Spreadsheets in Education (eJSiE)

Volume 10 | Issue 1 Article 3

July 2017

Shrinkage of the Sample Correlation Matrix of Returns Towards a Constant Correlation Target: A Pedagogic Illustration Based on Dow Jones Stock Returns

Clarence C. Y. Kwan
McMaster University, kwanc@mcmaster.ca

Shrinkage of the Sample Correlation Matrix of Returns Towards a Constant Correlation Target: A Pedagogic Illustration Based on Dow Jones Stock Returns

Abstract
This paper extends the introduction to shrinkage estimation in a recent paper from the same journal. The extension, which is in a portfolio investment context, is on shrinkage of the sample correlation matrix of returns towards a constant correlation target. Here, shrinkage estimation is about finding a weighted average of the sample correlation matrix and the target matrix, for a balance between reducing overall forecast errors and maintaining some existing idiosyncrasies in the individual correlations. Excel plays an important pedagogic role here. Besides illustrating the computations involved, the use of Excel also enables students to gain valuable hands-on experience in shrinkage estimation, by working with the Dow Jones stock returns in an Excel file accompanying this paper.

Keywords
sample correlation matrix, shrinkage estimation, constant correlation target

Cover Page Footnote
The author wishes to thank two anonymous reviewers for valuable comments and suggestions.

Shrinkage of the Sample Correlation Matrix of Returns Towards a Constant Correlation Target: A Pedagogic Illustration Based on Dow Jones Stock Returns

1 Introduction

Mean-variance portfolio theory is part of the core curriculum in financial education. Students learn from the theory that combinations of risky securities, such as common stocks, can improve an investment's risk-return trade-off. Portfolio selection models that are formulated to capture relevant institutional features and investment constraints adequately are useful for assisting practical portfolio decisions. The implementation of a portfolio selection model typically requires, as input parameters, a vector of expected returns and a covariance matrix of returns, which is symmetric. The covariance of matrix itself can be deduced from a vector of standard deviations of returns and a correlation matrix of returns, which is also symmetric, and vice versa.

The true values of the individual input parameters being unknown, their estimated values will have to be used instead. As the portfolio selection results are input sensitive, it is important that reliable estimates are used. A popular estimation method is by using a sample of past monthly or weekly return observations, under the assumption that each observation is a random draw from a stationary joint probability distribution. The choice of the length of a sample period is often a trade-off between satisfying the stationarity assumption (for using a short sample) and reducing the estimation errors (for using a long sample).

In practice, the use of past return observations to form expectations can also serve as a starting point in generating a set of acceptable input parameters. This is because insights of financial analysts are often deemed necessary for revising the estimated values of the input parameters. For an n-security case, the individual input parameters to be estimated include n expected returns, n standard deviations of returns, and $n(n-1)/2$ correlations of returns. These correlations are all off-diagonal elements in the upper or lower triangle of the $n \times n$ correlation matrix of returns. A relevant question, therefore, is whether it is practical for financial analysts to revise the entire set of estimated input parameters or only part of it.

Take, for example, the case where $n = 50$, which is not a large number for the size of a professionally managed stock portfolio. The insights of a group of analysts who track the financial data of the 50 companies considered, collectively, can lead to some improvements of the 50-element vector of sample average returns — as proxies for the corresponding expected returns — and the 50-element vector of sample standard deviations of returns. However, to revise the $1,225$ individual correlation coefficients, even partially, based on analysts' insights is a highly burdensome task. For any larger n, the enormity of the task will be even more overwhelming. The practical concerns that the correlations involved are too numerous for financial analysts to revise has led Elton and Gruber (1973) and Elton, Gruber, and Urich (1978) to comment that it is highly unlikely that estimations of the correlation matrix can be based on approaches other than those relying on past return observations.

The above authors and, subsequently, Chan, Karceski, and Lakonishok (1999), have provided empirical supports for using the average of sample correlations as a predictor of the individual correlations. Although such averaging is effective in reducing overall forecast errors, any idiosyncrasies in the individual correlations, inevitably, will be lost. For example, if some securities among the many securities considered are from the same industry, their sample correlations of returns are usually higher than the overall average. Their sample correlations beyond the overall average can be viewed as part of the idiosyncrasies in the set of securities considered, and such idiosyncrasies will be lost due to averaging. Shrinkage estimation of covariance and correlation matrices — a statistical approach introduced to the finance profession by Ledoit and Wolf (2003, 2004) — is a good remedy, which strikes a balance between reducing overall forecast errors and retaining some existing idiosyncrasies in the individual correlations.

Shrinkage estimation in the context here is about achieving the best weighted average of a sample covariance or correlation matrix and an approximate but structured matrix, for a given decision criterion. The latter matrix is called the shrinkage target, for which a symmetric matrix based on a constant correlation structure is a practical example. The weight that is assigned to the shrinkage target is known as the shrinkage intensity. However, unless some asymptotic properties are assumed for analytical convenience, the determination of the shrinkage intensity for a constant correlation target is still a tedious task. [See Kwan (2008) for the corresponding analytical details when no asymptotic properties are assumed.]

Using Excel for a numerical illustration, Kwan (2011) has provided an introduction to the shrinkage approach. For simplicity, each shrinkage target is characterized by zero correlations throughout. The current paper extends this introduction by shrinking the sample correlation matrix towards a constant correlation target instead. Here, the constant correlation is provided by the average of the individual sample correlations. Such an extension is relevant, both practically and pedagogically, for the following reason:

As the use of a zero correlation target always leads to attenuations in the individual correlations, their post-shrinkage average will always be lower than the original sample average. Besides the lack of empirical supports for a zero correlation target, it is also difficult to explain, from a pedagogic perspective, why attenuating the sample correlations can be expected to provide better estimates of the true correlations. In contrast, the use of a constant correlation target, which does not cause any change to the original sample average, is free from such concerns. With the post-shrinkage correlations being closer to the original sample average, a balance between two competing interests can be achieved. As indicated earlier, the balance is between reducing overall forecast errors and retaining some existing idiosyncrasies in the individual correlations.

This paper presents a pedagogic version of the above extension. A crucial simplifying assumption here is that any revisions of the sample standard deviations of returns for potential improvements are based on the judgement of the financial analysts involved. Under such an assumption, we can focus on shrinkage estimation of the individual correlations of returns, without the encumbrance of the usual statistical issues pertaining to their estimation via the corresponding sample variances and covariances as encountered previously in Kwan (2008).

As in Kwan (2011), Excel also plays an important pedagogic role in the current paper. For students who are unfamiliar with computer programming in Visual Basic for Applications (VBA), this paper illustrates, in a small scale example, an Excel-based computational approach that does not require its use. For versatility in accommodating various numbers of securities and return observations, this paper generates a user-defined Excel function for computing the shrinkage intensity, with the corresponding code in VBA. Access to the function can also be achieved via a keyboard shortcut, which will prompt the user for specific inputs as required for its individual arguments. Either way, to access the function requires that the corresponding Macro feature be enabled.

From a computational standpoint, shrinkage estimation with a constant correlation target accommodates a zero correlation target as a special case. Thus, the above-mentioned Excel function is coded in such a way that this special case is also covered. The choice between the two shrinkage targets is via one of the function's arguments. Further, the current paper also addresses the statistical issue of bias, as considered in Kwan (2008), for using sample correlations to estimate the corresponding true but unknown correlations. The same Excel function allows such bias to be either corrected or ignored at the discretion of the user. The choice is also via one of the function's arguments.

Given the above choices in the shrinkage target and in bias correction, there are four sets of shrinkage results to compare, for a given set of securities and the corresponding return observations. To facilitate some meaningful comparisons, this paper uses actual stock return data, instead of artificial data as in Kwan (2011). The full data set, which covers seven years of monthly return observations of the current 30 Dow Jones stocks, from January 2010 to December 2016, is contained in one of the two Excel files accompanying this paper.[1] To accommodate more readers, both Excel files have been saved in the 1997-2003 version, which has .xls as the extension of each filename.

The example in the remaining Excel file, which is for illustrating the computational details with or without the use of VBA, is based only on five of the Dow Jones stocks and six monthly return observations. Obviously, the number of observations involved is too low for any meaningful comparisons of the four sets of shrinkage results. Instead, the comparisons are based on the full set of Dow Jones stocks, with return observations ranging from 36 months to 84 months, all ending at December 2016. The first-mentioned Excel file, which contains the full data set, also includes a VBA-only example based on the 30 Dow Jones stocks and a 36-month sample period. Omitted from the file are other VBA-only examples with longer sample periods, which are similar from a computational standpoint.

The sample estimates as reported in this paper will show, in general, how the returns of major U.S. stocks are correlated and how such estimates are affected by the choice of the sample period. Students will see how shrinkage estimation affects the sample correlation matrix and whether improving the shrinkage target, as is done in this paper, makes any difference in the shrinkage results. Students will also see whether ignoring the statistical issue of bias is consequential. Further, given the available return observations in the first-mentioned Excel file, it will be a

[1] The underlying daily closing stock prices and any dividend data were collected from the Yahoo! Finance website (https://ca.finance.yahoo.com/), by first entering each company's ticker symbol for the search. Such data were available under the "Historical Data" tab.

good exercise for students to repeat the same computations for different sample periods. The exercise will help them gain some valuable hands-on experience with shrinkage estimation.

The data set in the first-mentioned Excel file accompanying this paper can easily be augmented to include some other stocks and/or to lengthen the sample period. Thus, students can also gain some hands-on experience in data collection for estimating the covariance matrix. However, as the scope of this paper is confined to illustrating the statistical approach of shrinkage estimation, whether the resulting correlation matrix will lead to meaningful improvements in the portfolio selection results still depends on the quality of other input parameters, which include a vector of expected returns and a vector of standard deviations or variances of returns.

The remainder of this paper is organized as follows: The analytical materials leading to the determination of the shrinkage intensity for a constant correlation target and a zero correlation target are presented in Sections 2 and 3, respectively. Section 4 first illustrates with a small scale case the computations involved, including the generation of a user-defined Excel function for computing the shrinkage intensity, as well as an approach not requiring the use of VBA. The codes in VBA and a subroutine for use in the illustration can be found in either Excel file accompanying this paper. Section 4 then summarizes and discusses the various shrinkage results, based on the full set of Dow Jones stocks and different sample periods. Section 5 provides some concluding remarks, as well as some suggestions for instructors.

2 Shrinkage Towards a Constant Correlation Target

This section, which is confined to shrinkage of the sample correlation matrix of returns towards a constant correlation target, covers the corresponding analytical and statistical tasks in five subsections. In the first subsection, we transform linearly the return observations to make the resulting sample covariance and correlation matrices indistinguishable from each other. Such transformations, which do not affect the sample correlation matrix, are for analytical convenience afterwards. In the second subsection, following Ledoit and Wolf (2003, 2004), we derive an explicit expression of the shrinkage intensity based on minimization of the expected value of a quadratic loss function. The third subsection shows how the shrinkage intensity can be estimated by using the transformed observations. In the fourth subsection, we verify that the estimated shrinkage intensity is always within its intended range. Finally, in the fifth subsection, we correct the statistical bias for using sample correlations to estimate the corresponding true but unknown correlations, as mentioned briefly in the introductory section.

