Spreadsheets in Education (eJSiE)

Volume 4 | Issue 3 Article 6

8-12-2011

An Introduction to Shrinkage Estimation of the Covariance Matrix: A Pedagogic Illustration

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

An Introduction to Shrinkage Estimation of the Covariance Matrix: A Pedagogic Illustration

Abstract
Shrinkage estimation of the covariance matrix of asset returns was introduced to the finance profession several years ago. Since then, the approach has also received considerable attention in various life science studies, as a remedial measure for covariance matrix estimation with insufficient observations of the underlying variables. The approach is about taking a weighted average of the sample covariance matrix and a target matrix of the same dimensions. The objective is to reach a weighted average that is closest to the true covariance matrix according to an intuitively appealing criterion. This paper presents, from a pedagogic perspective, an introduction to shrinkage estimation and uses Microsoft Excel™ for its illustration. Further, some related pedagogic issues are discussed and, to enhance the learning experience of students on the topic, some Excel-based exercises are suggested.

Keywords
Shrinkage estimation, sample covariance matrix

Cover Page Footnote
The author wishes to thank Y. Feng and two anonymous reviewers for helpful comments and suggestions. He also wishes to thank K. Brewer for advice on technical issues pertaining to Visual Basic for Applications (VBA).

An Introduction to Shrinkage Estimation of the Covariance Matrix: A Pedagogic Illustration[1]

Clarence C.Y. Kwan
DeGroote School of Business, McMaster University
(kwanc@mcmaster.ca)

Revised: August 2011

[1]The author wishes to thank Y. Feng and two anonymous reviewers for helpful comments and suggestions. He also wishes to thank K. Brewer for advice on technical issues pertaining to Visual Basic for Applications (VBA).

Abstract

Shrinkage estimation of the covariance matrix of asset returns was introduced to the finance profession several years ago. Since then, the approach has also received considerable attention in various life science studies, as a remedial measure for covariance matrix estimation with insufficient observations of the underlying variables. The approach is about taking a weighted average of the sample covariance matrix and a target matrix of the same dimensions. The objective is to reach a weighted average that is closest to the true covariance matrix according to an intuitively appealing criterion. This paper presents, from a pedagogic perspective, an introduction to shrinkage estimation and uses Microsoft Excel^{TM} for its illustration. Further, some related pedagogic issues are discussed and, to enhance the learning experience of students on the topic, some Excel-based exercises are suggested.

Keywords: Shrinkage estimation, sample covariance matrix.

An Introduction to Shrinkage Estimation of the Covariance Matrix: A Pedagogic Illustration

1 Introduction

For a given set of random variables, the corresponding covariance matrix is a symmetric matrix with its diagonal and off-diagonal elements being the individual variances and all pairwise covariances, respectively. When each variable involved is normalized to have a unit variance, such a matrix reduces to a correlation matrix. The usefulness of these matrices for multivariate investigations is well known across various academic disciplines. In finance, for example, the covariance matrix of asset returns is part of the input parameters for portfolio analysis to assist investment decisions. Likewise, in life science research, when statistical techniques are applied to analyze multivariate data from experiments, statistical inference can be made with information deduced from some corresponding covariance and correlation matrices.

As the true values of the individual matrix elements are unknown, they have to be estimated from samples of empirical or experimental data. To facilitate portfolio analysis in practice, for example, historical asset returns are commonly used to estimate the covariance matrix. Under the stationarity assumption of the probability distributions of asset returns, sample estimates of the covariance matrix are straightforward. The computations involved can easily be performed by using some of the built-in matrix functions in Microsoft Excel^{TM}.

As explained in Kwan (2010), for a covariance matrix of asset returns to be acceptable for portfolio analysis, it must be positive definite. A positive definite matrix is always invertible, but not vice versa. In the context of portfolio investment, with variance being a measure of risk, an invertible covariance matrix is required for portfolio selection models to reach portfolio allocation results. A positive definite covariance matrix always provides strictly positive variances of portfolio returns, regardless how investment funds are allocated among the assets considered. This feature ensures not only the presence of portfolio risk, but also the uniqueness of efficient portfolio allocation results as intended.

In case that the covariance matrix is estimated with insufficient observations, it will not be positive definite. Thus, for example, to estimate a 100×100 covariance matrix of monthly asset returns requires more than 100 monthly return observations, just to ensure that the number of observed returns exceed the number of the unknown matrix elements. To ensure further that estimation errors be small enough for the sample covariance matrix to be acceptable as part of the input parameters for portfolio analysis, additional observations are required. However, the validity of the stationarity assumption of asset return distributions will become a concern if an overly long sample period is used for the estimation.

The reliance on higher frequency observations over the same sample period, such as the use of weekly or daily asset returns, though effective in easing one's concerns about non-stationary asset return distributions, does have its drawbacks. Higher frequency observations tend to be noisier, thus affecting the quality of the estimated covariance matrix. For covariance matrix estimation that accounts for time-varying volatility of asset returns, some multivariate time series models are available.[1] However, such models are very complicated. Even the simplest versions of such models are well beyond the scope of any investment courses for business students. A challenge for instructors of investment courses that cover empirical issues in covariance matrix estimation, therefore, is whether it is feasible to go beyond sample estimates without using statistical methods that are burdensome for students.

Before seriously contemplating the above challenge, notice that, in life sciences and related fields where multivariate analysis is performed on experimental data, concerns about the adequacy of the sample covariance and correlation matrices are from a very different perspective. As indicated in Dobbin and Simon (2007) and Yao et al. (2008), biological studies often employ only small numbers of observations, because either there are few available experimental data to use or data collection is under budgetary or time constraints. Covariance matrices estimated with insufficient observations are known to be problematic. As explained by Schäfer and Strimmer (2005), in a study of gene association networks using graphical Gaussian models (GGM's), the partial correlation of any two genes, which measures their degree of association after removing the effect of other genes on them, can be deduced from the inverse of a covariance matrix. An implicit requirement is that the estimated covariance matrix be invertible.

In a recent cancer research study by Beerenwinkel et al. (2007), for example, pairwise partial correlations of 78 cancer genes (for a total of $78 \times 77 \div 2 = 3,003$ partial correlations) are based on only 35 available tumor samples. Obviously, the corresponding sample covariance matrix is not positive definite and thus is not invertible. The relatively low number of experimental observations has made it necessary for these authors to rely on some remedial measures. The remedial measures as reported by Schäfer and Strimmer (2005), Beerenwinkel et al. (2007), and Yao et al. (2008), as well as some other researchers facing similar challenges, are based on an innovative approach to

[1] See, for example, the survey articles by Bauwens, Laurent, and Rombouts (2006) and Silvennoinen and Teräsvirta (2009) for descriptions of various multivariate time series models for covariance matrix estimation.

estimate the covariance matrix, called shrinkage estimation, as introduced to the finance profession by Ledoit and Wolf (2003, 2004a, 2004b).

In investment settings, a weighted average of the sample covariance matrix of asset returns and a structured matrix of the same dimensions is viewed as shrinkage of the sample covariance matrix towards a target matrix. The shrinkage intensity is the weight that the target receives. In the Ledoit-Wolf studies, the alternative targets considered include an identity matrix, a covariance matrix based on the single index model (where the return of each asset is characterized as being linearly dependent on the return of a market index), and a covariance matrix based on the constant correlation model (where the correlation of returns between any two different assets is characterized as being the same). In each case, the corresponding optimal shrinkage intensity has been derived by minimizing an intuitively appealing quadratic loss function.[2]

For analytical convenience, the Ledoit-Wolf studies have relied on some asymptotic properties of the asset return data in model formulation. Although stationarity of asset return distributions is implicitly assumed, the corresponding analytical results are still based on observations of relatively long time series. Thus, the Ledoit-Wolf shrinkage approach in its original form is not intended to be a remedial measure for insufficient observations. To accommodate each life science case where the number of observations is far fewer than the number of variables involved, Schäfer and Strimmer (2005) have extended the Ledoit-Wolf approach to finite sample settings. The Schäfer-Strimmer study has listed six potential shrinkage targets for covariance and correlation matrices. They include an identity matrix, a covariance matrix based on the constant correlation model, and a diagonal covariance matrix with the individual sample variances being its diagonal elements, as well as three other cases related to these matrices.

