sigma}}_t(m_1) + a_2\widehat{\boldsymbol{\Sigma}}_t(m_2-m_1)\end{aligned} \tag{18}$$

We can easily generalize the above discussion when $M > 2$. Noticing that the new weights a_s depend on some of the old weights w_s we first define the new weights formally as:

$$
\begin{aligned}
a_1(\boldsymbol{w}_0) &\stackrel{\text{def}}{=} \sum_{s=1}^{M} \frac{m_1}{m_s} w_s \text{ with } \boldsymbol{w}_0 = [w_1, \ldots, w_M]^\top \\
a_2(\boldsymbol{w}_{-1}) &\stackrel{\text{def}}{=} \sum_{s=2}^{M} \frac{m_2 - m_1}{m_s} w_s \text{ with } \boldsymbol{w}_{-1} = [w_2, \ldots, w_M]^\top \\
a_3(\boldsymbol{w}_{-2}) &\stackrel{\text{def}}{=} \sum_{s=3}^{M} \frac{m_3 - m_2}{m_s} w_s \text{ with } \boldsymbol{w}_{-2} = [w_3, \ldots, w_M]^\top \\
\vdots \quad &\vdots \quad \vdots \\
a_M(\boldsymbol{w}_{-M+1}) &\stackrel{\text{def}}{=} \frac{m_M - m_{M-1}}{m_M} w_M \text{ with } \boldsymbol{w}_{-M+1} = w_M
\end{aligned}
\tag{19}
$$

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

$$
\begin{aligned}
\widehat{\boldsymbol{\Sigma}}_t(m_1) &\stackrel{\text{def}}{=} m_1^{-1} \sum_{i=t-m_1+1}^{t} \boldsymbol{r}_i \boldsymbol{r}_i^\top \\
\widehat{\boldsymbol{\Sigma}}_t(m_s - m_{s-1}) &\stackrel{\text{def}}{=} (m_s - m_{s-1})^{-1} \sum_{i=t-m_{s-1}+1}^{t-m_s} \boldsymbol{r} \boldsymbol{r}_i^\top
\end{aligned}
\tag{20}
$$

for $s = 2, \ldots, M$ with $m_0 \equiv 0$.

With these we can re-write the objective function of equation (17) as:

$$\begin{aligned} Q(\boldsymbol{w}) &= \mathrm{E}\left\| \boldsymbol{\Sigma} - \sum_{s=1}^{M} a_s(\boldsymbol{w}_{-s+1})\widehat{\boldsymbol{\Sigma}}_t(m_s - m_{s-1}) \right\|_F^2 \\ &= \sum_{i=1}^{N}\sum_{j=1}^{N}\left\{\sum_{s=1}^{M} a_s^2(\boldsymbol{w}_{-s+1})\mathrm{Var}\left[\widehat{\sigma}_{ij}(m_s - m_{s-1})\right]\right\} \\ &\quad + \sum_{i=1}^{N}\sum_{j=1}^{N}\left(\mathrm{E}\left[\sigma_{ij} - \sum_{s=1}^{M} a_s(\boldsymbol{w}_{-s+1})\widehat{\sigma}_{ij}(m_s - m_{s-1})\right]\right)^2 \end{aligned} \tag{21}$$

which does not involve the covariance terms $\mathrm{Cov}\left[\widehat{\sigma}_{ij}(m_1), \widehat{\sigma}_{ij}(m_2 - m_1)\right]$. The optimization of the objective function is most easily done numerically:

- compute the estimates for the variance terms, the expectation terms and the composite weights (for a given value of the original weights w_s), then impose the restrictions on the weights (either the original or the composite) and optimize the objective function directly.
- Under the i.i.d. assumption and the properties of the composite weights we can work without the expectation terms.

Do the optimal weights mean anything?

- It would be nice to have an explicit formulation and interpretation of the weights that come from averaging.
- The transformation from $w_s \mapsto a(\boldsymbol{w}_{-s+1})$ allows us to do so and, in the process, obtain an explicit expression of these composite weights that is amenable to a nice interpretation.
- Let us start with the first term in equation (21), i.e. ignoring the bias terms, and pass the double-summation inside the curly brackets to obtain:

$$Q(\boldsymbol{a}) = \sum_{s=1}^{M} a_s^2(\boldsymbol{w}_{-s+1})\pi(m_s - m_{s-1}) = \boldsymbol{a}^\top \boldsymbol{\Pi} \boldsymbol{a} \qquad (22)$$