2.1 Positive Linear Transformations of Return Observations

For an n-security case, let $\boldsymbol{R}_1, \boldsymbol{R}_2, \ldots, \boldsymbol{R}_n$ be the random returns of the individual securities. Let also $R_{i1}, R_{i2}, \ldots, R_{iT}$ be the returns observed at times $1, 2, \ldots, T$, respectively, for $i = 1, 2, \ldots, n$. For each security i, the sample mean return and the sample variance of returns are

$$\overline{\boldsymbol{R}}_i = \frac{1}{T} \sum\nolimits_{t=1}^{T} R_{it} \tag{1}$$

and

$$s_i^2 = \frac{1}{T-1} \sum\nolimits_{t=1}^{T} (R_{it} - \overline{\boldsymbol{R}}_i)^2, \tag{2}$$

respectively, where s_i is the sample standard deviation of returns. The sample covariance of returns between securities i and j, for $i, j = 1, 2, \ldots, n$, is

$$s_{ij} = \frac{1}{T-1} \sum\nolimits_{t=1}^{T} (R_{it} - \overline{\boldsymbol{R}}_i)(R_{jt} - \overline{\boldsymbol{R}}_j). \tag{3}$$

It is implicit that

$$s_i^2 = s_{ii}. \tag{4}$$

The sample correlation of returns between securities i and j, for $i, j = 1, 2, \ldots, n$, is

$$r_{ij} = \frac{s_{ij}}{s_i s_j}. \tag{5}$$

Let us perform a positive linear transformation on each of the n random variables, by defining

$$\boldsymbol{Z}_i = \frac{\boldsymbol{R}_i - \overline{\boldsymbol{R}}_i}{s_i}, \text{ for } i = 1, 2, \ldots, n. \tag{6}$$

We now have $Z_{i1}, Z_{i2}, \ldots, Z_{iT}$, instead of $R_{i1}, R_{i2}, \ldots, R_{iT}$, as the corresponding observed values at times $1, 2, \ldots, T$. With

$$Z_{it} = \frac{R_{it} - \overline{\boldsymbol{R}}_i}{s_i}, \text{ for } i = 1, 2, \ldots, n \text{ and } t = 1, 2, \ldots, T, \tag{7}$$

the sample mean and the sample variance of each $\boldsymbol{Z}_i$ are

$$\begin{aligned} \overline{\boldsymbol{Z}}_i &= \frac{1}{T} \sum\nolimits_{t=1}^{T} Z_{it} = \frac{1}{T} \sum\nolimits_{t=1}^{T} \frac{R_{it} - \overline{\boldsymbol{R}}_i}{s_i} \\ &= \frac{1}{s_i} \left[\left(\frac{1}{T} \sum\nolimits_{t=1}^{T} R_{it} \right) - \overline{\boldsymbol{R}}_i \right] = 0 \end{aligned} \tag{8}$$

and

$$\widehat{Var}(\boldsymbol{Z}_i) = \frac{1}{T-1}\sum_{t=1}^{T}(Z_{it}-\overline{\boldsymbol{Z}}_i)^2 = \frac{1}{s_i^2}\left[\frac{1}{T-1}\sum_{t=1}^{T}(R_{it}-\overline{\boldsymbol{R}}_i)^2\right] = 1, \quad (9)$$

respectively. The sample covariance between $\boldsymbol{Z}_i$ and $\boldsymbol{Z}_j$ is

$$\widehat{Cov}(\boldsymbol{Z}_i,\boldsymbol{Z}_j) = \frac{1}{T-1}\sum_{t=1}^{T}(Z_{it}-\overline{\boldsymbol{Z}}_i)(Z_{jt}-\overline{\boldsymbol{Z}}_j) = \frac{1}{s_i s_j}\left[\frac{1}{T-1}\sum_{t=1}^{T}(R_{it}-\overline{\boldsymbol{R}}_i)(R_{jt}-\overline{\boldsymbol{R}}_j)\right] = \frac{s_{ij}}{s_i s_j} = r_{ij}. \quad (10)$$

Thus, such transformations do not change the sample correlation matrix of the n variables; indeed, the sample correlation between $\boldsymbol{R}_i$ and $\boldsymbol{R}_j$, the sample correlation between $\boldsymbol{Z}_i$ and $\boldsymbol{Z}_j$, and the sample covariance between $\boldsymbol{Z}_i$ and $\boldsymbol{Z}_j$, for $i, j = 1, 2, \ldots, n$ and $i \neq j$, are indistinguishable.

2.2 Derivation of the Shrinkage Intensity

As both $\boldsymbol{Z}_i$ and $\boldsymbol{Z}_j$ have zero sample means, we can write

$$r_{ij} = \frac{1}{T-1}\sum_{t=1}^{T} Z_{it}Z_{jt}. \quad (11)$$

There are $n(n-1)/2$ off-diagonal elements in the upper triangle of the sample correlation matrix. Thus, the average of the sample correlations is

$$\overline{r} = \frac{2}{n(n-1)}\sum_{i=1}^{n-1}\sum_{j=i+1}^{n} r_{ij}. \quad (12)$$

The true correlation between $\boldsymbol{R}_i$ and $\boldsymbol{R}_j$, for $i, j = 1, 2, \ldots, n$ and $i \neq j$, is unknown. Let us denote it as ρ_{ij}.

With λ being the shrinkage intensity, the weight assigned to $\overline{r}$, the departure of the resulting correlation between $\boldsymbol{R}_i$ and $\boldsymbol{R}_j$ from the corresponding true correlation is $\lambda\overline{r} + (1-\lambda)r_{ij} - \rho_{ij}$. Following Ledoit and Wolf (2004), we look for the value of λ that minimizes the expected value of a quadratic loss function,

$$L(\lambda) = E\left\{\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}[\lambda\overline{r} + (1-\lambda)r_{ij} - \rho_{ij}]^2\right\}, \quad (13)$$

where $E(\cdot)$ is the expected value operator. Implicitly, the individual sample correlations and their average are treated as random variables. Here, the double summation is over the $n(n-1)/2$ off-diagonal elements in the upper triangle of the correlation matrix to be estimated. As $E(\cdot)$ is a linear operator, equation (13) is equivalent to

$$L(\lambda) = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} E\left\{[\lambda\overline{r} + (1-\lambda)r_{ij} - \rho_{ij}]^2\right\}. \tag{14}$$

Noting that $E(x^2) = Var(x) + [E(x)]^2$, for any random variable x, we can write

$$L(\lambda) = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \left\{Var[\lambda\overline{r} + (1-\lambda)r_{ij} - \rho_{ij}] + \left\{E[\lambda\overline{r} + (1-\lambda)r_{ij} - \rho_{ij}]\right\}^2\right\}, \tag{15}$$

which, upon simplifications, leads to

$$\begin{aligned} L(\lambda) &= \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \left\{\lambda^2 Var(\overline{r}) + (1-\lambda)^2 Var(r_{ij}) + 2\lambda(1-\lambda)Cov(\overline{r}, r_{ij})\right. \\ &\qquad \left. + [\lambda E(\overline{r}) + (1-\lambda)E(r_{ij}) - \rho_{ij}]^2\right\}. \end{aligned} \tag{16}$$

The optimal shrinkage intensity can be deduced from

$$\begin{aligned} \frac{dL(\lambda)}{d\lambda} &= \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \{2\lambda Var(\overline{r}) - 2(1-\lambda)Var(r_{ij}) + 2(1-2\lambda)Cov(\overline{r}, r_{ij}) \\ &\qquad + 2E(\overline{r} - r_{ij})[\lambda E(\overline{r} - r_{ij}) + E(r_{ij}) - \rho_{ij}]\} \\ &= 0. \end{aligned} \tag{17}$$

The result is

$$\lambda = \frac{\sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \left\{Var(r_{ij}) - Cov(\overline{r}, r_{ij}) - E(\overline{r} - r_{ij})[E(r_{ij}) - \rho_{ij}]\right\}}{\sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \{Var(\overline{r}) + Var(r_{ij}) - 2Cov(\overline{r}, r_{ij}) + [E(\overline{r} - r_{ij})]^2\}}. \tag{18}$$

In view of equation (12), we can write

$$\begin{aligned} \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} Var(\overline{r}) &= \frac{n(n-1)}{2} Var(\overline{r}) \\ &= \frac{n(n-1)}{2} \left[\frac{2}{n(n-1)}\right]^2 Var\left(\sum_{i=1}^{n-1} \sum_{j=i+1}^{n} r_{ij}\right) \qquad (19) \\ &= \frac{2}{n(n-1)} \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \sum_{k=1}^{n-1} \sum_{\ell=k+1}^{n} Cov(r_{ij}, r_{k\ell}). \qquad (20) \end{aligned}$$

Further, as

$$\begin{aligned} \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} Cov(\overline{r}, r_{ij}) &= \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} Cov\left\{\left[\frac{2}{n(n-1)} \sum_{k=1}^{n-1} \sum_{\ell=k+1}^{n} r_{k\ell}\right], r_{ij}\right\} \\ &= \frac{2}{n(n-1)} \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \sum_{k=1}^{n-1} \sum_{\ell=k+1}^{n} Cov(r_{ij}, r_{k\ell}), \qquad (21) \end{aligned}$$

we confirm that

$$\sum_{i=1}^{n-1}\sum_{j=i+1}^{n} Var(\bar{r}) = \sum_{i=1}^{n-1}\sum_{j=i+1}^{n} Cov(\bar{r}, r_{ij}). \tag{22}$$

Accordingly, equation (18) reduces to

$$\lambda = \frac{\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}\{\alpha_{ij} - E(\bar{r} - r_{ij})[E(r_{ij}) - \rho_{ij}]\}}{\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}\{\alpha_{ij} + [E(\bar{r} - r_{ij})]^2\}}, \tag{23}$$

where

$$\alpha_{ij} = Var(r_{ij}) - \frac{2}{n(n-1)}\sum_{k=1}^{n-1}\sum_{\ell=k+1}^{n} Cov(r_{ij}, r_{k\ell}). \tag{24}$$

2.3 Estimations of Various Expected Values, Variances, and Covariances

To compute the shrinkage intensity from equations (23) and (24), estimates of various expected values, variances, and covariances, as well as approximations of the true but unknown correlations, are required. Estimations of the variances and covariances involved — which draw on Schäfer and Strimmer (2005) for finite samples — can be explained by using familiar statistical concepts. Thus, they are presented below first. However, to estimate the true but unknown correlations without statistical bias is not as simple. To avoid unnecessary digressions, the corresponding material where the issue of bias is ignored is covered at the end of this subsection, along with estimations of the remaining terms in equation (23).