The emphasis of the Schäfer-Strimmer shrinkage approach is a special case where the target is a diagonal matrix. Shrinkage estimation of the covariance matrix for this case is relatively simple, from both analytical and computational perspectives. When all variables under consideration are normalized to have unit variances, the same shrinkage approach becomes that for the correlation matrix instead. Analytical complications in the latter case are caused by the fact that normalization of individual variables cannot be based on the true but unknown variances and thus has to be based instead on the sample variances, which inevitably have estimation errors. In order to retain the analytical features pertaining to shrinkage estimation of the covariance matrix, the Schäfer-

[2] The word *optimal* used throughout this paper is in an *ex ante* context. Whether an analytically determined shrinkage intensity, based on in-sample data, is *ex post* superior is an empirical issue that can only be assessed with out-of-sample data.

Strimmer study has assumed away any estimation errors in the variances when the same approach is applied directly to a set of normalized data.

Opgen-Rhein and Strimmer (2006a, 2006b, 2007a, 2007b) have extended the Schäfer-Strimmer approach by introducing a new statistic for gene ranking and by estimating gene association networks in dynamic settings to account for the time path of the data. In view of the analytical simplicity of the Schäfer-Strimmer version of the Ledoit-Wolf shrinkage approach, where the shrinkage target is a diagonal matrix, it has been directly applied to various other settings in life sciences and related fields. Besides the studies by Beerenwinkel et al. (2007) and Yao et al. (2008) as referenced earlier, the following are further examples:

With shrinkage applied to the covariance matrix for improving the GGM's, Werhli, Grzegorczyk, and Husmeier (2006) have reported a favorable comparison of the shrinkage GGM approach over a competing approach, called relevance networks, in terms of the accuracy in reconstructing gene regulatory networks. In a study of information-based functional brain mapping, Kriegeskorte, Goebel, and Bandettini (2006) have reported that shrinkage estimation with a diagonal target improves the stability of the sample covariance matrix. Dabney and Storey (2007), also with the covariance matrix estimated with shrinkage, have proposed an improved centroid classifier for high-dimensional data and have demonstrated that the new classifier enhances the prediction accuracy for both simulated and actual microarray data. More recently, in a study of gene association networks, Tenenhaus et al. (2010) have used the shrinkage GGM approach as one of the major benchmarks to assess partial correlation networks that are based on partial least squares regression.

The research influence of the Ledoit-Wolf shrinkage approach, however, is not confined to life science fields. The approach as reported in Ledoit's working papers well before its journal publications already attracted attention of other finance researchers. It was among the approaches for risk reduction in large investment portfolios adopted by Jaggannathan and Ma (2003). More recently, Disatnik and Benninga (2007) have compared empirically various shrinkage estimators (including portfolios of estimators) of high-dimensional covariance matrices based on monthly stock return data. In an analytical setting, where shrinkage estimation of covariance and correlation matrices are with targets based on the average correlation of asset returns, Kwan (2008) has accounted for estimation errors in all variances when shrinking the sample correlation matrix, thus implicitly allowing the analytical expression of the Schäfer-Strimmer shrinkage intensity to be refined.

2 A Pedagogic Approach and the Role of Excel in the Illustration of Shrinkage Estimation

In view of the attention that shrinkage estimation has received in the various studies, of particular interest to us educators is whether the topic is now ready for its introduction to the classroom. With the help of Excel tools, this paper shows that it is indeed ready. In order to avoid distractions by analytical complications, this paper has its focus on shrinkage estimation of the covariance matrix, with the target being a diagonal matrix. Specifically, the diagonal elements of the target matrix are the corresponding sample variances of the underlying variables. It is implicit, therefore, that the shrinkage approach here pertains only to the covariances. Readers who are interested in the analytical details of more sophisticated versions of shrinkage estimation can find them directly in Ledoit and Wolf (2003, 2004a, 2004b) and, for extensions to finite sample settings, in Schäfer and Strimmer (2005) and Kwan (2008).

This paper utilizes Excel tools in various ways to help students understand shrinkage estimation better. Before formally introducing optimal shrinkage estimation, we establish in Section 3 that a weighted average of a sample covariance matrix and a structured target, such as a diagonal matrix, is always positive definite. To avoid digressions, analytical support for some materials in Section 3 is provided in Appendix A. An Excel example, with a scroll bar for manually making weight changes, illustrates that, even for a covariance matrix estimated with insufficient observations, a non-zero weight for the target matrix will always result in a positive definite weighted average. The positive definiteness of the resulting matrix is confirmed by the consistently positive sign of its leading principal minors (that is, the determinants of its leading principal submatrices). The Excel function MDETERM, which is for computing the determinants of matrices, is useful for the illustration. As any effects on the leading principal minors due to weight changes are immediately displayed in the worksheet, the idea of shrinkage estimation will become less abstract and more intuitive to students.

Optimal shrinkage is considered next in Section 4, with analytical support provided in Appendix B. As mentioned briefly earlier, the idea is based on minimization of a quadratic loss function. Here, we take a weighted average of the sample covariance matrix, which represents a noisy but unbiased estimate of the true covariance matrix, and a target matrix, which is biased. Loss is defined as the expected value of the sum of all squared deviations of the resulting matrix elements from the corresponding true values. We search for a weighted average that corresponds to the lowest loss. As there is only one unknown parameter in the quadratic loss function, which is the shrinkage intensity (i.e., the weight that the target matrix receives), its optimal value can easily be found.

We then continue with the same Excel example to illustrate the computational task in shrinkage estimation. In order to accommodate students with different prior experience with Excel, we use alternative ways to compute the optimal shrinkage intensity. The most intuitive approach, which is cumbersome for higher-dimensional cases, is to perform a representative computation for a cell in an Excel worksheet and to use copy-and-paste to duplicate the cell formula elsewhere repeatedly, with minor revisions when required. Students who have at least some rudimentary knowledge of Visual Basic for Applications (VBA) can recognize that the use of a simple user-defined function is more efficient for performing the same computations. As shown in Appendix D, the VBA program here uses three nested “for next” loops for the intended task. For convenience to users, the function has only two arguments; they are for cell references of the required data for the computations involved. Once defined, the function works just like any built-in Excel functions. Further, to enhance the user-friendliness of the function, a simple Excel Sub procedure that calls the function directly is also included. The procedure uses input boxes for entering cell references of the required data for the computations and the intended location to show the final result.

Section 5 addresses some pedagogic issues and suggests some exercises for students. To enhance the learning experience of students, we consider it important for instructors to assign some discipline-based Excel exercises on shrinkage estimation. Such exercises not only will provide students with some hands-on experience with the specific technique, but also will enable them to explore various estimation issues pertaining to the covariance matrix. In addition, some simulation-based exercises are also suggested, with analytical support provided in Appendix C.

As this paper is not meant to provide quick recipes for students to perform shrinkage estimation, it pays special attention to the underlying concepts. It also provides an alternative formulation of the same approach that shrinks the sample correlation matrix towards an identity matrix. To allow students to attempt shrinkage estimation of the correlation matrix, a simplified version that has been used in the various life science studies is suggested. The practical relevance of such an approach, as well as the attendant analytical complications of a more sophisticated version, will also be addressed in Section 5. Finally, Section 6 provides some concluding remarks.

Before proceeding to Section 3, notice that, although this paper is primarily a pedagogic paper, with Excel tools utilized to illustrate shrinkage estimation that has been established elsewhere, it still has a research contribution. As it will become clear in Section 4, the contribution pertains to the implementation of the shrinkage approach. Specifically, by drawing on a well-known statistical relationship between the variance of a random variable and the expected value of the square of the same variable, this paper is able to remove the upward bias in the estimated shrinkage intensity, pertaining to finite samples, which still exists in the literature of shrinkage estimation.