where we defined $\boldsymbol{a} \stackrel{\text{def}}{=} [a_1(\boldsymbol{w}_0), a_2(\boldsymbol{w}_{-1}), \ldots, a_M(\boldsymbol{w}_{-M+1})]^\top$ as the vector of composite weights and
$\boldsymbol{\Pi} \stackrel{\text{def}}{=} \text{diag}\left[\pi(m_1), \pi(m_2 - m_1), \ldots, \pi(m_M - m_{M-1})\right]$ 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 $\boldsymbol{a}$:

$$\Lambda(\boldsymbol{a}) = \boldsymbol{a}^\top \boldsymbol{\Pi} \boldsymbol{a} + 2\lambda(1 - \boldsymbol{e}^\top \boldsymbol{a}) \tag{23}$$

where $\boldsymbol{e}$ is a vector of ones. The solution is:

$$\boldsymbol{a}^* \stackrel{\text{def}}{=} \operatorname{argmin} \Lambda(\alpha) \equiv \frac{\boldsymbol{\Pi}^{-1}\boldsymbol{e}}{\boldsymbol{e}^\top \boldsymbol{\Pi}^{-1}\boldsymbol{e}} \tag{24}$$

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

$$a_s^* \stackrel{\text{def}}{=} \frac{\pi^{-1}(m_s - m_{s-1})}{\sum_{s=1}^{M} \pi^{-1}(m_s - m_{s-1})} \tag{25}$$

i.e., *the weights assigned to the rolling window covariances for averaging are inversely proportional to the asymptotic variances.* Higher estimation risk vis-a-vis the true covariance leads to a lower weight in constructing the averaged estimate. Note also tht this result justifies the heuristics presented in equation (13) which work essentially on the same premise but using one estimate at a time.

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

$$\Lambda(\boldsymbol{a}) = \boldsymbol{a}^\top \boldsymbol{\Pi} \boldsymbol{a} + \sum_{i=1}^{N}\sum_{j=1}^{N}\left[\sigma_{ij} - \boldsymbol{E}_{ij}^\top \boldsymbol{a}\right]^2 + 2\lambda(1 - \boldsymbol{e}^\top \boldsymbol{a}) \tag{26}$$

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

$$\begin{aligned} \frac{1}{2}\frac{\partial \Lambda(\boldsymbol{a})}{\partial \boldsymbol{a}} &= \left(\boldsymbol{\Pi} + \sum_{i=1}^{N}\sum_{j=1}^{N} \boldsymbol{E}_{ij}\boldsymbol{E}_{ij}^\top\right)\boldsymbol{a} - \sum_{i=1}^{N}\sum_{j=1}^{N} \boldsymbol{E}_{ij}\sigma_{ij} - \lambda \boldsymbol{e} \\ &= \boldsymbol{V}\boldsymbol{a} - \boldsymbol{b} - \lambda \boldsymbol{e} \end{aligned} \tag{27}$$

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 condition for the Lagrange multiplier we end up with the new solution:

$$\boldsymbol{a}^* = \operatorname{argmin} \Lambda(\boldsymbol{a}) = \boldsymbol{V}^{-1}\boldsymbol{b} + \left(1 - \boldsymbol{e}^\top \boldsymbol{V}^{-1}\boldsymbol{b}\right) \frac{\boldsymbol{V}^{-1}\boldsymbol{e}}{\boldsymbol{e}^\top \boldsymbol{V}^{-1}\boldsymbol{e}} \tag{28}$$

Note that there are three parts in the new weights: 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 (25). Note that when we do not take into account the bias term the solution in the above equation collapses to that of equation (26) and, therefore, both equations have the same interpretation.

OK, the math is nice but does it work?

- We first consider the data generating process (DGP) of Patton and Sheppard (2008) 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.
- Then, we consider another DGP that conforms a bit more closely with the idea of covariance averaging and is amenable to an analysis where N is large.
- For DGP #1 We take $N = 2$ and have:

$$\begin{aligned} \boldsymbol{r}_t &\stackrel{\text{def}}{=} \boldsymbol{\Sigma}_t^{1/2}\boldsymbol{\epsilon}_t \\ \boldsymbol{\epsilon}_t &\stackrel{\text{def}}{=} \textstyle\sum_{k=1}^{78} \xi_{kt} \quad \text{with} \quad \xi_{kt} \sim N(0, 78^{-1}) \\ \boldsymbol{\Sigma}_t &\stackrel{\text{def}}{=} 0.05\bar{\boldsymbol{\Sigma}} + 0.85\boldsymbol{\Sigma}_{t-1} + 0.10\boldsymbol{r}_{t-1}\boldsymbol{r}_{t-1}^\top \end{aligned} \tag{29}$$

with $\bar{\boldsymbol{\Sigma}}$ being the unconditional covariance matrix (with unit diagonal).