2.3.1 Variances

To estimate $Var(r_{ij})$, for $i = 1, 2, \ldots, n-1$ and $j = i+1, i+2, \ldots, n$, let $\mathbf{W}_{ij} = \boldsymbol{Z}_i\boldsymbol{Z}_j$. Let also $W_{ij1}, W_{ij2}, \ldots, W_{ijT}$ be the T observations of $\mathbf{W}_{ij}$. The sample mean of $\mathbf{W}_{ij}$ in terms of the T observations of $\boldsymbol{Z}_i$ and $\boldsymbol{Z}_j$is

$$\overline{\mathbf{W}}_{ij} = \frac{1}{T}\sum_{t=1}^{T} Z_{it}Z_{jt}. \tag{25}$$

Combining equations (11) and (25) leads to

$$r_{ij} = \frac{T}{T-1}\overline{\mathbf{W}}_{ij}. \tag{26}$$

Using $\widehat{Var}(\cdot)$ and $\widehat{Cov}(\cdot,\cdot)$ for sample variances and covariances of the random variables involved, respectively, we can write

$$\widehat{Var}(r_{ij}) = \frac{T^2}{(T-1)^2}\widehat{Var}(\overline{\mathbf{W}}_{ij}) \tag{27}$$

and

$$\begin{aligned}\widehat{Var}(\overline{\boldsymbol{W}}_{ij}) &= \widehat{Var}\left(\frac{1}{T}\sum\nolimits_{t=1}^{T} W_{ijt}\right) = \frac{1}{T^2}\widehat{Cov}\left(\sum\nolimits_{t=1}^{T} W_{ijt}, \sum\nolimits_{u=1}^{T} W_{iju}\right) \\ &= \frac{1}{T^2}\sum\nolimits_{t=1}^{T}\sum\nolimits_{u=1}^{T}\widehat{Cov}\left(W_{ijt}, W_{iju}\right). \end{aligned} \tag{28}$$

With $W_{ij1}, W_{ij2}, \ldots, W_{ijT}$ being T random draws from the probability distribution of $\boldsymbol{W}_{ij}$, it is implicit that

$$\widehat{Cov}\left(W_{ijt}, W_{ijt}\right) = \widehat{Var}(\boldsymbol{W}_{ij}), \text{ for } t = 1, 2, \ldots, T, \tag{29}$$

and, in the absence of serial correlations,

$$\widehat{Cov}\left(W_{ijt}, W_{iju}\right) = 0, \text{ for } t \neq u. \tag{30}$$

It follows directly from

$$\widehat{Var}(\overline{\boldsymbol{W}}_{ij}) = \frac{1}{T^2}\left[T\,\widehat{Var}(\boldsymbol{W}_{ij})\right] = \frac{1}{T}\widehat{Var}(\boldsymbol{W}_{ij}) \tag{31}$$

that

$$\widehat{Var}(r_{ij}) = \frac{T}{(T-1)^2}\widehat{Var}(\boldsymbol{W}_{ij}). \tag{32}$$

As

$$\widehat{Var}(\boldsymbol{W}_{ij}) = \frac{1}{T-1}\sum\nolimits_{t=1}^{T}(W_{ijt} - \overline{\boldsymbol{W}}_{ij})^2, \tag{33}$$

we can estimate each $Var(r_{ij})$ in equation (24) with

$$\widehat{Var}(r_{ij}) = \frac{T}{(T-1)^3}\sum\nolimits_{t=1}^{T}(W_{ijt} - \overline{\boldsymbol{W}}_{ij})^2. \tag{34}$$

2.3.2 Covariances

To estimate each $Cov(r_{ij}, r_{k\ell})$ in equation (24), we also let $\boldsymbol{W}_{k\ell} = \boldsymbol{Z}_k\boldsymbol{Z}_\ell$, for $k = 1, 2, \ldots, n-1$ and $\ell = k+1, k+2, \ldots, n$. With $W_{k\ell 1}, W_{k\ell 2}, \ldots, W_{k\ell T}$ being the T observations of $\boldsymbol{W}_{k\ell}$, the corresponding sample mean is

$$\overline{\boldsymbol{W}}_{k\ell} = \frac{1}{T}\sum\nolimits_{t=1}^{T} Z_{kt}Z_{\ell t}. \tag{35}$$

It follows from equation (26) that

$$\widehat{Cov}(r_{ij}, r_{k\ell}) = \frac{T^2}{(T-1)^2}\widehat{Cov}(\overline{\boldsymbol{W}}_{ij}, \overline{\boldsymbol{W}}_{k\ell}), \tag{36}$$

where

$$\widehat{Cov}(\overline{\boldsymbol{W}}_{ij}, \overline{\boldsymbol{W}}_{k\ell}) = \widehat{Cov}\left(\frac{1}{T}\sum\nolimits_{t=1}^{T} W_{ijt}, \frac{1}{T}\sum\nolimits_{u=1}^{T} W_{k\ell u}\right)$$
$$= \frac{1}{T^2}\sum\nolimits_{t=1}^{T}\sum\nolimits_{u=1}^{T}\widehat{Cov}(W_{ijt}, W_{k\ell u}). \tag{37}$$

As $W_{ij1}, W_{ij2}, \ldots, W_{ijT}$ and $W_{k\ell 1}, W_{k\ell 2}, \ldots, W_{k\ell T}$ are random draws from the probability distributions of $\boldsymbol{W}_{ij}$ and $\boldsymbol{W}_{k\ell}$, respectively, it is implicit that

$$\widehat{Cov}\left(W_{ijt}, W_{k\ell t}\right) = \widehat{Cov}(\boldsymbol{W}_{ij}, \boldsymbol{W}_{k\ell}), \text{ for } t = 1, 2, \ldots, T, \tag{38}$$

and, in the absence of serial correlations,

$$\widehat{Cov}\left(W_{ijt}, W_{k\ell u}\right) = 0, \text{ for } t \neq u. \tag{39}$$

Thus, equation (36) can be written as

$$\widehat{Cov}(r_{ij}, r_{k\ell}) = \frac{T^2}{(T-1)^2}\left[\frac{T}{T^2}\widehat{Cov}(\boldsymbol{W}_{ij}, \boldsymbol{W}_{k\ell})\right] = \frac{T}{(T-1)^2}\widehat{Cov}(\boldsymbol{W}_{ij}, \boldsymbol{W}_{k\ell}). \tag{40}$$

Once $\widehat{Cov}(\boldsymbol{W}_{ij}, \boldsymbol{W}_{k\ell})$ is expressed explicitly in terms of the observations, equation (40) becomes

$$\widehat{Cov}(r_{ij}, r_{k\ell}) = \frac{T}{(T-1)^3}\sum\nolimits_{t=1}^{T}(W_{ijt} - \overline{\boldsymbol{W}}_{ij})(W_{k\ell t} - \overline{\boldsymbol{W}}_{k\ell}). \tag{41}$$

2.3.3 Expected Values

To get equations (23) and (24) ready for use, we also need various sample means for the corresponding expected values, as well as an estimate of each true but unknown correlation ρ_{ij}, for $i = 1, 2, \ldots, n-1$ and $j = i+1, i+2, \ldots, n$. Let us start with the expected value $E(r_{ij})$. Given equation (26), where $\overline{\boldsymbol{W}}_{ij}$ is already a sample mean, the sample mean of r_{ij} is

$$\overline{r}_{ij} = \frac{T}{T-1}\overline{\boldsymbol{W}}_{ij}, \tag{42}$$

As $\overline{r}_{ij} = r_{ij}$, we can directly substitute r_{ij} for $E(r_{ij})$ when using equations (23) and (24) to compute the optimal shrinkage intensity. With $\overline{r}$ being the average of the $n(n-1)/2$ individual sample correlations, we can also directly substitute $\overline{r}$ for $E(\overline{r})$ there as well.

To estimate each unknown ρ_{ij}, for $i \neq j$, is not a simple task. This is because, for a finite sample, $E(r_{ij})$ is a biased estimator of ρ_{ij}. The nature of the bias is that the sample correlation tends to understate the true correlation in magnitude. The bias becomes smaller, if a larger sample of return observations is used for the estimation.[2] For analytical convenience, we use $E(r_{ij})$ as an estimator of ρ_{ij} for now. Bias correction will be considered in the final subsection below.

With $E(r_{ij}) - \rho_{ij}$ being zero, equation (23) reduces to

$$\lambda = \frac{\sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \alpha_{ij}}{\sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \{\alpha_{ij} + [E(\overline{r} - r_{ij})]^2\}}. \tag{43}$$

As we can use r_{ij} and $\overline{r}$ for $E(r_{ij})$ and $E(\overline{r})$, respectively, the estimated λ is given by

$$\widehat{\lambda} = \frac{\sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \widehat{\alpha}_{ij}}{\sum_{i=1}^{n-1} \sum_{j=i+1}^{n} [\widehat{\alpha}_{ij} + (\overline{r} - r_{ij})^2]}, \tag{44}$$

where

$$\widehat{\alpha}_{ij} = \widehat{Var}(r_{ij}) - \frac{2}{n(n-1)} \sum_{k=1}^{n-1} \sum_{\ell=k+1}^{n} \widehat{Cov}(r_{ij}, r_{k\ell}). \tag{45}$$

Here, r_{ij}, $\overline{r}$, $\widehat{Var}(r_{ij})$, and $\widehat{Cov}(r_{ij}, r_{k\ell})$ can be computed by using equations (11), (12), (34), respectively.

2.4 Acceptability of the Estimated Shrinkage Intensity

For the shrinkage results based on equations (44) and (45) to be acceptable, the condition of $0 < \widehat{\lambda} < 1$ must be satisfied. To satisfy such a condition, in turn, requires that $\sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \widehat{\alpha}_{ij}$ be positive. The sign of this double summation is the same as that of

$$\begin{aligned} & \frac{n(n-1)}{2} \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \widehat{\alpha}_{ij} \\ = \; & \frac{n(n-1)}{2} \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \widehat{Var}(r_{ij}) - \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \sum_{k=1}^{n-1} \sum_{\ell=k+1}^{n} \widehat{Cov}(r_{ij}, r_{k\ell}). \end{aligned} \tag{46}$$

For ease of exposition below, let

$$A = \frac{n(n-1)}{2} \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \widehat{Var}(r_{ij}) \tag{47}$$

and

$$B = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \sum_{k=1}^{n-1} \sum_{\ell=k+1}^{n} \widehat{Cov}(r_{ij}, r_{k\ell}). \tag{48}$$

[2] For the purpose of deriving equation (23), whether there is any statistical bias in each sample correlation r_{ij} as an estimator of the true but unknown correlation ρ_{ij} is not an issue. It becomes an issue, only when the expected value of each r_{ij} is used for the corresponding ρ_{ij} in the implementation of equation (23). Thus, this is the place, not any earlier, to consider the issue.

The double summation in A and the quadruple summation in B cover, in total, $n(n-1)/2$ terms and $[n(n-1)/2]^2$ terms, respectively.

We provide below a direct proof of $A - B > 0$. Before proceeding, notice that, due to the presence of the multiplicative factor $n(n-1)/2$, each of the sample variance terms in A appears $n(n-1)/2$ times. Notice also that, as $\widehat{Cov}(r_{ij}, r_{ij}) = \widehat{Var}(r_{ij})$, $n(n-1)/2$ of the $[n(n-1)/2]^2$ terms in B, where both $k = i$ and $\ell = j$, are sample variances. As A and B have the same number of additive terms, the proof below is based on comparisons between individual pairs of terms from each sum of terms.

For any pair of $\widehat{Var}(r_{ij})$ and $\widehat{Var}(r_{k\ell})$, as

$$\left[\sqrt{\widehat{Var}(r_{ij})} - \sqrt{\widehat{Var}(r_{k\ell})}\right]^2 \geq 0, \tag{49}$$

we have

$$\widehat{Var}(r_{ij}) + \widehat{Var}(r_{k\ell}) \geq 2\sqrt{\widehat{Var}(r_{ij})}\sqrt{\widehat{Var}(r_{k\ell})}. \tag{50}$$

Provided that the two random variables r_{ij} and $r_{k\ell}$ are not perfectly and positively correlated, we also have

$$\sqrt{\widehat{Var}(r_{ij})}\sqrt{\widehat{Var}(r_{k\ell})} > \widehat{Cov}(r_{ij}, r_{k\ell}). \tag{51}$$

which leads to

$$\widehat{Var}(r_{ij}) + \widehat{Var}(r_{k\ell}) > 2\widehat{Cov}(r_{ij}, r_{k\ell}) \tag{52}$$

or, equivalently,

$$\widehat{Var}(r_{ij}) + \widehat{Var}(r_{k\ell}) > \widehat{Cov}(r_{ij}, r_{k\ell}) + \widehat{Cov}(r_{k\ell}, r_{ij}). \tag{53}$$

It is only when both $k = i$ and $\ell = j$ that we have equality of the two sides.