3 The Sample Covariance Matrix and its Shrinkage Towards a Diagonal Target

Consider a set of n random variables, labeled as $\widetilde{R}_1, \widetilde{R}_2, \ldots, \widetilde{R}_n$. For each variable $\widetilde{R}_i$, where $i = 1, 2, \ldots, n$, we have T observations, labeled as $R_{i1}, R_{i2}, \ldots, R_{iT}$. Each observation t actually consists of the set of observations of $R_{1t}, R_{2t}, \ldots, R_{nt}$, for $t = 1, 2, \ldots, T$. Thus, the set of observations for these random variables can be captured by an $n \times T$ matrix with each element (i, t) being R_{it}, for $i = 1, 2, \ldots, n$ and $t = 1, 2, \ldots, T$. The sample variance of variable i and the sample covariance of variables i and j are

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

and

$$s_{ij} = \frac{1}{T-1} \sum\nolimits_{t=1}^{T} \left(R_{it} - \overline{R}_i\right)\left(R_{jt} - \overline{R}_j\right), \tag{2}$$

respectively, where $\overline{R}_i = \frac{1}{T}\sum_{t=1}^{T} R_{it}$ and $\overline{R}_j = \frac{1}{T}\sum_{t=1}^{T} R_{jt}$ are the corresponding sample means.[3] Notice that the sample covariance s_{ii} is the same as the sample variance s_i^2. The $n \times n$ matrix, where each element (i, j) being s_{ij}, for $i, j = 1, 2, \ldots, n$, is the sample covariance matrix, labeled here as $\widehat{\mathbf{V}}$. Notice also that $\widehat{\mathbf{V}}$ is symmetric, with $s_{ij} = s_{ji}$, for $i, j = 1, 2, \ldots, n$.

3.1 Covariance Matrix Estimation with Insufficient Observations

For the sample covariance matrix $\widehat{\mathbf{V}}$ to be positive semidefinite, we must have $\boldsymbol{x}'\widehat{\mathbf{V}}\boldsymbol{x} \geq 0$, for any n-element column vector $\boldsymbol{x}$, where the prime indicates matrix transposition. For $\widehat{\mathbf{V}}$ to be also positive definite, $\boldsymbol{x}'\widehat{\mathbf{V}}\boldsymbol{x}$ must be strictly positive for any $\boldsymbol{x}$ with at least one non-zero element. We show in Appendix A that $\widehat{\mathbf{V}}$ is always positive semidefinite. For $\widehat{\mathbf{V}}$ to be positive definite, some conditions must be satisfied. As shown pedagogically in Kwan (2010), to be positive definite, the sample covariance matrix $\widehat{\mathbf{V}}$ must have a positive determinant. We also show in Appendix A that, if $\widehat{\mathbf{V}}$ is estimated with insufficient observations (that is, with $T \leq n$), its determinant is always zero. If so, it is not positive definite.

[3] Here and in what follows, we have assumed that students are already familiar with summation signs and basic matrix operations. For students with inadequate algebraic skills, the materials in this section are best introduced after they have acquired some hands-on experience with Excel functions pertaining to matrix operations.

Notice that the sample covariance matrix $\widehat{\boldsymbol{V}}$ is not always invertible even if it is estimated with sufficient observations. To ensure its invertibility, the following conditions must hold: First, no $\widetilde{R}_i$ can be a constant, as this situation will result in both row i and column i of $\widehat{\boldsymbol{V}}$ being zeros. Second, no $\widetilde{R}_i$ can be replicated by a linear combination of any of the remaining $n-1$ variables. This replication will result in row (column) i of $\widehat{\boldsymbol{V}}$ being a linear combination of some other rows (columns), thus causing its determinant to be zero.[4]

3.2 A Weighted Average of the Sample Covariance Matrix and a Diagonal Matrix

Suppose that none of $\widetilde{R}$, for $i = 1, 2, \ldots, n$, are constants. This ensures that s_{ii} be positive, for $i = 1, 2, \ldots, n$. Now, let $\widehat{\mathbf{D}}$ be an $n \times n$ diagonal matrix with each element (i, i) being s_{ii}. For any n-element column vector $\boldsymbol{x}$ with at least one non-zero element, the matrix product $\boldsymbol{x}'\widehat{\boldsymbol{D}}\,\boldsymbol{x}$ is always strictly positive. This is because, with x_i being element i of vector $\boldsymbol{x}$, we can write $\boldsymbol{x}'\widehat{\boldsymbol{D}}\,\boldsymbol{x}$ explicitly as $\sum_{i=1}^{n} x_i^2 s_{ii}$, which is strictly positive, as long as at least one of $x_1, x_2, \ldots, x_n$ is different from zero.

The idea of shrinkage estimation of the covariance matrix is to take a weighted average of $\widehat{\boldsymbol{V}}$ and $\widehat{\mathbf{D}}$. With λ being the weight assigned to $\widehat{\mathbf{D}}$, we can write the weighted average as

$$\widehat{\mathbf{C}} = (1 - \lambda)\,\widehat{\boldsymbol{V}} + \lambda\widehat{\mathbf{D}}. \tag{3}$$

We have already established that $\boldsymbol{x}'\widehat{\boldsymbol{V}}\,\boldsymbol{x} \geq 0$ and $\boldsymbol{x}'\widehat{\boldsymbol{D}}\,\boldsymbol{x} > 0$, for any n-element column vector $\boldsymbol{x}$ with at least one non-zero element. Therefore, for $0 < \lambda < 1$, the weighted average $\widehat{\mathbf{C}}$ is positive definite, regardless of whether $\widehat{\boldsymbol{V}}$ is estimated with insufficient observations, because

$$\boldsymbol{x}'\widehat{\boldsymbol{C}}\;\boldsymbol{x} = (1 - \lambda)\boldsymbol{x}'\widehat{\boldsymbol{V}}\,\boldsymbol{x} + \lambda\boldsymbol{x}'\widehat{\boldsymbol{D}}\,\boldsymbol{x} > 0. \tag{4}$$

Notice that the case of $\lambda = 0$ is where no shrinkage is applied to the sample covariance matrix. This case retains the original $\widehat{\boldsymbol{V}}$ as an estimate for the covariance matrix, together with all analytical problems that $\widehat{\boldsymbol{V}}$ may carry with it. In contrast, the case of $\lambda = 1$, which indicates instead complete shrinkage of all pairwise covariances, simply ignores the existence of any covariances of the random

[4] In the context of portfolio investment, for example, a constant $\widetilde{R}_i$ in the first situation denotes the presence of a risk-free asset. With the sample variance and covariances of returns pertaining to asset i indicated by $s_{ij} = 0$, for $j = 1, 2, \ldots, n$, the determinant of $\widehat{\boldsymbol{V}}$ is inevitably zero. The second situation, with $\widetilde{R}_i$ being equivalent to a linear combination of the random returns of some other assets under consideration, is where asset i is a portfolio of such assets. This situation also includes the special case where $\widetilde{R}_i = a + b\widetilde{R}_j$, with a and $b(\neq 0)$ being parameters. In such a case, as the random return of asset i is perfectly correlated with the random return of asset j, another asset under portfolio consideration, the sample covariance matrix containing both assets i and j will not have a full rank and thus is not invertible.

variables considered. Cases where $\lambda < 0$ or $\lambda > 1$ are meaningless, from the perspective of shrinkage estimation. Therefore, the shrinkage intensity that λ represents is intended to be for $0 < \lambda < 1$. It will be shown in Section 4 that, under the formulation of quadratic loss minimization, the optimal shrinkage intensity will always correspond to $0 < \lambda < 1$.

3.3 An Excel Example

To illustrate covariance matrix estimation, let us consider a simple Excel example where there are seven random variables ($n = 7$) and six observations ($T = 6$). The example is shown in Figure 1. Although the number of observations is obviously much too low for a meaningful estimation, our purpose here is to illustrate first how insufficient observations, with $T < n$, affect the positive definiteness of the sample covariance matrix. We then illustrate in the same Excel example that a weighted average of the sample covariance matrix and a diagonal matrix (with its diagonal elements containing the corresponding sample variances) is always positive definite.

The 6×7 blocks of cells, B3:H8, show the set of observations, with each column there containing the observations for each variable. With the individual sample means shown in B10:H10, a set of mean-removed observations is then generated, as shown in B14:H19. Specifically, the formula for cell B10, which is =AVERAGE(B3:B8), is pasted to cells B10:H10; likewise, the formula for cell B14, which is =B3-B$10, is pasted to cells B14:H19. For the remainder of the worksheet, copy-and-paste operations are also used to generate analogous cell formulas. The individual cell formulas can be found in the supplementary Excel file (shrink.xls).

The 7×7 sample covariance matrix, as shown in cells B22:H28, is generated for computational convenience according to equation (A3) in Appendix A. To do so, we combine Excel functions for matrix transposition (TRANSPOSE) and multiplication (MMULT), as well as for counting numbers (COUNT), which is {=MMULT(TRANSPOSE(B14:H19),B14:H19)/(COUNT(B14:B19)-1)}. Notice that, to perform MMULT, as well as other matrix operations, the destination cells must be selected in advance. To enter the function, the "Shift+Ctrl+Enter" keys are pressed simultaneously. As expected, the sample covariance matrix thus generated is symmetric.