For each of the parameter value combinations 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 various covariances, denoted generically by $\widehat{\mathbf{\Sigma}}_t^s(r)$, for method s.
3. Repeat the above 3 steps for a number of $R = 1000$ replications and then compute the mean distance of the covariance estimates vis-a-vis the true model covariance for all available values of τ, i.e.:

$$\bar{D}_R^s(h) \stackrel{\text{def}}{=} \frac{1}{R} \sum_{r=1}^{R} \|\mathbf{\Sigma}_{t+h} - \widehat{\mathbf{\Sigma}}_t^s(r)\|_F^2, \text{ for } h = 1, 2, ..\tau \qquad (30)$$

This last statistic is what we report, across different selections for B – the number and lengths of rolling windows.

- Next consider a simpler model, which has no realized covariance terms but 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. The form of the model now is:

$$
\begin{aligned}
\boldsymbol{r}_t &\stackrel{\text{def}}{=} \boldsymbol{\Sigma}_t^{1/2}\boldsymbol{\epsilon}_t \quad \text{with} \quad \boldsymbol{\epsilon}_t \sim N(0, \boldsymbol{I}_N) \\
\boldsymbol{\Sigma}_t^* &\stackrel{\text{def}}{=} \sum_{j=1}^{M} w_j^X(\alpha)\boldsymbol{r}_{t-j}\boldsymbol{r}_{t-j}^\top \\
\boldsymbol{\Sigma}_t &\stackrel{\text{def}}{=} 0.1\bar{\boldsymbol{\Sigma}} + 0.9\boldsymbol{\Sigma}_t^*
\end{aligned}
\tag{31}
$$

  where $w_j^X(\alpha)$ are the exponential weights of equation (12), with α 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 $N = 50$

Results from DGP#1

Covariance Averaging using Bivariate Scalar Diagonal VECH with $\rho = 0$

Estimator	B=(5, 20, 50)				B=(5, 20, 50, 100)				B=(50, 100, 200, 400)				B=(5, 20, 50, 100, 200, 400)			
	t	$t+1$	$t+5$	$t+10$	t	$t+1$	$t+5$	$t+10$	t	$t+1$	$t+5$	$t+10$	t	$t+1$	$t+5$	$t+10$
Realised 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
Optimised 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
Optimised 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
Optimised 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
Optimised 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
Optimised 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
Grid Search Weights	0.72	0.65	0.91	1.02	0.55	0.49	0.80	0.95	0.80	0.81	0.87	0.94	0.46	0.39	0.69	0.93

Results from DGP#2

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
Optimised 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
Optimised 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
Optimised 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
Optimised 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
Optimised 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
Grid Search Weights	0.93	0.83	0.84	1.03	0.21	0.23	1.22	0.94	1.02	1.24	1.90	1.48	0.89	1.14	1.69	1.16

Performance rankings based on DGP#1

Covariance Averaging Performance using Bivariate Scalar Diagonal VECH across all ρ

	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
Optimised 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
Optimised 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
Optimised 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
Optimised 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
Optimised 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
Grid Search 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	1.0	1.0	0.0	0.0

Performance rankings based on DGP#2

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
Optimised 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
Optimised 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
Optimised 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
Optimised 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
Optimised 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
Grid Search Weights	0.0	0.0	0.0	0.5	0.5	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0

Performance rankings overall

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
Optimised 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
Optimised 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
Optimised 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
Optimised 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
Optimised 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
Grid Search Weights	0.00	0.00	0.00	0.14	0.14	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.71	0.71	0.00	0.00

Would you use the method in real life?!

- While the forecasting performance of covariance averaging is very good this has potential application more in hedging strategies rather than portfolio problems. We thus examine the real-world performance of covariance averaging in the context of a GMV portfolio.
- Furthermore, we consider in some detail the combined effect of covariance estimation and portfolio rebalancing - the latter in various forms that we describe later.
- The standard GMV problem is given below:

$$\begin{aligned} \text{Minimize:} \quad & \boldsymbol{x}_t^\top \boldsymbol{\Sigma}_t \boldsymbol{x}_t \\ \text{subject to:} \quad & \boldsymbol{e}^\top \boldsymbol{x}_t = 1 \\ & 0 \le b_L \le x_{ti} \le b_U \le 1 \end{aligned} \tag{32}$$

where $\boldsymbol{x}_t$ changes at rebalancing/re-optimization times.