After netting out the $n(n-1)/2$ sample variance terms in A and B where both $k = i$ and $\ell = j$, we have

$$\frac{1}{2}\left\{\left[\frac{n(n-1)}{2}\right]^2 - \frac{n(n-1)}{2}\right\} = \frac{(n+1)(n)(n-1)(n-2)}{8} \tag{54}$$

pairs of sample variances in A left, for comparisons with the corresponding pairs of sample covariances in B. Given inequality (53), the positive sign of $A - B$ is assured. To illustrate, let us consider the case where $n = 4$. In this case, as $n(n-1)/2 = 6$, $[n(n-1)/2]^2 = 36$, and $(n+1)(n)(n-1)(n-2)/8 = 15$, we can write, explicitly,

$$A = 6\widehat{Var}(r_{12}) + 6\widehat{Var}(r_{13}) + 6\widehat{Var}(r_{14}) + 6\widehat{Var}(r_{23}) + 6\widehat{Var}(r_{24}) + 6\widehat{Var}(r_{34}) \tag{55}$$

and

$$\begin{aligned} B \;=\; & \widehat{Var}(r_{12}) + \widehat{Cov}(r_{12}, r_{13}) + \widehat{Cov}(r_{12}, r_{14}) + \widehat{Cov}(r_{12}, r_{23}) + \widehat{Cov}(r_{12}, r_{24}) + \widehat{Cov}(r_{12}, r_{34}) \\ & + \widehat{Cov}(r_{13}, r_{12}) + \widehat{Var}(r_{13}) + \widehat{Cov}(r_{13}, r_{14}) + \widehat{Cov}(r_{13}, r_{23}) + \widehat{Cov}(r_{13}, r_{24}) + \widehat{Cov}(r_{13}, r_{34}) \\ & + \widehat{Cov}(r_{14}, r_{12}) + \widehat{Cov}(r_{14}, r_{13}) + \widehat{Var}(r_{14}) + \widehat{Var}(r_{14}, r_{23}) + \widehat{Cov}(r_{14}, r_{24}) + \widehat{Cov}(r_{14}, r_{34}) \\ & + \widehat{Cov}(r_{23}, r_{12}) + \widehat{Cov}(r_{23}, r_{13}) + \widehat{Cov}(r_{23}, r_{14}) + \widehat{Var}(r_{23}) + \widehat{Cov}(r_{23}, r_{24}) + \widehat{Cov}(r_{23}, r_{34}) \\ & + \widehat{Cov}(r_{24}, r_{12}) + \widehat{Cov}(r_{24}, r_{13}) + \widehat{Cov}(r_{24}, r_{14}) + \widehat{Cov}(r_{24}, r_{23}) + \widehat{Var}(r_{24}) + \widehat{Cov}(r_{24}, r_{34}) \\ & + \widehat{Cov}(r_{34}, r_{12}) + \widehat{Cov}(r_{34}, r_{13}) + \widehat{Cov}(r_{34}, r_{14}) + \widehat{Cov}(r_{34}, r_{23}) + \widehat{Cov}(r_{34}, r_{24}) + \widehat{Var}(r_{34}). \end{aligned} \tag{56}$$

After netting out the 6 sample variance terms in A and B, we have 15 pairs of sample variances in A and the corresponding 15 pairs of sample covariances in B for comparisons. Given inequality (53), it is straightforward to establish that $A - B > 0$.

2.5 Bias Correction

Even under the stationarity assumption of security return distributions, to estimate ρ_{ij} for $i \neq j$ without relying on asymptotic properties is not a simple task. This is because $E(r_{ij})$ is a biased estimator of ρ_{ij}. [See, for example, Zimmerman, Zumbo, and Williams (2003).] The nature of the bias is that the sample correlation tends to understate the true correlation in magnitude. The bias attenuates as the sample period lengthens. However, the use of a longer sample period is susceptible to violations of the stationarity assumption. As the analytical task to find an unbiased estimator is beyond the scope of this paper, only an approximate result from the statistics literature is duplicated here instead.

As reported in Zimmerman, Zumbo, and Williams (2003), the following estimator of ρ_{ij}, attributed to Olkin and Pratt (1958), is able to eliminate most of the bias in the sample correlation r_{ij} of normally distributed variables i and j:

$$\widehat{\rho}_{ij} = r_{ij} \left[1 + \frac{1 - r_{ij}^2}{2(T-3)} \right]. \tag{57}$$

With the bias in the estimation of ρ_{ij} accounted for, the estimated λ based on equation (23) is given by

$$\widehat{\lambda} = \frac{\sum_{i=1}^{n-1} \sum_{j=i+1}^{n} (\widehat{\alpha}_{ij} - \widehat{\gamma}_{ij})}{\sum_{i=1}^{n-1} \sum_{j=i+1}^{n} [\widehat{\alpha}_{ij} + (\overline{r} - r_{ij})^2]}. \tag{58}$$

Here,

$$\widehat{\gamma}_{ij} = (r_{ij} - \overline{r}) \left[\frac{r_{ij}(1 - r_{ij}^2)}{2(T-3)} \right] \tag{59}$$

is the estimated value of $E(\overline{r} - r_{ij})[E(r_{ij}) - \rho_{ij}]$ in equation (23).

Does bias correction lead to a lower or higher shrinkage intensity? The answer hinges on the sign of the sum $\sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \widehat{\gamma}_{ij}$ or, equivalently, the sign of the sum $\sum_{i=1}^{n-1} \sum_{j=i+1}^{n} [r_{ij}(1 - r_{ij}^2)](r_{ij} - \overline{r})$. The latter sum can be viewed as a weighted sum of the $n(n-1)/2$ values of $(r_{ij} - \overline{r})$, for $i = 1, 2, \ldots, n-1$ and $j = i+1, i+2, \ldots, n$, with the corresponding weight for each $(r_{ij} - \overline{r})$ term provided by $r_{ij}(1 - r_{ij}^2)$. The weighted average of the $n(n-1)/2$ values of $(r_{ij} - \overline{r})$ is the weighted sum divided by the sum of the $n(n-1)/2$ weights. Although both the sum $\sum_{i=1}^{n-1} \sum_{j=i+1}^{n} (r_{ij} - \overline{r})$ and the average, which is the sum divided by $n(n-1)/2$, are zeros, the sign of the weighted sum and the sign of the weighted average will depend on the weights involved.

To find the sign of the weighted sum, let us first sort the $n(n-1)/2$ individual values of r_{ij} in an ascending order. Notice that sorting r_{ij} is equivalent to sorting the corresponding values of $(r_{ij} - \overline{r})$. If there is any negative value of r_{ij}, as $\overline{r}$ is positive, the corresponding $[r_{ij}(1 - r_{ij}^2)](r_{ij} - \overline{r}_{ij})$ term will always be positive, thus contributing to the attainment of a positive weighted sum. Strictly zero values of r_{ij} will have no contributions to the weighted sum. The eventual sign of the weighted sum still depends on how the corresponding weight $r_{ij}(1 - r_{ij}^2)$ progresses from one sorted value of $(r_{ij} - \overline{r})$ to the next. If the sorted values of r_{ij} correspond to progressively higher weights, then the weighted sum will be positive. Here is why:

Suppose for now that none of the values of r_{ij} are negative. As the sum $\sum_{i=1}^{n-1} \sum_{j=i+1}^{n} (r_{ij} - \overline{r})$ is zero, the average of the $n(n-1)/2$ values of $(r_{ij} - \overline{r})$ will also be zero. If it turns out that $r_{ij}(1 - r_{ij}^2)$ always increases from one sorted value of $(r_{ij} - \overline{r})$ to the next, the weighted average of the $n(n-1)/2$ terms of $(r_{ij} - \overline{r})$ will be greater than the average. In this scenario, the weighted average will be positive. Then, the weighted sum, which is $\sum_{i=1}^{n-1} \sum_{j=i+1}^{n} [r_{ij}(1 - r_{ij}^2)](r_{ij} - \overline{r})$, will be positive as well.

Now, suppose instead that there are some negative values of r_{ij}. As the sum $\sum_{i=1}^{n-1} \sum_{j=i+1}^{n} (r_{ij} - \overline{r})$ is zero, the partial sum where each r_{ij} is positive will be positive. So will the average of the terms involved in the partial sum. For the partial sum, each weight as provided by $r_{ij}(1 - r_{ij}^2)$

is positive. In the above-mentioned scenario, the weighted average of the terms involved will be greater than the average. As the average is positive, the weighted average will also be positive. Then, so will the part of the sum $\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}[r_{ij}(1-r_{ij}^2)](r_{ij}-\overline{r})$ where r_{ij} is positive. As the part of the sum where r_{ij} is negative is positive, the sum of the two parts will be positive too.

To establish a sufficient condition of progressively higher weights for the sorted values of r_{ij}, let $f(x) = x(1-x^2)$ be a continuous function of x, defined for $0 \leq x < 1$. The first derivative of the function is positive for $0 \leq x < \sqrt{3}/3$ (≈ 0.5774), is zero for $x = \sqrt{3}/3$, and is negative for $x > \sqrt{3}/3$. For a set of values of x matching the given set of sorted values of r_{ij}, the corresponding values of $r_{ij}(1-r_{ij}^2)$ will increase from one sorted value of $(r_{ij}-\overline{r})$ to the next, provided that $0 \leq r_{ij} < \sqrt{3}/3$. A direct implication is that, if none of the $n(n-1)/2$ values of r_{ij} are greater than $\sqrt{3}/3$, as the weighted sum $\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}[r_{ij}(1-r_{ij}^2)](r_{ij}-\overline{r})$ is positive, bias correction will lead to a lower shrinkage intensity.

Does the above sufficient condition always hold? For implementing a portfolio selection model, it is good practice to avoid considering, for the same portfolio, any securities whose returns are highly correlated, as the effectiveness of portfolio diversification will be weakened by the presence of such securities. Given this practice, sample correlations of returns for portfolio decisions are mostly likely lower than $\sqrt{3}/3$, but higher sample correlations are still possible. As long as the sample correlations involved are predominantly lower than $\sqrt{3}/3$, it is unlikely that the outliers, which cause violations of the sufficient condition, will lead to a sign reversal of the weighted sum $\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}[r_{ij}(1-r_{ij}^2)](r_{ij}-\overline{r})$. However, whenever the sufficient condition is violated, it is advisable to check the impact of bias correction on the shrinkage results.