Next, we verify whether the sample covariance matrix is positive definite by checking the signs of all its leading principal minors. By definition, the i-th leading principal minor of an $n \times n$ matrix is the determinant of the submatrix containing its first i rows and its first i columns, for $i = 1, 2, \ldots, n$. Thus, there are seven leading principal minors in a 7×7 sample covariance matrix. As illustrated in Kwan (2010), all leading principal minors of a positive definite matrix are positive, and a positive definite matrix has all positive leading principal minors.

	A	B	C	D	E	F	G	H	I	J
1										
2	Obs\Var	1	2	3	4	5	6	7		
3	1	10	12	9	-2	17	8	12		
4	2	-9	-11	2	-5	-7	2	-2		
5	3	16	5	8	5	18	8	9		
6	4	6	-3	6	-13	1	4	2		
7	5	1	4	-9	5	8	-16	-1		
8	6	12	-1	2	22	11	6	10		
9										
10	Mean	6	1	3	2	8	2	5		
11										
12	Mean Removed									
13	Obs\Var	1	2	3	4	5	6	7		
14	1	4	11	6	-4	9	6	7		
15	2	-15	-12	-1	-7	-15	0	-7		
16	3	10	4	5	3	10	6	4		
17	4	0	-4	3	-15	-7	2	-3		
18	5	-5	3	-12	3	0	-18	-6		
19	6	6	-2	-1	20	3	4	5		
20										
21	Cov Mat	1	2	3	4	5	6	7		Lead Pr Min
22	1	80.4	47.4	28.6	44.8	75.8	39.6	46.6		80.4
23	2	47.4	62.0	10.4	16.2	68.2	4.0	32.2		2738.04
24	3	28.6	10.4	43.2	-20.6	19.0	56.8	25.4		87071.056
25	4	44.8	16.2	-20.6	141.6	52.8	-2.0	32.0		5836124.52
26	5	75.8	68.2	19.0	52.8	92.8	22.4	48.8		11958160.4
27	6	39.6	4.0	56.8	-2.0	22.4	83.2	37.6		-1.74E-08
28	7	46.6	32.2	25.4	32.0	48.8	37.6	36.8		2.3747E-22
29										
30	Scroll Bar	2000	(from 0 to 10000)							
31	Shrink Int	0.200								
32										
33	Covariance Matrix after Shrinkage									
34		1	2	3	4	5	6	7		Lead Pr Min
35	1	80.40	37.92	22.88	35.84	60.64	31.68	37.28		80.4
36	2	37.92	62.00	8.32	12.96	54.56	3.20	25.76		3546.8736
37	3	22.88	8.32	43.20	-16.48	15.20	45.44	20.32		129639.83
38	4	35.84	12.96	-16.48	141.60	42.24	-1.60	25.60		13673532.3
39	5	60.64	54.56	15.20	42.24	92.80	17.92	39.04		396669294
40	6	31.68	3.20	45.44	-1.60	17.92	83.20	30.08		1.2585E+10
41	7	37.28	25.76	20.32	25.60	39.04	30.08	36.80		1.3575E+11
42										
43										
44										

Figure 1 An Excel Example Illustrating Weighted Averages of the Sample Covariance Matrix and a Diagonal Matrix.

As Excel has a function (MDETERM) for computing the determinant, it is easy to find all leading principal minors of a given matrix. A simple way is to use cut-and-paste operations for the task. With the formula for cell J22, which is =MDETERM($B$22:B22), first pasted to cells K23, L24, M25, N26, O27, and P28 diagonally, we can subsequently move these cells back to column J to allow J22:J28 to contain all seven leading principal minors. Not surprisingly, the first five leading principal minors are all positive, indicating that the 5×5 sample covariance matrix based on the first five random variables ($n = 5$) has been estimated with sufficient observations ($T = 6$). When the remaining variables are added successively, we expect the determinants of the corresponding 6×6 and 7×7 sample covariance matrices to be zeros. However, due to rounding errors in the computations, two small non-zero values are reached instead. As the product of the seven sample variances in the example is about 8.66×10^{12}, the last two leading minors, which are -1.74×10^{-8} and 2.37×10^{-22}, are indeed very small in magnitude.

To illustrate the effect of shrinkage, we insert a scroll bar to the worksheet from the Insert tab of the Developer menu. With the scroll bar in place, we can adjust the shrinkage intensity manually and observe the corresponding changes to the estimated covariance matrix and the seven leading principal minors. Figure 1 shows in cell B31 the case where $\lambda = 0.200$, indicating that a 20.0% weight is assigned to the diagonal matrix. The formula for cell B35, which is =IF(COUNT($B$22:B$22)=COUNT($B$22:$B22),B22,(1-$B$31)*B22), is copied to cells B35:H41. Notice that, as expected, the magnitudes of all covariances have been attenuated.

The seven leading principal minors of the covariance matrix after shrinkage are computed in the same manner as those in cells J22:J28. All of them are now positive. Although not shown in Figure 1, we have confirmed that, for any $0 < \lambda < 1$, the corresponding leading principal minors are all positive, illustrating that shrinkage is effective as a remedial measure for insufficient observations in sample estimation of the covariance matrix.

4 Shrinkage Estimation of the Covariance Matrix

For an $n \times n$ covariance matrix, there are n variances and $n(n-1)$ covariances in total. To introduce the idea of optimal shrinkage, let us start with a simple case where shrinkage is confined to only one of these covariances. Suppose that the sample covariance in question is s_{ij}, where $j \neq i$, and that the corresponding true covariance is σ_{ij}. Suppose also that a weighted average is taken between zero and s_{ij} with weights being λ and $1 - \lambda$, respectively.

The squared deviation of the weighted average from σ_{ij}, which is $[(1-\lambda)s_{ij} - \sigma_{ij}]^2$, can be viewed as a loss. With s_{ij} being a random variable, we are interested in finding a particular λ that provides the lowest expected value of the squared deviation, labeled as $E\{[(1-\lambda)s_{ij}-\sigma_{ij}]^2\}$. Here, E is the expected value operator, with $E(\cdot)$ indicating the expected value of the random variable $(\cdot)$ in question. Notice that the reliance on quadratic loss minimization is quite common because of its analytical convenience. A well-known example is linear regression, where the best fit according to the ordinary-least-squares approach is that the sum of squared deviations of the observations from the fitted line is minimized.

The same idea can be extended to account for all individual covariances. As the covariance matrix is symmetric, we only have to consider the $n(n-1)/2$ covariances in its upper triangle, where $j > i$ (or, equivalently, in its lower triangle, where $j < i$). Analytically, we seek a common weighting factor λ that minimizes the expected value of the sum of squared deviations, with each being $[(1-\lambda)s_{ij}-\sigma_{ij}]^2$, for $i = 1, 2, \ldots, n-1$ and $j = i+1, i+2, \ldots, n$. The loss function under this formulation, $E\left\{\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}[(1-\lambda)s_{ij}-\sigma_{ij}]^2\right\}$, can be expressed simply as $E\left\{\sum_{j>i}[(1-\lambda)s_{ij}-\sigma_{ij}]^2\right\}$, with an implicit understanding that $\sum_{j>i}$ stands for a double summation to account for all $n(n-1)/2$ cases of $j > i$.

As shown in Appendix B, the optimal shrinkage intensity based on minimization of the loss function is

$$\lambda = \frac{\sum_{j>i} Var(s_{ij})}{\sum_{j>i}[Var(s_{ij})+\sigma_{ij}^2]}. \tag{5}$$

Notice that, the variance $Var(\cdot)$ of any random variable $(\cdot)$, defined as $E\{[(\cdot)-E(\cdot)]^2\}$, can also be written as $E[(\cdot)^2]-[E(\cdot)]^2$. Then, with $Var(s_{ij}) = E(s_{ij}^2)-[E(s_{ij})]^2$ and $E(s_{ij}) = \sigma_{ij}$, equation (5) is equivalent to

$$\lambda = \frac{\sum_{j>i} Var(s_{ij})}{\sum_{j>i} E(s_{ij}^2)}. \tag{6}$$

To determine the optimal shrinkage intensity λ with either equation (5) or equation (6) requires that $Var(s_{ij})$ and σ_{ij}^2 [or, equivalently, $E(s_{ij}^2)$] for all $j > i$ be estimated. Before addressing the estimation issues, notice that both the numerator and the denominator in the expression of λ in equation (5) are positive. With the denominator being greater, we must have $0 < \lambda < 1$ as intended. This analytical feature ensures positive weights on both the sample covariance matrix and the diagonal target. It also ensures that the resulting covariance matrix be positive definite, as illustrated in Section 3.