- In our analysis we consider 5 different cases for rebalancing: rebalancing at optimization time only (or no-rebalancing in the interim period); rebalancing based on a time threshold; rebalancing based on weight threshold; and a combination of 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.
- *Time rebalancing* refers to the weights' return into their initial values in a pre-specified number of days.
- *Threshold rebalancing* refers to the weights' return into their initial values in the case when a single's weight change is greater than a specified threshold (in percentage).

1. Start with an initial wealth of 1 million dollars and set the transaction cost of buying/selling one share at 0.005 dollars. At the beginning of the algorithm we decide of the re-optimisation period every E trading days.
2. Using a rolling window we set an in-sample of $N_{roll} = m_M + 1$ price observations.
3. We calculate the returns, the covariance estimators and the portfolio weights.
4. We find the number of shares that can be purchased and the positions we need to open/close. and the overall portfolio transaction costs are calculated.
5. Inbetween E portfolio re-optimisations the time and threshold rebalancing takes place.
6. Using the $N_{roll} + 1$ out-of-sample prices and the corresponding returns we calculate the overall portfolio value, its return and then wealth that will be carried over in the next round. item[6] Using the new wealth we start again from step 2 of the algorithm.

R1. *Time Rebalancing.*

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

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

R1c. The above procedure is carried-out $\lfloor E/R_{TR} \rfloor$ times.

R2. *Threshold Rebalancing.*

R2a. Set the threshold parameter, R_{ThR}. 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, as before, all transaction weights are calculated and current wealth is again computed.

R3. *Time and Threshold Rebalancing.* It combines the two approaches!

We use the following parametrizations for the optimization & rebalancing procedures described above:

- The bounds on the weights $\boldsymbol{x}_t$ are set to be in these 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. $E = 180$, $R_{TR} = 60$, $R_{ThR} = 5\%$
- The total, therefore, number of runs for each combination of the above is $36 \times N_B$, where N_B is the number of rolling window combinations that we have.

Summarizing the empirical output

- Since there is a lot of output being generated we present results on meta-data analysis based on aggregation across portfolio cases and methods. Here is how we do it.
- For each portfolio case, say i, we compute the performance measures. Let a representative such measure be called P_{ijk} when is based on rebalancing method j and estimation approach k.
- We compute three types of averaged statistics:
  1. The first statistic pools the data across all i, then evaluates across all j and aggregates across all k (ALL);
  2. The second statistic takes the data for each i, then evaluates across all j and then aggregates across all k (WITHIN);
  3. The third statistic pools the data across j, then evaluates across k and then aggregates across i (BETWEEN).
- We employ a GMM-approach to estimate mean differences (of a method against the benchmarks) for the above measures and type, and *the percentage of times that this difference is in favour of the new methods is counted and tracked*.

Data

- Daily data from S&P500 stocks from to 1990 to October 2012.
- Three groups of 8, 20 and 40 stocks - the latter based on (most recent) largest capitalization.
- Need to try other meaningful group combinations!

Table 1A. Success rates across all runs - SP08				
	Sample-Full	**Sample-Half**	**LW-Full**	**LW-Half**
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%

Table 1B. Success rates across all runs - SP20				
	Sample-Full	**Sample-Half**	**LW-Full**	**LW-Half**
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%

Table 1C. Success rates across all runs - SP40				
	Sample-Full	**Sample-Half**	**LW-Full**	**LW-Half**
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 2A. Success rates within methods - SP08	Sample-Full	Sample-Half	LW-Full	LW-Half
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%

Table 2B. Success rates within methods - SP20	Sample-Full	Sample-Half	LW-Full	LW-Half
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%

Table 2C. Success rates within methods - SP40	Sample-Full	Sample-Half	LW-Full	LW-Half
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 3A. Cumulative return differences in between methods - SP40