3 Shrinkage Towards a Zero Correlation Target

For a zero correlation target, which is an identity matrix, the corresponding analytical materials also start with positive linear transformations of the return observations. With $\overline{r}$ substituted by zero, equation (13) becomes

$$L(\lambda) = E\left\{\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}[(1-\lambda)r_{ij}-\rho_{ij}]^2\right\} = \sum_{i=1}^{n-1}\sum_{j=i+1}^{n} E\left\{[(1-\lambda)r_{ij}-\rho_{ij}]^2\right\}. \quad (60)$$

Noting again that $E(x^2) = Var(x) + [E(x)]^2$, for any random variable x, we can write

$$L(\lambda) = \sum_{i=1}^{n-1}\sum_{j=i+1}^{n}\left\{Var[(1-\lambda)r_{ij}-\rho_{ij}] + \left\{E[(1-\lambda)r_{ij}-\rho_{ij}]\right\}^2\right\}, \quad (61)$$

which, upon simplifications, leads to

$$L(\lambda) = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \left\{(1-\lambda)^2 Var(r_{ij}) + [(1-\lambda)E(r_{ij}) - \rho_{ij}]^2\right\}. \tag{62}$$

The optimal shrinkage intensity can be deduced from

$$\frac{dL(\lambda)}{d\lambda} = -2 \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \left\{(1-\lambda) Var(r_{ij}) + [(1-\lambda)E(r_{ij}) - \rho_{ij}]E(r_{ij})\right\} = 0. \tag{63}$$

The result is

$$\lambda = \frac{\sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \left\{Var(r_{ij}) + E(r_{ij})[E(r_{ij}) - \rho_{ij}]\right\}}{\sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \{Var(r_{ij}) + [E(r_{ij})]^2\}}. \tag{64}$$

Equation (64) is much simpler than equations (23) and (24), as no more covariance terms between any sample correlations are involved here.

If the bias in the estimation of ρ_{ij} with $E(r_{ij})$ is ignored, equation (64) reduces to

$$\lambda = \frac{\sum_{i=1}^{n-1} \sum_{j=i+1}^{n} Var(r_{ij})}{\sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \{Var(r_{ij}) + [E(r_{ij})]^2\}}, \tag{65}$$

which implies that $0 < \lambda < 1$, as expected. The estimated shrinkage intensity is

$$\widehat{\lambda} = \frac{\sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \widehat{Var}(r_{ij})}{\sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \left[\widehat{Var}(r_{ij}) + r_{ij}^2\right]}, \tag{66}$$

where $\widehat{Var}(r_{ij})$ is estimated by using equation (34). If bias correction is based on equation (57), we have

$$\widehat{\lambda} = \frac{\sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \left[\widehat{Var}(r_{ij}) - \widehat{\gamma}_{ij}\right]}{\sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \left[\widehat{Var}(r_{ij}) + r_{ij}^2\right]} \tag{67}$$

instead, where $\widehat{\gamma}_{ij}$ is also given by equation (59).

From a computational standpoint, equation (58) accommodates equation (44) as a special case, by substituting each $\widehat{\gamma}_{ij}$ term there with a zero. Likewise, equation (67) accommodates equation (66) as a special case, with the same substitutions. Equation (58) also accommodates equation (67) as a special case, by substituting the random variable $\overline{r}$ there with a zero (which is non-random). This substitution leads to $\widehat{\alpha}_{ij} = \widehat{Var}(r_{ij})$ in equation (58). In turn, equation (44) accommodates equation (66) as a special case, by substituting the random variable $\overline{r}$ there with a zero. All together, based on the analytical materials in this paper, we have four different ways to shrink the sample correlation matrix, as there are two choices between shrinkage targets and two choices regarding bias correction.

4 An Excel Illustration

This section starts with a small scale example based on a small subset of the current Dow Jones stocks and a short sample period. The example is intended to provide an Excel illustration of the above-mentioned four different ways to shrink the sample correlation matrix of returns. From a computational standpoint, there are three approaches to perform the same task. Specifically, they include an approach not involving VBA, an approaches requiring a user-defined Excel function coded in VBA, and a variation of the latter approach, which uses an Excel Macro to prompt the user for inputs, instead of entering directly the set of arguments for the same Excel function.

This section then uses the two VBA-based approaches to produce shrinkage results for the current 30 Dow Jones stocks. The sample periods, all ending at December 2016, include $T = 36$, 48, 60, 72, and 84 months (ranging from three to seven years). By covering such sample periods, we illustrate how the choice of the length of a sample period affects the shrinkage results. The same illustration will also give students a good idea of how the returns of major U.S. stocks are correlated, how the choice of a shrinkage target affects the end results, and to what extent is bias correction relevant in practice.

4.1 A Small Scale Example

Figure 1, which is based on one of the two Excel files accompanying this paper, illustrates the computations involved in Sections 2 and 3, by using an example where $n = 5$ and $T = 6$. The five Dow Jones stocks in the illustration are Apple (AAPL), American Express (AXP), Cisco Systems (CISO), Proctor & Gamble (PG), and Visa (V). The six-month sample period is from July to December 2016.

All formulas pertaining to the computations in Figure 1 can be found in the corresponding Excel file, and thus only some selected formulas are explicitly described below. The monthly return observations for this small subset of Dow Jones stocks are shown in B3:F8, with the corresponding ticker symbols and dates indicated in B2:F2 and A3:A8, respectively. As the six-month sample period is obviously too short for the corresponding shrinkage results to be meaningful, its use is primarily for illustrating the computations involved. Notice that the monthly return observations in B3:F8 are displayed in a reverse chronological order, corresponding to

	A	B	C	D	E	F	G
1	Return Observations						
2		AAPL	AXP	CSCO	PG	V	
3	12/30/2016	0.047955	0.028318	0.013414	0.019646	0.009053	
4	11/30/2016	-0.021578	0.084613	-0.028031	-0.05	-0.060902	
5	10/31/2016	0.004334	0.042161	-0.02459	-0.025404	-0.002297	
6	9/30/2016	0.065504	-0.023483	0.008906	0.027946	0.02225	
7	8/31/2016	0.023606	0.017375	0.029807	0.020096	0.038309	
8	7/29/2016	0.090063	0.060895	0.073196	0.018779	0.052312	
9							
10	Mean	0.03498067	0.03497983	0.012117	0.00184383	0.0097875	
11	Standard Deviations	0.0409916	0.03730638	0.03745621	0.03177549	0.03980126	
12							
13	Transformed Observations						
14		AAPL	AXP	CSCO	PG	V	
15	12/30/2016	0.31651203	-0.1785709	0.0346271	0.56024836	-0.0184542	
16	11/30/2016	-1.3797625	1.3304204	-1.0718649	-1.6315667	-1.7760617	
17	10/31/2016	-0.7476329	0.19249166	-0.9799977	-0.857511	-0.303621	
18	9/30/2016	0.74462418	-1.5671002	-0.0857268	0.82145596	0.3131182	
19	8/31/2016	-0.2774878	-0.4718988	0.47228482	0.57441021	0.71659785	
20	7/29/2016	1.34374699	0.6946578	1.63067747	0.53296318	1.06842085	
21							
22	Mean	7.401E-17	1.665E-16	-7.401E-17	-3.701E-17	7.401E-17	
23	Standard Deviations	1	1	1	1	1	
24							
25	Sample Correlation Matrix						
26		AAPL	AXP	CSCO	PG	V	
27	AAPL	1	-0.4277215	0.84377759	0.84761104	0.82833899	
28	AXP	-0.4277215	1	-0.1153242	-0.7247835	-0.500944	
29	CSCO	0.84377759	-0.1153242	1	0.73570665	0.85089053	
30	PG	0.84761104	-0.7247835	0.73570665	1	0.87720909	
31	V	0.82833899	-0.500944	0.85089053	0.87720909	1	
32					Ave Correl	0.3214761	
33							
34	Shrinkage Intensity						
35		Without VBA		Using Function		Using Macro	
36				SHRINK		Ctrl + s	
37	Constant Correlation Target						
38							
39	Bias Corrected	0.22774811		0.22774811		0.22774811	
40							
41	Bias Ignored	0.27889835		0.27889835		0.27889835	
42							

Figure 1: An Excel Illustration of the Computations Involved in Shrinkage of the Sample Correlation Matrix Towards a Constant Correlation Target or a Zero Correlation Target, with or without Bias Correction

	A	B	C	D	E	F	G
43	Zero Correlation Target						
44							
45	Bias Corrected	0.24677483		0.24677483		0.24677483	
46							
47	Bias Ignored	0.28948748		0.28948748		0.28948748	
48							
49	Correlation Matrix After Shrinkage						
50	Constant Correlation Target						
51	Bias Corrected	AAPL	AXP	CSCO	PG	V	
52	AAPL	1	-0.2570932	0.7248244	0.72778479	0.71290192	
53	AXP	-0.2570932	1	-0.0158438	-0.4864999	-0.3136394	
54	CSCO	0.7248244	-0.0158438	1	0.64136642	0.73031738	
55	PG	0.72778479	-0.4864999	0.64136642	1	0.75064195	
56	V	0.71290192	-0.3136394	0.73031738	0.75064195	1	
57					Ave Correl	0.3214761	
58							
59	Constant Correlation Target						
60	Bias Ignored	AAPL	AXP	CSCO	PG	V	
61	AAPL	1	-0.2187715	0.69810855	0.70087286	0.68697576	
62	AXP	-0.2187715	1	0.00649866	-0.4329835	-0.2715724	
63	CSCO	0.69810855	0.00649866	1	0.62017842	0.70323771	
64	PG	0.70087286	-0.4329835	0.62017842	1	0.72221607	
65	V	0.68697576	-0.2715724	0.70323771	0.72221607	1	
66					Ave Correl	0.3214761	
67							
68	Zero Correlation Target						
69	Bias Corrected	AAPL	AXP	CSCO	PG	V	
70	AAPL	1	-0.3221706	0.63555452	0.63844197	0.62392578	
71	AXP	-0.3221706	1	-0.0868651	-0.5459252	-0.3773237	
72	CSCO	0.63555452	-0.0868651	1	0.55415277	0.64091216	
73	PG	0.63844197	-0.5459252	0.55415277	1	0.66073597	
74	V	0.62392578	-0.3773237	0.64091216	0.66073597	1	
75					Ave Correl	0.2421439	
76							
77	Zero Correlation Target						
78	Bias Ignored	AAPL	AXP	CSCO	PG	V	
79	AAPL	1	-0.3039015	0.59951454	0.60223825	0.58854522	
80	AXP	-0.3039015	1	-0.0819393	-0.5149678	-0.355927	
81	CSCO	0.59951454	-0.0819393	1	0.52272878	0.60456837	
82	PG	0.60223825	-0.5149678	0.52272878	1	0.62326804	
83	V	0.58854522	-0.355927	0.60456837	0.62326804	1	
84					Ave Correl	0.2284128	

Figure 1: An Excel Illustration of the Computations Involved in Shrinkage of the Sample Correlation Matrix Towards a Constant Correlation Target or a Zero Correlation Target, with or without Bias Correction (Continued)