4.1 Estimation Issues

The estimation of $Var(s_{ij})$ can follow the same approach as described in Kwan (2009), which draws on Schäfer and Strimmer (2005). The idea is as follows: We first introduce a random variable $\widetilde{w}_{ij}$, with observations being the product of mean-removed $\widetilde{R}_i$ and $\widetilde{R}_j$. That is, the observations of $\widetilde{w}_{ij}$, labeled as w_{ijt}, are $(R_{it} - \overline{R}_i)(R_{jt} - \overline{R}_j)$, for $t = 1, 2, \ldots, T$. The sample mean of $\widetilde{w}_{ij}$ is

$$\overline{w}_{ij} = \frac{1}{T} \sum\nolimits_{t=1}^{T} (R_{it} - \overline{R}_i)(R_{jt} - \overline{R}_j), \tag{7}$$

which, when combined with equation (2), leads to

$$s_{ij} = \frac{T}{T-1} \overline{w}_{ij}. \tag{8}$$

In view of equation (8), the sampling variances of s_{ij} and $\overline{w}_{ij}$, labeled as $\widehat{Var}(s_{ij})$ and $\widehat{Var}(\overline{w}_{ij})$, respectively, are related by

$$\widehat{Var}(s_{ij}) = \frac{T^2}{(T-1)^2} \widehat{Var}(\overline{w}_{ij}). \tag{9}$$

As the distribution of the sample mean of a random variable based on T observations has a sampling variance that is only $1/T$ of the sampling variance of the variable, it follows from equation (9) that[5]

$$\widehat{Var}(s_{ij}) = \frac{T}{(T-1)^2} \widehat{Var}(\widetilde{w}_{ij}) = \frac{T}{(T-1)^3} \sum\nolimits_{t=1}^{T} (w_{ijt} - \overline{w}_{ij})^2. \tag{10}$$

This equation allows each of the variance terms, $Var(s_{ij})$, in equation (5) to be estimated.

In various studies involving the shrinkage of the sample covariance matrix towards a diagonal target, including those based on the Schäfer-Strimmer approach as referenced in Section 1, each $E(s_{ij}^2)$ in equation (6) has been approximated directly by the square of the corresponding point estimate s_{ij}. However, recall that $E(s_{ij}^2) = Var(s_{ij}) + [E(s_{ij})]^2$. Such an approximation, which implicitly assumes the equality of $E(s_{ij}^2)$ and $[E(s_{ij})]^2$, has the effect of understating the denominator in the expression of λ in equation (6). In turn, it has the effect of overstating the optimal shrinkage intensity λ.

To avoid the above bias, this paper stays with equation (5) instead for any subsequent computations. With $E(s_{ij}) = \sigma_{ij}$, the sample covariance s_{ij} provides an unbiased estimate of σ_{ij}. The optimal shrinkage intensity can then be reached by estimating each σ_{ij}^2 in equation (5) with the square of the corresponding s_{ij}. To show the improvement here, let λ^* be the estimated λ according to equation (5), where each σ_{ij}^2 is estimated by s_{ij}^2. Let also $\lambda^{\#}$ be the estimated λ according to equation (6), where each $E(s_{ij}^2)$ is approximated directly by s_{ij}^2. Denoting $\alpha = \sum_{j>i} \widehat{Var}(s_{ij})$ and

[5] See, for example, Kwan (2009) for a pedagogic illustration of this statistical concept.

$\beta = \sum_{j>i} s_{ij}^2$, we can write $\lambda^* = \alpha/(\alpha+\beta)$ and $\lambda^{\#} = \alpha/\beta$. As $1/\lambda^* = 1+\beta/\alpha = 1+\lambda^{\#}$, it follows that $\lambda^{\#} = \lambda^*/(1-\lambda^*) > \lambda^*$.

4.2 An Excel Example

The same Excel example in Figure 1 is continued in Figure 2. We now illustrate how equation (5) can be used to determine the optimal shrinkage intensity. For this task, we first scale the sample covariance matrix in B22:H28 by the factor $(T-1)/T$, which is 5/6, to obtain a 7×7 matrix consisting of $\overline{w}_{ij}$, for $i, j = 1, 2, \ldots, 7$. The resulting matrix is placed in B46:H52. This step requires the cell formula for B46, which is =B22*(COUNT($B$3:$B$8)-1)/COUNT($B$3:$B$8), to be pasted to B46:H52.

We then paste the cell formula for B54, which is =IF(COUNT($B$22:B$22)=COUNT($B$22:$B22),"",B22*B22), to B54:H60, so that the squares of the sample covariances can be summed and stored in A115 later. The idea of using an IF statement here to allow a blank cell is that, as the sample variances are not required for equation (5), there is no need to place their squares in the diagonal cells of the 7×7 block B54:H60. Notice that, although the summations in equation (5) are for $7 \times 6 \div 2$ $(= 21)$ cases where $j > i$ for computational efficiency, symmetry of the covariance matrix allows us to reach the same shrinkage intensity by considering all 42 cases of $j \neq i$ instead. As $w_{ijt} = w_{jit}$, for $i, j = 1, 2, \ldots, 7$ and $t = 1, 2, \ldots, 6$, displaying all 42 cases explicitly will enable us to recognize readily any errors in subsequent cell formulas pertaining to the individual w_{ijt}.

The individual cases of $(w_{ijt} - \overline{w}_{ij})^2$, for $i \neq j$, are stored in B62:H109. For this task, we paste the cell formula for B62, which is =IF(COUNT($B14:B14)=1,"",($B14*B14-B$46)^2), to B62:H67; the cell formula for B69, which is =IF(COUNT($B14: B14)=2,"",($C14*B14-B$47)^2), to B69:H74; the cell formula for B76, which is =IF(COUNT($B14:B14)=3,"",($D14*B14-B$48)^2), to B76:H81; and so on. The use of an IF statement here to allow a blank cell is to omit all cases of $(w_{iit} - \overline{w}_{ii})^2$. Notice that each column of six cells here containing $(w_{ijt} - \overline{w}_{ij})^2$ displays the same corresponding numbers as that containing $(w_{jit} - \overline{w}_{ji})^2$, as intended.

The sum of all cases of $\widehat{Var}(s_{ij})$, which is =SUM(B62:H110)*COUNT(B14:B19)/(COUNT(B14:B19)-1)^3 according to equation (10), is stored in A112. The optimal shrinkage intensity, which is =A112/(A112+A115), as stored in A119. The covariance matrix after shrinkage and its leading principal minors, which are shown in B123:H129 and J123:J129, respectively, can be established in the same manner as those in B35:H41 and J35:J41. As expected, the seven leading principal minors are all positive; the covariance matrix after shrinkage is positive definite, notwithstanding the fact that the sample covariance matrix itself is based on insufficient observations.