No Rebalance	**Sample-Full**	**Sample-Half**	**LW-Full**	**LW-Half**
Proportion	92.31%	69.23%	92.31%	76.92%
Mean Difference	-7.73%	-1.16%	-7.31%	-2.35%
Time only	**Sample-Full**	**Sample-Half**	**LW-Full**	**LW-Half**
Proportion	92.31%	76.92%	84.62%	76.92%
Mean Difference	-6.05%	-1.41%	-4.86%	-1.77%
Thresh. #1 only	**Sample-Full**	**Sample-Half**	**LW-Full**	**LW-Half**
Proportion	15.38%	38.46%	38.46%	69.23%
Mean Difference	7.35%	1.26%	1.67%	-3.61%
Thresh. #2 only	**Sample-Full**	**Sample-Half**	**LW-Full**	**LW-Half**
Proportion	84.62%	84.62%	84.62%	84.62%
Mean Difference	-7.75%	-3.77%	-8.43%	-7.40%
T & T #1	**Sample-Full**	**Sample-Half**	**LW-Full**	**LW-Half**
Proportion	92.31%	76.92%	92.31%	76.92%
Mean Difference	-8.25%	-3.84%	-8.83%	-3.98%
T &T #2	**Sample-Full**	**Sample-Half**	**LW-Full**	**LW-Half**
Proportion	84.62%	84.62%	84.62%	84.62%
Mean Difference	-2.76%	-4.83%	-4.19%	-6.39%

Table 3B. Drawdown differences in between methods - SP40

No Rebalance	**Sample-Full**	**Sample-Half**	**LW-Full**	**LW-Half**
Proportion	84.62%	53.85%	76.92%	53.85%
Mean Difference	1.18%	-0.02%	1.09%	0.02%
Time only	**Sample-Full**	**Sample-Half**	**LW-Full**	**LW-Half**
Proportion	0.00%	46.15%	0.00%	53.85%
Mean Difference	1.34%	0.24%	1.09%	-0.06%
Thresh. #1 only	**Sample-Full**	**Sample-Half**	**LW-Full**	**LW-Half**
Proportion	7.69%	38.46%	7.69%	38.46%
Mean Difference	1.55%	-0.02%	1.60%	0.45%
Thresh. #2 only	**Sample-Full**	**Sample-Half**	**LW-Full**	**LW-Half**
Proportion	23.08%	38.46%	15.38%	38.46%
Mean Difference	1.50%	0.30%	1.60%	0.09%
T & T #1	**Sample-Full**	**Sample-Half**	**LW-Full**	**LW-Half**
Proportion	7.69%	30.77%	7.69%	30.77%
Mean Difference	2.16%	0.75%	2.21%	0.71%
T &T #2	**Sample-Full**	**Sample-Half**	**LW-Full**	**LW-Half**
Proportion	23.08%	38.46%	15.38%	46.15%
Mean Difference	1.46%	0.23%	1.59%	-0.04%

Table 3C. Sharpe ratio differences in between methods - SP40

No Rebalance	**Sample-Full**	**Sample-Half**	**LW-Full**	**LW-Half**
Proportion	92.31%	23.08%	92.31%	30.77%
Mean Difference	-0.01	0.00	-0.01	0.00
Time only	**Sample-Full**	**Sample-Half**	**LW-Full**	**LW-Half**
Proportion	76.92%	69.23%	76.92%	69.23%
Mean Difference	-0.01	0.00	-0.01	-0.01
Thresh. #1 only	**Sample-Full**	**Sample-Half**	**LW-Full**	**LW-Half**
Proportion	23.08%	23.08%	38.46%	53.85%
Mean Difference	0.00	0.00	0.00	0.00
Thresh. #2 only	**Sample-Full**	**Sample-Half**	**LW-Full**	**LW-Half**
Proportion	76.92%	69.23%	76.92%	76.92%
Mean Difference	-0.01	0.00	-0.01	0.00
T & T #1	**Sample-Full**	**Sample-Half**	**LW-Full**	**LW-Half**
Proportion	92.31%	53.85%	92.31%	53.85%
Mean Difference	-0.01	0.00	-0.01	0.00
T &T #2	**Sample-Full**	**Sample-Half**	**LW-Full**	**LW-Half**
Proportion	76.92%	61.54%	76.92%	76.92%

All in all...

- Introduce covariance averaging based on rolling window averages of the sample covariance.
- Provide a host of heuristic and optimizing procedures to select the weights that should be attached to the sample covariances.
- Show that the proposed approach has merit in fitting and forecasting the unknown covariance and in consistently improving portfolio allocation results, especially in large(r) problems involving many assets.
- Open questions that are not addressed: *be more specific* as to which methods of covariance averaging tops all others; how does covariance averaging performs in forecasting with real series; can covariance averaging be fruitfully used in other inference problems; can we have a complete theory based on underlying properties of the time series;