	H	I	J	K	L	M	N	O	P	Q	R
1		# Var		# Obs							
2		5		6							
3											
4		# Var - 1		# Obs - 1							
5		4		5							
6											
7	Sum of (Wij t-Wij Bar)^2, for t = 1 to T										
8		I,j									
9		1,2	1,3	1,4	1,5	2,3	2,4	2,5	3,4	3,5	4,5
10		4.8814	4.5803	3.4292	5.3529	3.3646	4.428	5.4471	2.3436	3.8466	5.8185
11											
12		Sum Variances									
13		2.0876									
14											
15	Sum of (Wij t-Wij Bar)*(Wkl t-Wkl Bar), for t = 1 to T										
16	i.j\k,l	1,2	1,3	1,4	1,5	2,3	2,4	2,5	3,4	3,5	4,5
17	1,2	4.8814	0.7856	-2.79	-2.012	3.3111	4.5346	4.6741	-1.092	-0.319	-3.508
18	1,3	0.7856	4.5803	2.3721	4.0353	0.6609	0.1448	-0.074	2.489	3.8169	2.5695
19	1,4	-2.79	2.3721	3.4292	3.9381	-1.996	-2.929	-3.302	2.4169	2.6485	4.0899
20	1,5	-2.012	4.0353	3.9381	5.3529	-1.437	-2.569	-3.056	3.1144	4.2236	4.9283
21	2,3	3.3111	0.6609	-1.996	-1.437	3.3646	3.0858	3.99	-1.384	-0.467	-3.172
22	2,4	4.5346	0.1448	-2.929	-2.569	3.0858	4.428	4.6237	-1.376	-1.024	-3.915
23	2,5	4.6741	-0.074	-3.302	-3.056	3.99	4.6237	5.4471	-2.058	-1.548	-4.874
24	3,4	-1.092	2.489	2.4169	3.1144	-1.384	-1.376	-2.058	2.3436	2.5788	3.1857
25	3,5	-0.319	3.8169	2.6485	4.2236	-0.467	-1.024	-1.548	2.5788	3.8466	3.6084
26	4,5	-3.508	2.5695	4.0899	4.9283	-3.172	-3.915	-4.874	3.1857	3.6084	5.8185
27											
28		Sum Covariances									
29		5.0561									
30											
31	r: Sample Correl; r Bar - r: Average Correl - Sample Correl; Bias: r*(1-r^2)/2/(T-3)										
32		1,2	1,3	1,4	1,5	2,3	2,4	2,5	3,4	3,5	4,5
33	r	-0.428	0.8438	0.8476	0.8283	-0.115	-0.725	-0.501	0.7357	0.8509	0.8772
34	r Bar - r	0.7492	-0.522	-0.526	-0.507	0.4368	1.0463	0.8224	-0.414	-0.529	-0.556
35	Bias	-0.058	0.0405	0.0398	0.0433	-0.019	-0.057	-0.063	0.0562	0.0391	0.0337
36											
37	Shrinkage Target			Sum (r Bar - r)^2			Sum r^2				
38	1	0		4.0904			5.1238				
39											
40	Bias Correction			Sum (r Bar - r)(Bias)			Sum (-r)(Bias)				
41	1	0		-0.2901			-0.3080				
42											

Figure 1: An Excel Illustration of the Computations Involved in Shrinkage of the Sample Correlation Matrix Towards a Constant Correlation Target or a Zero Correlation Target, with or without Bias Correction (Continued)

$T = 6, 5, 4, \ldots, 1$.

The first step of the computational task is to transform the return observations, so that the transformed observations for each stock will have a zero mean and a unit standard deviation. The transformed observations are displayed in B15:F20. The sample means and standard deviations before and after the transformations are displayed in B10:F11 and B22:F23, respectively. The transformed observations in B15:F20 are then used directly to produce the sample correlation matrix, as displayed in B27:F31.

For computational convenience, two Excel functions for matrix operations (MMULT and TRANSPOSE) are nested, so that the transpose of the 6×5 matrix in B15:F20 can be multiplied to the matrix itself to produce a 5×5 matrix. Specifically, after selecting the block of cells B27:F31 and entering the formula =MMULT(TRANSPOSE(B15:F20),B15:F20)/(COUNT(B15:B20)-1), which is based on equation (11), we must press the Shift+Ctrl+Enter keys together to obtain a 5×5 sample correlation matrix of returns. Here, the use of function COUNT is for providing the number of return observations. The average of the sample correlations, as shown in F32, is computed as the sum of the 25 cells in B27:F31 net of the sum of the diagonal elements (which is 5), divided by the total number of off-diagonal elements (which is 20). Notice that this approach is more convenient than taking the average of the 10 elements in the upper triangle of the 5×5 sample correlation matrix.

Three alternative approaches are used to compute the shrinkage intensity. The first approach does not require the use of VBA. The two remaining approaches do; they differ in that one of them computes the shrinkage intensity directly via a user-defined function, called SHRINK, and the other one uses a Macro to prompt the user for inputs. Regardless of the approach involved, the shrinkage intensity is computed in four different ways, as described in the final paragraph of Section 3. The computed values of the shrinkage intensity based on the non-VBA approach are displayed in B39, B41, B45, and B47. These values correspond to the four cases where a constant correlation target or a zero correlation target is used, with or without bias correction in each case. The computed values based on the two VBA approaches are displayed in the corresponding cells in columns D and F.

As expected, the results from the three alternative approaches are consistent with each other. The post-shrinkage correlation matrices based on the four computed values of shrinkage intensity are displayed in B52:F56, B61:F65, B70:F74, and B79:F83. The corresponding post-shrinkage average correlations are displayed in F57, F66, F75, and F84. To show how different values of the shrinkage intensity can be reached, let us start with the non-VBA approach in the following.

For $n = 5$, there are $n(n-1)/2 = 10$ off-diagonal elements in the upper triangle of the correlation matrix. Thus, the double summation $\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}\widehat{Var}(r_{ij})$ covers a total of 10 terms, for $(i,j) = (1,2)$, $(1,3)$, $(1,4)$, $(1,5)$, $(2,3)$, $(2,4)$, $(2,5)$, $(3,4)$, $(3,5)$, and $(4,5)$. For each pair of i and j, we compute $\widehat{Var}(r_{ij})$ by using equation (34). The 10 computed values of $\sum_{t=1}^{T}(W_{ijt} - \overline{\boldsymbol{W}}_{ij})^2$, where $T = 6$, are displayed in I10:R10. Their sum, multiplied by $T/(T-1)^3 = 6/5^3$, is $\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}\widehat{Var}(r_{ij})$; the result is displayed in I13.

The sum $\sum_{t=1}^{6}(W_{ijt} - \overline{\boldsymbol{W}}_{ij})^2$ for each pair of i and j can be computed directly by using a cell formula. We can either nest the Excel functions MMULT and TRANSPOSE for matrix operations or use the Excel function SUM for the sum of products. Let us use $(i,j) = (1,2)$ for an illustration. The formula for I10 is either =MMULT(TRANSPOSE($B$15:$B$20*C$15:C$20-$K$5/$K$2*C$27),$B$15:$B$20*C$15:C$20-$K$5/$K$2*C$27) or =SUM(($B$15:$B$20*C$15:C$20-$K$5/$K$2*C$27)*($B$15:$B$20*C$15:C$20-$K$5/$K$2*C$27)). For either formula to work as intended, the Shift+Ctrl+Enter keys must be pressed together. The part ($B$15:$B$20*C$15:C$20-$K$5/$K$2*C$27) in either formula is to generate directly a column of numbers corresponding to $W_{12t} - \overline{\boldsymbol{W}}_{12}$, for $t = 1, 2, \ldots, 6$, where $\overline{\boldsymbol{W}}_{12} = r_{12} \times 5/6$ according to equation (26). The difference in the two formulas is in how $\sum_{t=1}^{6}(W_{12t} - \overline{\boldsymbol{W}}_{12})^2$ is computed.

For a constant correlation target, the quadruple summation $\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}\sum_{k=1}^{n-1}\sum_{\ell=k+1}^{n}\widehat{Cov}$ $(r_{ij}, r_{k\ell})$ consists of 100 terms, with $(i,j) = (1,2)$, $(1,3)$, $(1,4)$, $(1,5)$, $(2,3)$, $(2,4)$, $(2,5)$, $(3,4)$, $(3,5)$, and $(4,5)$, and with the same coverage for (k,ℓ). To use equation (41) for computing $\widehat{Cov}(r_{ij}, r_{k\ell})$, the corresponding sum $\sum_{t=1}^{6}(W_{ijt} - \overline{\boldsymbol{W}}_{ij})(W_{k\ell t} - \overline{\boldsymbol{W}}_{k\ell})$ for each case of (i,j) and (k,ℓ) is computed first. The 100 sums are displayed in the 10×10 block I17:R26.

To illustrate the computations involved, let us consider the case where $(i,j) = (1,2)$ and $(k,\ell) = (4,5)$, as displayed in R17. The cell formula there can be either =MMULT(TRANSPOSE ($B$15:$B$20*$C$15:$C$20-$K$5/$K$2*$C$27),$E$15:$E$20*F$15:F$20-$K$5/$K$2*F$30) or =SUM(($B$15:$B$20*$C$15:$C$20-$K$5/$K$2*$C$27)*($E$15:$E$20*F$15:F$20-$K$5/$K$2 *F$30)), depending on how $\sum_{t=1}^{6}(W_{12t} - \overline{\boldsymbol{W}}_{12})(W_{45t} - \overline{\boldsymbol{W}}_{45})$ is computed. Analogous to the sample variance in I10, the parts ($B$15:$B$20*$C$15:$C$20-$K$5/$K$2*$C$27) and ($E$15:$E$20 *F$15:F$20-$K$5/$K$2*F$30) in either formula in R17 are for generating the columns of numbers $(W_{12t} - \overline{\boldsymbol{W}}_{12})$ and $(W_{45t} - \overline{\boldsymbol{W}}_{45})$, respectively, for $t = 1, 2, \ldots, 6$. The sum of the 100 terms in I17:R26, multiplied by $T/(T-1)^3 = 6/5^3$, is $\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}\sum_{k=1}^{n-1}\sum_{\ell=k+1}^{n}\widehat{Cov}\ (r_{ij}, r_{k\ell})$; the result is displayed in I29. Notice that, for a zero correlation target, the computations leading to this quadruple summation are unnecessary.

To compute the four values of $\widehat{\lambda}$, for the two alternative shrinkage targets with or without bias correction in each case, values of $\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}(\overline{r}-r_{ij})^2$, $\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}r_{ij}^2$, $\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}(\overline{r}-r_{ij})(r_{ij})(1-r_{ij}^2)/[2(T-3)]$, and $-\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}r_{ij}^2(1-r_{ij}^2)/[2(T-3)]$ are also required. These four double summations, for $n=5$, can be computed by first setting up the following three rows of numbers: r_{ij}, $(\overline{r}-r_{ij})$, and $r_{ij}(1-r_{ij}^2)/[2(T-3)]$ for the 10 cases of (i,j) as described earlier. They are displayed in I33:R35. The corresponding values of the four double summations are displayed in K38, N38, K41, and N41.

These values can be reached by either nesting the Excel functions MMULT and TRANSPOSE for matrix operations or using the Excel function SUM for the sum of products as described earlier. For example, to compute $\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}(\overline{r}-r_{ij})(r_{ij})(1-r_{ij}^2)/[2(T-3)]$ in K41, the formula can be either =MMULT(I34:R34,TRANSPOSE(I35:R35)) or =SUM(I34:R34*I35:R35). The computed values of $\widehat{\lambda}$ based on equations (44), (58), (66), and (67) are displayed in B41, B39, B47, and B45, respectively. This step completes the illustration of the non-VBA approach.

A crucial point in the determination of the shrinkage intensity is that, for a finite sample of return observations for n securities, each of the $n(n-1)/2$ sample correlations in the upper triangle of the sample correlation matrix, which is subject to estimation error, is in itself a random variable from a statistical standpoint. For a zero correlation target, the determination of the shrinkage intensity requires that the $n(n-1)/2$ sample variances of such random variables be computed. For a constant correlation target, however, the corresponding task requires also that the $[n(n-1)/2]^2$ sample covariances between such random variables be computed.

In the former case, the computational task can be achieved in three nested loops, with the innermost loop providing the sample variance for each combination of securities as specified in the remaining loops. In the latter case, five nested loops are also required, as four loops are needed to select pairs of random variables for sample covariance computations. The VBA coding here has captured such features.