	A	B	C	D	E	F	G	H	I	J
45	w ij bar	1	2	3	4	5	6	7		
46	1	67.00	39.50	23.83	37.33	63.17	33.00	38.83		
47	2	39.50	51.67	8.67	13.50	56.83	3.33	26.83		
48	3	23.83	8.67	36.00	-17.17	15.83	47.33	21.17		
49	4	37.33	13.50	-17.17	118.00	44.00	-1.67	26.67		
50	5	63.17	56.83	15.83	44.00	77.33	18.67	40.67		
51	6	33.00	3.33	47.33	-1.67	18.67	69.33	31.33		
52	7	38.83	26.83	21.17	26.67	40.67	31.33	30.67		
53	Sq Cov									
54	1		2246.76	817.96	2007.04	5745.64	1568.16	2171.56		
55	2	2246.76		108.16	262.44	4651.24	16.00	1036.84		
56	3	817.96	108.16		424.36	361.00	3226.24	645.16		
57	4	2007.04	262.44	424.36		2787.84	4.00	1024.00		
58	5	5745.64	4651.24	361.00	2787.84		501.76	2381.44		
59	6	1568.16	16.00	3226.24	4.00	501.76		1413.76		
60	7	2171.56	1036.84	645.16	1024.00	2381.44	1413.76			
61	Square of Mean-Removed w 1jt									
62	1		20.25	0.03	2844.44	738.03	81.00	117.36		
63	2		19740.25	78.03	4578.78	26190.03	1089.00	4378.03		
64	3		0.25	684.69	53.78	1356.69	729.00	1.36		
65	4		1560.25	568.03	1393.78	3990.03	1089.00	1508.03		
66	5		2970.25	1308.03	2738.78	3990.03	3249.00	78.03		
67	6		2652.25	890.03	6833.78	2040.03	81.00	78.03		
68	Square of Mean-Removed w 2jt									
69	1	20.25		3287.111	3306.25	1778.028	3927.111	2516.694		
70	2	19740.25		11.11111	4970.25	15170.03	11.11111	3268.028		
71	3	0.25		128.4444	2.25	283.3611	427.1111	117.3611		
72	4	1560.25		427.1111	2162.25	831.3611	128.4444	220.0278		
73	5	2970.25		1995.111	20.25	3230.028	3287.111	2010.028		
74	6	2652.25		44.44444	2862.25	3948.028	128.4444	1356.694		
75	Square of Mean-Removed w 3jt									
76	1	0.03	3287.11		46.69	1456.69	128.44	434.03		
77	2	78.03	11.11		584.03	0.69	2240.44	200.69		
78	3	684.69	128.44		1034.69	1167.36	300.44	1.36		
79	4	568.03	427.11		774.69	1356.69	1708.44	910.03		
80	5	1308.03	1995.11		354.69	250.69	28448.44	2584.03		
81	6	890.03	44.44		8.03	354.69	2635.11	684.69		
82	Square of Mean-Removed w 4jt									
83	1	2844.444	3306.25	46.69444		6400	498.7778	2988.444		
84	2	4578.778	4970.25	584.0278		3721	2.777778	498.7778		
85	3	53.77778	2.25	1034.694		196	386.7778	215.1111		
86	4	1393.778	2162.25	774.6944		3721	802.7778	336.1111		
87	5	2738.778	20.25	354.6944		1936	2738.778	1995.111		
88	6	6833.778	2862.25	8.027778		256	6669.444	5377.778		

Figure 2 An Excel Example Illustrating the Determination of the Optimal Shrinkage Intensity.

	A	B	C	D	E	F	G	H	I	J
89	Square of Mean-Removed w 5jt									
90	1	738.03	1778.03	1456.69	6400.00		1248.44	498.78		
91	2	26190.03	15170.03	0.69	3721.00		348.44	4138.78		
92	3	1356.69	283.36	1167.36	196.00		1708.44	0.44		
93	4	3990.03	831.36	1356.69	3721.00		1067.11	386.78		
94	5	3990.03	3230.03	250.69	1936.00		348.44	1653.78		
95	6	2040.03	3948.03	354.69	256.00		44.44	658.78		
96	Square of Mean-Removed w 6jt									
97	1	81.00	3927.11	128.44	498.78	1248.44		113.78		
98	2	1089.00	11.11	2240.44	2.78	348.44		981.78		
99	3	729.00	427.11	300.44	386.78	1708.44		53.78		
100	4	1089.00	128.44	1708.44	802.78	1067.11		1393.78		
101	5	3249.00	3287.11	28448.44	2738.78	348.44		5877.78		
102	6	81.00	128.44	2635.11	6669.44	44.44		128.44		
103	Square of Mean-Removed w 7jt									
104	1	117.36	2516.69	434.03	2988.44	498.78	113.78			
105	2	4378.03	3268.03	200.69	498.78	4138.78	981.78			
106	3	1.36	117.36	1.36	215.11	0.44	53.78			
107	4	1508.03	220.03	910.03	336.11	386.78	1393.78			
108	5	78.03	2010.03	2584.03	1995.11	1653.78	5877.78			
109	6	78.03	1356.69	684.69	5377.78	658.78	128.44			
110										
111	Sum of All Estimated Var S ij									
112	25786.91									
113										
114	Sum of All S ij Squared									
115	66802.72									
116										
117	Optimal Shrinkage Intensity									
118	Using Above Results			Using Function (SHRINK)			Using Macro (Shortcut: Ctrl + s)			
119	0.278508			0.278508			0.278508			
120										
121	Covariance Matrix after Shrinkage									
122		1	2	3	4	5	6	7		Lead Pr Min
123	1	80.40	34.20	20.63	32.32	54.69	28.57	33.62		80.4
124	2	34.20	62.00	7.50	11.69	49.21	2.89	23.23		3815.24605
125	3	20.63	7.50	43.20	-14.86	13.71	40.98	18.33		144483.064
126	4	32.32	11.69	-14.86	141.60	38.09	-1.44	23.09		16512423.7
127	5	54.69	49.21	13.71	38.09	92.80	16.16	35.21		639771526
128	6	28.57	2.89	40.98	-1.44	16.16	83.20	27.13		2.6442E+10
129	7	33.62	23.23	18.33	23.09	35.21	27.13	36.80		3.7954E+11
130										
131										
132										

Figure 2 An Excel Example Illustrating the Determination of the Optimal Shrinkage Intensity (Continued).

Although the computations as illustrated above are straightforward, the size of the worksheet will increase drastically if the sample covariance matrix is based on many more observations. The increase is due to the need for a total of n^2T cells to store the individual values of $(w_{ijt} - \overline{w}_{ij})^2$, for $i, j = 1, 2, \ldots, n$ and $t = 1, 2, \ldots, T$. For students who have at least some rudimentary knowledge of generating user-defined functions in Excel, we can bypass the use of these n^2T cells for the intended computations.

By programming in Visual Basic for Applications (VBA), we can compute directly, via a user-defined function with three nested "for next" loops, the cumulative sum of $(w_{ijt} - \overline{w}_{ij})^2$ as the integers i, j, and t are to increase over their corresponding ranges. Specifically, we place in the computer memory an initial sum of zero. We vary i from 1 to $n - 1$, which is 6 in the example. For each case of i, we vary j from $i + 1$ to n, which is 7 in the example. For each case of j, we vary t from 1 to T, which is 6 in the example. Starting with $i = 1$, $j = i + 1 = 2$, and $t = 1$, we add $(w_{ijt} - \overline{w}_{ij})^2 = (w_{121} - \overline{w}_{12})^2$ to the initial sum, which is zero, thus resulting in a cumulative sum of $(w_{121} - \overline{w}_{12})^2$. The subsequent terms to be accumulated to the sum are, successively, $(w_{122} - \overline{w}_{12})^2$, $(w_{123} - \overline{w}_{12})^2$, $\ldots$, $(w_{126} - \overline{w}_{12})^2$, just to exhaust all 6 cases of t for $i = 1$ and $j = 2$. Then, we continue with $i = 1$ and $j = i + 2 = 3$ and add all 6 cases of $(w_{13t} - \overline{w}_{13})^2$ to the cumulative sum. The procedure continues until all 6 cases of $(w_{67t} - \overline{w}_{67})^2$ are finally accounted for.

We can use two of the above three nested loops in the same user-defined function to compute the sum of s_{ij}^2, for $i, j = 1, 2, \ldots, n$ and $i \neq j$, as well. The idea is to have a separate cumulative sum, which is also initialized to be zero in the computer memory. By going through all cases of i from 1 to $n - 1$ and, for each i, all cases of j from $i + 1$ to n, the cumulative sum will cover all $n(n - 1)/2$ cases of s_{ij}^2. With such a function in place, there will be no need for B54:H60 and B62:H109 in the worksheet to display the individual values of s_{ij}^2 and $(w_{ijt} - \overline{w}_{ij})^2$, respectively.

The VBA code of this user-defined function, named SHRINK, as described above, is provided in Appendix D. The cell formula for D119, which uses the function, is =SHRINK(B14:H19,B46:H52). The two arguments of the function are the cell references of the $T \times n$ matrix of all mean-removed observations and the $n \times n$ matrix with each element (i, j) being $\overline{w}_{ij}$. In essence, by retrieving the row numbers and the column numbers of the stored data in the worksheet, we are able to use the information there to provide the cumulative sums of s_{ij}^2 and $(w_{ijt} - \overline{w}_{ij})^2$. As expected, the end results based on the two alternative approaches, as shown in A119 and D119, are the same.

A more intuitive way to call the function SHRINK is to use a Sub procedure, which is an Excel macro defined by the user, that allows the user to provide, via three input boxes, the cells for the two arguments as required for the function to work, as well as the cell for displaying the end result. The VBA code of this Sub procedure is also provided in Appendix D. Again, as expected, the end result as shown in G119 is the same to those in A119 and D119.