To compute $\widehat{\lambda}$ with a VBA approach, we can, in principle, use directly the return observations in B3:F8 as the only input. However, the VBA coding is much less tedious, if we use instead, as inputs, both the transformed observations in B15:F20 and the sample correlation matrix in

B27:F31, as well as the user's choice of the shrinkage target and of bias correction. The VBA coding is primarily for the most tedious case, where both shrinkage of the sample correlation matrix is towards the constant correlation target and the bias is corrected. The corresponding results for the three remaining cases where a zero correlation target is used and/or the bias is ignored can be reached, by bypassing any tedious but unnecessary computations and by multiplying the less tedious but unnecessary computational components with zeros.

The VBA code for the user-defined function SHRINK can be accessed by selecting the menu item Visual Basic under the Developer tab. The function has four arguments, with the first two arguments being the locations of the cells for the transformed observations and the sample covariance matrix, and with each of the remaining arguments being a binary $(0, 1)$ choice in how $\widehat{\lambda}$ is computed. Among the four cells in D39, D41, D45, and D47, the formula in D39, for example, which is =SHRINK(B15:F20,B27:F31,1,1), indicates that a constant correlation target is used and that bias is corrected. Any changes to the binary codes in the third and fourth arguments will lead to changes in how $\widehat{\lambda}$ is computed, and the corresponding results are displayed in D41, D45, and D47.

A variant of the above approach, which requires the user-defined function SHRINK, is to use an Excel Macro to prompt the user for the four arguments of the function. The VBA code is listed under the subroutine "Sub ViaFunction()." The Macro has been set up for the keyboard shortcut Ctrl+s. If the shortcut key is lost, it can easily be reinstated by selecting the menu item Macro under the Developer tab and entering the shortcut key again in Options.

Once the keys Ctrl+s are pressed, the user is prompted the following, which is the first of the five prompts: "Select the Cells Containing All Transformed Observations." The response ought to be B15:F20. The user is then prompted the following: "Select the Cells Containing the Sample Correlation Matrix." The response ought to be B27:F31. The next two prompts, "Shrinkage Target? 1 for Constant Correlation Target; 0 for Zero Correlation Target" and "Bias Correction? 1 for Yes; 0 for No," require a binary response in each case. Finally, the user is prompted the following: "Select the Cell for Displaying the Output." For example, if the responses to the third and fourth prompts are both 1, then the response to the final prompt is F39. Other responses to the last three of the five prompts correspond to the displays in F41, F45, and F47.

4.2 Shrinkage Results Based on the Current 30 Dow Jones Stocks

We now turn our attention to shrinkage estimation of the 30×30 correlation matrix of returns. It is for the full set of Dow Jones stocks and for different lengths of the sample period, ranging from $T = 36$ to $T = 84$ at 12-month intervals, all having December 2016 as the final month of return observations. Table 1 summarizes the pre-shrinkage and post-shrinkage results. As indicated in the introductory section, the choice of the length of a sample period is often a trade-off between satisfying the stationarity assumption of the joint probability distribution of returns and reducing the estimation errors. By estimating the correlation matrix with past return observations, we implicitly accept the stationarity assumption, even for a sample period that is as lengthy as 84 months. No attempt is made here to establish the optimal length of a sample period from a statistical standpoint.

Table 1: A Summary of Pre-Shrinkage and Post-Shrinkage Results for the Current 30 Dow Jones Stocks Based on Different Numbers of Monthly Return Observations

# of Monthly Return Observations	36	48	60	72	84
(A) Sample Correlations					
Average	0.3391	0.3259	0.3079	0.3202	0.3564
Maximum	0.8957	0.8805	0.8653	0.8669	0.8523
Minimum	−0.2229	−0.1392	−0.2173	−0.1425	−0.0452
# of Sample Correlations Above $\sqrt{3}/3$	33	25	16	25	30
Percentage (out of 435)	7.59%	5.75%	3.68%	5.75%	6.90%
# of Negative Sample Correlations	22	17	24	18	1
Percentage (out of 435)	5.06%	3.91%	5.52%	4.14%	0.23%
(B) Estimated Shrinkage Intensity					
Constant Correlation Target					
Bias Corrected	0.4654	0.4249	0.3702	0.3279	0.3269
Bias Ignored	0.4703	0.4288	0.3739	0.3309	0.3291
Zero Correlation Target					
Bias Corrected	0.1957	0.1606	0.1432	0.1152	0.0856
Bias Ignored	0.2048	0.1677	0.1491	0.1202	0.0899
(C) Post-Shrinkage Average Correlation					
Constant Correlation Target					
Bias Corrected	0.3391	0.3259	0.3079	0.3202	0.3564
Bias Ignored	0.3391	0.3259	0.3079	0.3202	0.3564
Zero Correlation Target					
Bias Corrected	0.2727	0.2736	0.2638	0.2833	0.3259
Bias Ignored	0.2696	0.2713	0.2620	0.2817	0.3244

4.2.1 Panel (A): Sample Correlations

Table 1 has three panels. As shown in Panel (A), the average sample correlations ($\overline{r}$) for $T = 36$, 48, 60, 72, and 84 are between 0.3079 and 0.3564, with the lowest and highest averages corresponding to $T = 60$ and 84, respectively. Among the five overlapping sample periods, the range of sample correlations for $T = 36$, from -0.2229 to 0.8957, is the widest; in contrast, the corresponding range for $T = 84$, from -0.0452 to 0.8523, is the narrowest. While the highest sample correlations, which range from 0.8523 to 0.8957, are all between Goldman Sachs (GS) and JPMorgan Chase (JPM), the lowest sample correlations, which range from -0.0452 to -0.2229, are between various other companies.

Both the number of sample correlations exceeding $\sqrt{3}/3$ and the number of negative sample correlations, among the 435 off-diagonal elements in the upper triangle of the 30×30 correlation matrix — which are relevant in determining how bias correction affects the shrinkage results, as explained in Subsection 2.5 — are also included in Panel (A). So are the corresponding figures in percentage terms. Only a small number of the sample correlations (for no more than 7.59% of the 435 elements) is above the $\sqrt{3}/3$ threshold, for any length of the sample period considered.

For $T = 36$ to $T = 72$, there are also some comparable numbers of negative sample correlations. As there are many more sample correlations that are below the $\sqrt{3}/3$ threshold, we can safely expect bias correction to have a net negative effect on the shrinkage intensity, in view of the explanation in Subsection 2.5. Indeed, the shrinkage results for either a constant correlation target or a zero correlation target, as shown in Panel (B), do confirm this negative effect of bias correction.

4.2.2 Panel (B): Estimated Shrinkage Intensity

Panel (B) also shows that, for either shrinkage target with or without bias correction, an increase in the length of the sample period (T) always corresponds to a decrease in the estimated shrinkage intensity ($\widehat{\lambda}$). While the effect of bias correction on $\widehat{\lambda}$ is small, the corresponding effect of the choice of the shrinkage target is far more substantive. With bias corrected, $\widehat{\lambda}$ for a constant correlation target decreases from 0.4654 to 0.3269, as T increases from 36 to 84. For a zero correlation target, the corresponding decrease in $\widehat{\lambda}$ is from 0.1957 to 0.0856 instead.

In the following, we first explain why $\widehat{\lambda}$ decreases with increasing T. To explain why $\widehat{\lambda}$ is higher for a constant correlation target than for a zero correlation target requires a discussion of the results in Panel (C). Thus, the latter explanation will be provided later. To explain why $\widehat{\lambda}$ decreases with increasing T, a key point is that, as T increases, the individual sample correlations tend to be closer to the corresponding true but unknown correlations. Intuitively, as T increases, the need for quality improvements in the individual sample correlations from shrinkage will lessen, thus resulting in progressively lower values of $\widehat{\lambda}$.

Here is a more detailed explanation based on the analytical materials in Sections 2 and 3: For ease of exposition, we start with the case where a zero correlation target is used. As the effect of bias correction on the shrinkage results tends to be marginal, the corresponding detailed explanation is omitted. Under the stationarity assumption of the joint probability distribution of returns, increases in T are expected to cause each $\widehat{Var}(\boldsymbol{W}_{ij})$, for $i = 1, 2, \ldots, n-1$ and $j = i+1, i+2, \ldots, n$, to vary in a less drastic manner, when compared with what such increases will affect the corresponding $\widehat{Var}(r_{ij})$. This is all because of the multiplicative factor $T/(T-1)^2$ that directly connects $\widehat{Var}(\boldsymbol{W}_{ij})$ and $\widehat{Var}(r_{ij})$ in equation (32).

For $T = 36$, 48, 60, 72, and 84, the values of $T/(T-1)^2$ are 0.02939, 0.02173, 0.01724, 0.01428, and 0.01219, respectively. In an ideal scenario where there are no changes in $\widehat{Var}(\boldsymbol{W}_{ij})$ as T increases, the values of $\widehat{Var}(r_{ij})$ at $T = 48$, 60, 72, and 84 will be 73.94%, 58.65%, 48.60%, and 41.49%, respectively, of its value at $T = 36$. In a more realistic scenario, however, there will be changes in $\widehat{Var}(\boldsymbol{W}_{ij})$ as T increases. Even so, a downward trend in $\widehat{Var}(r_{ij})$ can still be expected. This is because the strong attenuation effect of $T/(T-1)^2$ as T increases will prevent a reversal of the downward trend from happening.

Given how the individual values of $\widehat{Var}(r_{ij})$ varies with T in general, the double summation $\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}\widehat{Var}(r_{ij})$ is also expected to exhibit a downward trend as T increases. As long as the stationarity assumption holds, the individual values of r_{ij}^2 are expected to vary, but not in a drastic way, as T increases; so are the double summation $\sum_{i=1}^{n-1}\sum_{j=i+1}^{n} r_{ij}^2$ in the denominator on the right hand side of equation (66). Again, it is the strong attenuation effect of $T/(T-1)^2$ on $\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}\widehat{Var}(r_{ij})$ that determines how the estimated shrinkage intensity $\widehat{\lambda}$ based on equation (66) varies with T.

To explain why $\widehat{\lambda}$ decreases as T increases, let us consider the function $g(x) = x/(x+c)$, where both the variable x and the constant c are positive. The first derivative of $g(x)$, which is $c/(x+c)^2$, is also positive. An implication is that a decrease in x corresponds to a decrease in the function. Equation (66) has the same analytical form as $g(x)$, with x representing $\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}\widehat{Var}(r_{ij})$ and c approximating $\sum_{i=1}^{n-1}\sum_{j=i+1}^{n} r_{ij}^2$. If an increase in T leads to a decrease in $\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}\widehat{Var}(r_{ij})$, it will also lead to a lower $\widehat{\lambda}$, provided that $\sum_{i=1}^{n-1}\sum_{j=i+1}^{n} r_{ij}^2$ varies only moderately without a trend, so that it can be approximated by a positive constant.