5 Related Pedagogic Issues and Exercises for Students

In this paper, we have illustrated that shrinkage estimation for potentially improving the quality of the sample covariance matrix can be covered in classes where estimation issues of the covariance matrix are taught. From a pedagogic perspective, it is useful to address the issue of estimation errors in the sample covariance matrix before introducing shrinkage estimation. Students ought to be made aware of the fact that the sample covariance matrix is only an estimate of the true but unknown covariance matrix. As each of the sample variances and covariances of the set of random variables in question is itself a sample statistic, knowledge of the statistical concept of sampling distribution will enable students to appreciate more fully what $Var(s_{ij})$ is all about. Simply stated, the sampling distribution is the probability distribution of a sample statistic. By recognizing that each s_{ij} is a sample statistic, students will also recognize that it has a distribution and that the corresponding $Var(s_{ij})$ represents the second moment of the distribution.

For a given set of random variables, how well each sample covariance s_{ij} is estimated can be assessed in terms of standard error, which is the square root of the sampling variance that $\widehat{Var}(s_{ij})$ represents, relative to the point estimate itself. For example, as displayed in C22 of the Excel worksheet in Figure 2, we have $s_{12} = 47.4$. According to equation (10), the sampling variance of s_{12}, which is $\widehat{Var}(s_{12})$, can be computed by multiplying the sum of the six cells in C62:C67 with the factor $T/(T-1)^3$, where $T = 6$, The standard error of s_{12}, which is $\sqrt{\widehat{Var}(s_{12})} = \sqrt{1,293.29} = 35.96$, is quite large relative to the point estimate of $s_{12} = 47.4$. Given that the covariance estimation here is based on only six observations, a large estimation error is hardly a surprise. However, this example does illustrate the presence of estimation errors in the sample covariance matrix.

To enhance the learning experience of students, some exercises are suggested below. The first two types of exercises are intended to help students recognize, in a greater depth, the presence of errors in covariance matrix estimation and the impact of various factors on the magnitudes of such errors. The third type of exercises extends the shrinkage approach to estimating the correlation matrix. Such an extension is particularly relevant when observations of the underlying random variables are in different measurement units, as often encountered in life science studies.

5.1 Estimation of the Covariance Matrix with Actual Observations

Students can benefit greatly from covariance matrix estimation by utilizing actual observations that are relevant in their fields of studies. Depending on the fields involved, the observations can be in the form or empirical or experimental data. In finance, for example, stock returns can be generated from publicly available stock price and dividend data. Small-scale exercises for students, such as those involving the sample covariance matrix of monthly or weekly returns of the 30 Dow Jones stocks or some subsets of such stocks, estimated with a few years of monthly observations, ought to be manageable for students. The size of the corresponding Excel worksheet for each exercise will be drastically reduced if a user-defined function, similar to the function SHRINK above, is written for the computations of the sampling variances of s_{ij}, for $i, j = 1, 2, \ldots, n$, with $n \leq 30$. From such exercises, students will have hands-on experience with how the sampling variance that each $\widehat{Var}(s_{ij})$ represents tends to vary as the number of observations for the estimation increases.

5.2 Estimation of the Covariance Matrix with Simulated Data

In view of concerns about the non-stationarity of the underlying probability distributions in empirical studies or budgetary constraints for making experimental observations, the reliance on simulations, though tending to be tedious, is a viable way to generate an abundant amount of usable data for pedagogic purposes. An attractive feature of using simulated data is that we can focus on some specific issues without the encumbrance of other confounding factors. For example, with everything else being the same, the more observations we have, the closer tends to be between the sample covariance matrix and the true one. The use of a wide range of numbers of random draws will allow us to assess how well shrinkage estimation really helps in small to large sample situations. Further, simulated data are useful for examining issues of whether shrinkage estimation is more effective or less effective if the underlying variables are highly correlated, or if the magnitudes of the variances of the underlying variables are highly divergent. It will be up to the individual instructors to decide which specific issues are to be explored by their students.

To facilitate a simulation study, suppose that, under the assumption of a multivariate normal distribution, the true covariance matrix of the underlying random variables is known. From this distribution, we can make as many random draws as we wish. These random draws as simulated observations, in turn, will allow us to estimate the covariance matrix. To illustrate the idea of simulations in the context of covariance matrix estimation, consider a four-variable case, where the true covariance matrix is

$$\boldsymbol{V} = \begin{bmatrix} 36 & 12 & 24 & 36 \\ 12 & 64 & 32 & 24 \\ 24 & 32 & 100 & 48 \\ 36 & 24 & 48 & 144 \end{bmatrix}, \tag{11}$$

where each element is denoted by σ_{ij}, for $i, j = 1, 2, 3, 4$. The positive definiteness of $\boldsymbol{V}$ can easily be confirmed with Excel, by verifying the positive sign of each of its four leading principal minors. The analytical details pertaining to the generation of simulated observations for estimating $\boldsymbol{V}$, from the underlying probability distribution, are provided in Appendix C.

In this illustration, where $n = 4$, we consider $T = 5, 6, 7, \ldots, 50$. The lowest T is 5, because it is the minimum number of observations to ensure that the 4×4 sample covariance matrix be positive definite. For each T, the random draws to generate a set of simulated observations R_{it}, for $i = 1, 2, 3, 4$ and $t = 1, 2, \ldots, T$, and the corresponding sample covariance matrix $\widehat{\boldsymbol{V}}$ are repeated 100 times. The average of the 100 values of each estimated covariance s_{ij} is taken, for all $j > i$. Let $\overline{s}_{12}$, $\overline{s}_{13}$, $\overline{s}_{14}$, $\overline{s}_{23}$, $\overline{s}_{24}$, and $\overline{s}_{34}$ be such averages.

For each value of the shrinkage intensity, with $\lambda = 0.2$, 0.4, and 0.6, we take the difference between $|\overline{s}_{ij} - \sigma_{ij}|$ and $|\lambda\overline{s}_{ij} - \sigma_{ij}|$ for all $j > i$. We then take the average of all such differences, denoted in general by

$$\overline{D} = \frac{2}{n(n-1)} \sum\nolimits_{j>i} \left(|\overline{s}_{ij} - \sigma_{ij}| \, - \, |\lambda\overline{s}_{ij} - \sigma_{ij}| \right). \tag{12}$$

Here, $n(n-1)/2$ is the number of covariances in the upper triangle of the covariance matrix. This average can be interpreted as the difference in the mean absolute deviations for the two estimation methods. A positive $\overline{D}$ indicates that shrinkage provides an improvement. Figure 3 shows graphically how $\overline{D}$ varies with T for the three values of λ.

Some general patterns can be noted. The effectiveness of shrinkage to improve the quality of the estimated covariance matrix declines as the number of observations increases. The higher the number of observations, the more counter-productive is the reliance on heavy shrinkage to attenuate the errors in the sample covariances. However, if the number of observations is low, shrinkage is beneficial in spite of the strong bias that exists in the shrinkage target. The lower the number of observations, a higher shrinkage intensity will tend to result in a greater improvement in the estimated covariances.

5.3 Estimation of the Correlation Matrix

For analytical convenience, the pedagogic illustration in this paper has been confined to shrinkage estimation of the sample covariance matrix towards a diagonal target matrix with its diagonal ele-

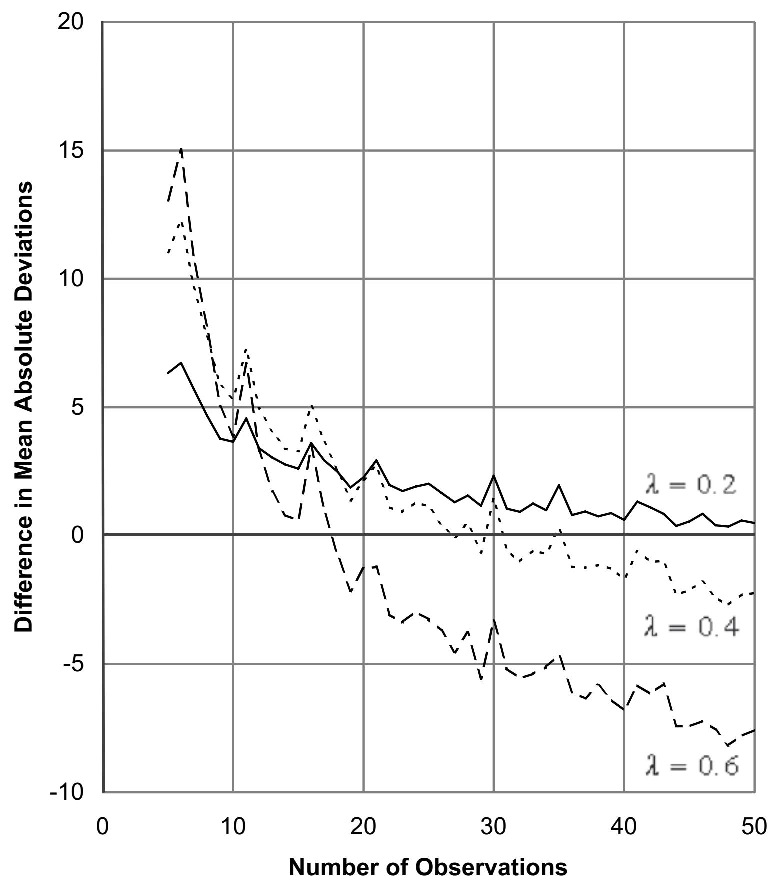

Figure 3 Differences in Mean Absolute Deviations from the True Covariances for Three Shrinkage Intensities from s Simulation Study, with a Positive Difference Indicating an Improvement by Shrinkage Estimation.