We now extend the above explanation to the case where a constant correlation target is used instead. Equation (44), which is used to estimate the corresponding shrinkage intensity, also has the same analytical form as $g(x)$. The difference here is that x represents $\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}\widehat{\alpha}_{ij}$ and c approximates $\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}(\overline{r}-r_{ij})^2$, which under the stationarity assumption is expected to vary only moderately without a trend as T changes. Thus, what needs to be established here is that $\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}\widehat{\alpha}_{ij}$ — which is known to be positive, as shown in Subsection 2.4 — has a downward trend as T increases. The same connection between $\widehat{Var}(\boldsymbol{W}_{ij})$ and $\widehat{Var}(r_{ij})$ in equation (32) also exists between $\widehat{Cov}(\boldsymbol{W}_{ij}, \boldsymbol{W}_{k\ell})$ and $\widehat{Cov}(r_{ij}, r_{k\ell})$ in equation (40). It is still the same dominating multiplicative factor $T/(T-1)^2$ that causes $\widehat{\alpha}_{ij}$ and then $\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}\widehat{\alpha}_{ij}$ to decrease as T increases, thus resulting in a downward trend in the estimated shrinkage intensity.

4.2.3 Panel (C): Post-Shrinkage Average Correlation

Panel (C) shows, under the column for each T, the post-shrinkage average correlation for each of the four cases covered in Panel (B). For a constant correlation target, the post-shrinkage average correlation remains the same as the corresponding average sample correlation, regardless of whether the bias is corrected or ignored. For a zero correlation target, in contrast, the post-shrinkage average correlation is lower.

The result that shrinkage of the sample correlation matrix towards a constant correlation target has no effect on the average correlation is as expected. This is because shrinkage estimation in the context of this paper is about taking a weighted average of the sample correlation matrix and the target matrix. The average of all off-diagonal elements of the sample correlation matrix is $\overline{r}$. So is the average of the corresponding elements in a constant correlation matrix, which serves as the target matrix here. Regardless of the weights that are assigned to the two matrices, the average of all off-diagonal elements of the resulting matrix must also be $\overline{r}$.

The matrix that is used as a zero correlation target for shrinkage estimation is an identity matrix. Taking a weighted average of the sample correlation matrix and an identity matrix, which has all zero off-diagonal elements, will lead to a matrix with uniformly attenuated off-diagonal elements. Thus, the post-shrinkage average correlation will inevitably be lower than the original average of the sample correlations. This is what is shown in the last two rows of Panel (C).

The ranges of post-shrinkage correlations can easily be deduced from the displayed information in Table 1. Given the nature of shrinkage estimation, such ranges will always be narrower than the corresponding ranges of the sample correlations. For example, as indicated earlier, the highest sample correlation, which is between Goldman Sachs (GS) and JPMorgan Chase (JPM), is 0.8957 for $T = 36$. Let us label it as $r_{GS,JPM}$. For a constant correlation target where $\overline{r} = 0.3391$, the estimated shrinkage intensity $\widehat{\lambda}$ with the bias corrected is 0.4654. The corresponding post-shrinkage correlation is $\widehat{\lambda}\overline{r} + (1 - \widehat{\lambda})r_{GS,JPM} = 0.4654 \times 0.3391 + (1 - 0.4654) \times 0.8957 = 0.6367$, which is closer to $\overline{r}$ than $r_{GS,JPM}$ is. Also for $T = 36$, the lowest sample correlation is between DuPont (DD) and Wal-Mart (WMT); it is -0.2229. Let us label it as $r_{DD,WMT}$. For the same shrinkage target with the bias corrected, the corresponding post-shrinkage correlation is $\widehat{\lambda}\overline{r} + (1 - \widehat{\lambda})r_{DD,WMT} = 0.4654 \times 0.3391 + (1 - 0.4654) \times (-0.2229) = 0.0386$, which is closer to $\overline{r}$ than $r_{DD,WMT}$ is. Accordingly, the range of post-shrinkage correlations, 0.5981 ($= 0.6367 - 0.0386$), is much narrower than the original range of sample correlation, 1.1186 [$= 0.8957 - (-0.2229)$].

We now return to Panel (B) and explain why the use of a zero correlation target, instead of a constant correlation target, always leads to a lower estimated shrinkage intensity. Regardless of which shrinkage target is involved, the true but unknown correlation of returns ρ_{ij} between each pair of securities, i and j, is always present in the quadratic loss function. The sample correlation r_{ij}, with or without some minor adjustments, is used for its estimation, depending on whether the bias is corrected or ignored. Either way, given the objective of expected quadratic loss minimization, the post-shrinkage correlations are not supposed to be far away from the corresponding sample correlations. For a zero correlation target, however, the greater the shrinkage intensity, the more attenuated will be the individual correlations, and the further away will be the post-shrinkage correlations from the corresponding sample correlations. Thus, there is a natural tendency to avoid assigning an excessive weight on the shrinkage target; that is, the resulting $\widehat{\lambda}$ will tend to be lower.

5 Concluding Remarks and Suggestions for Instructors

This paper has extended the introduction to shrinkage estimation in Kwan (2011). The extension is on shrinking the sample correlation matrix of returns towards a constant correlation target, as justified in Ledoit and Wolf (2004) and Kwan (2008). The extension is intended to be simpler than the corresponding approach in Kwan (2008), which has also used the same target for shrinkage estimation without relying on any asymptotic properties of the finite samples involved. Shrinkage targets based on other structured correlation matrices, which are more refined analytically, have not been attempted in this paper.

Given the pedagogic objectives of the extension, the analytical materials involved have been presented in considerable detail, as the intended readers of this paper include also students. Special attention has been on reducing the analytical and computational burden of the tasks involved, whenever possible, while still maintaining an analytical emphasis. In particular, by assuming that insights of financial analysts can lead to improvements in the quality of the estimated variances, this paper has been able to bypass the usual statistical issues pertaining to the estimation of the individual correlations via the use of the corresponding sample variances and covariances, as encountered in Kwan (2008).

The idea of shrinkage estimation is very simple; in the context here, it is about taking a weighted average of the sample correlation matrix and the target matrix. The weight that is assigned to the latter matrix is known as the shrinkage intensity. By using a constant correlation target, we seek to achieve a balance between reducing the overall estimation errors and maintaining some existing idiosyncrasies in the individual correlations. However, to achieve such a balance analytically is a tedious task, even for this seemingly simple target. Thus, Excel has played an important pedagogic role in this paper. Central to the Excel illustration in this paper is the user-defined function SHRINK for computing the shrinkage intensity. The function not only can accommodate the zero correlation target in Kwan (2011), but also can allow the user to correct the statistical bias for using the sample correlations as estimators of the corresponding true correlations.

The use of Dow Jones data in this paper is intended to illustrate how the returns of U.S. major stocks are correlated and how shrinkage estimation affects the sample correlation matrix of returns. Students can gain some valuable hands-on experience with shrinkage estimation, by replicating some of the results in Table 1 with the Dow Jones data in one of the two Excel files accompanying this paper. For example, to replicate the results there for a 48-month sample period, all that is required is to lengthen the 36-month sample period there by 12 months and to repeat the same Excel-based computations.

In advanced investment courses that cover portfolio theory in depth, students are taught the importance of having high quality input parameters for implementing portfolio selection models. However, the issue as to how input quality can be improved is either addressed only briefly or considered to be beyond the scope of the courses involved. As a result, although students are made aware of the existence of the issue, they may not be taught how to address it properly in practical settings. The hands-on experience with shrinkage estimation that students gain from using the Dow Jones data will enable them to understand better not only the statistical techniques involved, but also the underlying concepts of portfolio investments. The latter benefit is equally important from a pedagogic perspective, as the reliance on matrix algebra in the coverage of portfolio theory for analytical convenience tends to mask the intuition of the portfolio concepts involved.

There are exchange-traded funds (ETF's) that track the Dow Jones Industrial Average (DJIA). There are also investment strategies based on the Dow Jones stocks, one of which — commonly known as Dogs of the Dow — is to invest in 10 of the Dow Jones stocks with the highest dividend yields. Students learn from mean-variance portfolio theory that, the lower the correlations of returns between securities, the greater are the benefits from portfolio diversification. Given the high correlations between some Dow Jones stocks as shown in the above-mentioned Excel file, it may seem counter-intuitive to students who have learned portfolio concepts that DJIA-based ETF's, which contain some highly correlated stocks, are of interest as investment portfolios to many investors.

In addition to discussions of the issue in class, instructors can also assign Excel-based projects for students to investigate whether investing in subsets of the Dow Jones stocks, especially those with high-correlation cases removed, can result in better portfolio performance in terms of risk-return trade-off. The Dow Jones data in the above-mentioned Excel file, augmented to include some former Dow Jones stocks and/or to increase the number of return observations if necessary, are suitable for use in such projects. To incorporate shrinkage estimation into such projects, students can compare investment results based on pre-shrinkage and post-shrinkage correlation matrices for various subsets of the Dow Jones stocks.

Finally, some remarks on computational times are in order. It takes much more computer time to produce the shrinkage results for a constant correlation target than for a zero correlation target. This is because, in VBA coding for the former case, five nested loops are required in the computation of the shrinkage intensity and, for the latter cases, only three nested loops are required. Suppose that, in both cases, the number of securities is n and the number of return observations is T. To go through five nested loops requires $[n(n-1)/2]^2T$ steps of computations; to go through three nested loops, only $[n(n-1)/2]T$ steps are involved. For $n = 30$ and $T = 36$, for example, the corresponding numbers are $6,812,100$ and $15,660$; for $n = 30$ and $T = 84$, they are $15,894,900$ and $36,540$ instead. In terms of computational times, it takes in the former case 2 minutes 17 seconds and 5 minutes 18 seconds (on a desktop computer at 3.16 GHz and 4.00 GB RAM) for the user-defined function SHRINK to compute the shrinkage intensity for $T = 36$ and $T = 84$, respectively. In contrast, in the latter case, the corresponding task can be completed almost instantaneously for either sample period.

References

Chan, L.K.C., Karceski, J., and Lakonishok, J., (1999). On Portfolio Optimization: Forecasting Covariances and Choosing the Risk Model. *Review of Financial Studies*, 12(5), 937-974.

Elton, E.J., and Gruber, M.J., (1973). Estimating the Dependence Structure of Share Prices — Implications for Portfolio Selection. *Journal of Finance*, 28(5), 1203-1232.

Elton, E.J., Gruber, M.J., and Padberg, M.W., (1976). Simple Criteria for Optimal Portfolio Selection. *Journal of Finance*, 31(5), 1341-1357.

Elton, E.J., Gruber, M.J., and Urich, T.J., (1978). Are Betas Best? *Journal of Finance*, 33(5), 1375-1384.

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(4), 236-244.

Kwan, C.C.Y., (2011). An Introduction to Shrinkage Estimation of the Covariance Matrix: A Pedagogic Illustration, *Spreadsheets in Education*,4(3), Article 6.

Ledoit, O., and Wolf, M., (2003). Improved Estimation of the Covariance Matrix of Stock Returns with an Application to Portfolio Selection. *Journal of Empirical Finance*, 10(5), 603-621.

Ledoit, O., and Wolf, M., (2004). Honey, I Shrunk the Sample Covariance Matrix. *Journal of Portfolio Management*, 30(4), 110-119.

Olkin, I., Pratt, J.W., (1958). Unbiased Estimation of Certain Correlation Coefficients. *Annals of Mathematical Statistics*, 29(1), 201–211.

Schäfer, J., and Strimmer, K., (2005). A Shrinkage Approach to Large-Scale Covariance Matrix Estimation and Implications for Functional Genomics. *Statistical Applications in Genetics and Molecular Biology*, 4(1), Article 32.

Zimmerman, D.W., Zumbo, B.D., Williams, R.H., (2003). Bias in Estimation and Hypothesis Testing of Correlation. *Psicológica*, 24(1), 133–158.