ments being the corresponding sample variances. An extension of the same approach to shrinkage estimation of the correlation matrix towards an identity matrix is simple, provided that the simplifying assumption in Schäfer and Strimmer (2005) is also imposed. That is, if estimation errors in all sample variances are ignored.

The idea is that, if we start with a set of n random variables, $\widetilde{R}_1/s_1, \widetilde{R}_2/s_2, \ldots, \widetilde{R}_n/s_n$, the element (i, j) of the $n \times n$ sample covariance matrix will be $s_{ij}/(s_i s_j)$, for $i, j = 1, 2, \ldots, n$. That is, if each of the n random variables is normalized by the corresponding sample standard deviation s_i, the resulting sample covariance matrix is the same as the sample correlation matrix of the original random variables. For this set of normalized random variables, shrinkage estimation of its covariance matrix towards a diagonal target matrix is equivalent to shrinking the sample correlation matrix of the original random variables, $\widetilde{R}_1, \widetilde{R}_2, \ldots, \widetilde{R}_n$, towards an identity matrix.

For shrinkage estimation of the correlation matrix instead, equation (5) can be written as

$$\lambda = \frac{\sum_{j>i} Var(r_{ij})}{\sum_{j>i}[Var(r_{ij}) + \rho_{ij}^2]}, \tag{13}$$

where $r_{ij} = s_{ij}/(s_i s_j)$ is the sample correlation of the original random variables i and j, with ρ_{ij} representing their true but unknown correlation. Under the assumption that $s_1, s_2, \ldots, s_n$ are without estimation errors as in the Schäfer-Strimmer study, each $Var(r_{ij})$ in equation (13) is simply $Var(s_{ij})$, divided by the product of the sample estimates s_{ii} and s_{jj}. Likewise, ρ_{ij}^2 can be approximated by $s_{ij}^2/(s_{ii}s_{ii})$, the square of the sample estimate s_{ij}, also divided by the product of the sample estimates s_{ii} and s_{jj}. In view of the simplicity of this revised formulation of shrinkage estimation, its small-scale implementation with actual observations is also suitable as an exercise for students. However, to relax the above assumption by recognizing the presence of estimation errors in the individual sample variances is a tedious analytical exercise.[6]

Why is shrinkage estimation of the sample correlation matrix relevant in practice? In portfolio investment settings, for example, to generate input data for portfolio analysis to guide investment decisions, if the individual expected returns and variances of returns are, in whole or in part, based on the insights of the security analysts involved, the correlation matrix is the remaining input whose estimation requires historical return data. If so, although the correlation matrix can still

[6] Besides the issue of estimation errors in the sample variances that already complicates the estimation of $Var(r_{ij})$, there is a further analytical issue. As indicated in Zimmerman, Zumbo, and Williams (2003), the sample correlation provides a biased estimate of the true but unknown correlation. However, as Olkin and Pratt (1958) show, the bias can easily be corrected if the underlying random variables are normally distributed. The correction for bias will make equation (13) more complicated. See, for example, Kwan (2008, 2009) for analytical details pertaining to the above issues.

be deduced from the shrinkage results of the covariance matrix, to estimate the correlation matrix instead is more direct.

Another justification for directly shrinking the sample correlation matrix does not apply to portfolio investment settings. Rather, it pertains to experimental settings, such as those in the various life science studies, where different measurement units are used for the underlying variables. Measurement units inevitably affect the magnitudes of the elements in each sample covariance matrix. When a quadratic loss function is used to determine the optimal shrinkage intensity, sample covariances with larger magnitudes tend to receive greater attention. Thus, to avoid the undesirable effects of the choice of measurement units on the optimal shrinkage results, it becomes necessary to normalize the random variables involved. However, as indicated earlier, doing so will also lead to analytical complications. Nevertheless, such issues ought to be mentioned when shrinkage estimation is introduced to the classroom.

6 Concluding Remarks

This paper has illustrated a novel approach, called shrinkage estimation, for potentially improving the quality of the sample covariance matrix for a given set of random variables. The approach, which was introduced to the finance profession, including its practitioners, only a few years ago, has also received considerable attention in some life science fields where invertible covariance and correlation matrices are used to analyze multivariate experimental data. Although the implementation of the approach can be based on various analytical formulations, with some formulations being analytically cumbersome, the two most common versions as reported in various life science studies are surprisingly simple. Specifically, in one version, an optimal weighted average is sought between the sample covariance matrix and a diagonal matrix with the diagonal elements being the corresponding sample variances. The other version involves the sample correlation matrix and an identity matrix of the same dimensions instead, under some simplifying assumptions.

This paper has considered, from a pedagogic perspective, the former version, which involves the sample covariance matrix. In order to understand shrinkage estimation properly, even for such a simple version, an important concept for students to have is that the sample covariance matrix, which is estimated with observations of the random variables considered, is subject to estimation errors. Once students are aware of this statistical feature and know its underlying reason, they can understand why a sample covariance of two random variables is a sample statistic and what the sampling variance of such a statistic represents. The use of Excel to illustrate shrinkage estimation will allow students to follow the computational steps involved, thus facilitating a better understanding of the underlying principle of the approach.

The role of Excel in this pedagogic illustration is indeed important. As all computational results are displayed on the worksheets involved, students can immediately see how shrinkage estimation improves the quality of the sample covariance matrix. For example, in cases where estimations are based on insufficient observations, students can easily recognize, from the displayed values of the leading principal minors, that the corresponding sample covariance matrix is problematic. They can also recognize that shrinkage estimation is a viable remedial measure. What is attractive about using Excel for illustrative purposes is that, as students will not be distracted by the attendant computational chores, they can focus on understanding the shrinkage approach itself.

As having hands-on experience is important for students to appreciate better what shrinkage estimation can do to improve the quality of the sample covariance matrix, it is useful to assign relevant Excel-based exercises to students. In investment courses, for example, shrinkage estimation of the covariance matrix of asset returns can be in the form of some stand-alone exercises for students. It can also be part of a project for students that compares portfolio investment decisions based on different characterizations of the covariance matrix. From the classroom experience of the author as an instructor of investment courses, the hands-on experience that students have acquired from Excel-based exercises is indeed valuable. Such hands-on experience have enabled students not only to be more proficient in various Excel skills, but also to understand the corresponding course materials better. This paper, which has provided a pedagogic illustration of shrinkage estimation, is intended to make classroom coverage of such a useful analytical tool less technically burdensome for students.

References

Bauwens, L., Laurent, S., and Rombouts, J.V.K., (2006). Multivariate GARCH models: a survey. *Journal of Applied Econometrics*, 21(1), 79-109.

Beerenwinkel, N., Antal, T., Dingli, D., Traulsen, A., Kinzler, K.W., Velculescu, V.E., Vogelstein, B., and Nowak, M.A., (2007). Genetic Progression and the Waiting Time to Cancer. *PLoS Computational Biology*, 3(11), 2239-2246.

Dabney, A.R., and Storey, J.D., (2007). Optimality Driven Nearest Centroid Classification from Genomic Data. *PLoS ONE*, 10, e1002.

Disatnik, D.J., and Benninga, S., (2007). Shrinking the Covariance Matrix: Simpler Is Better. *Journal of Portfolio Management,* Summer, 55-63.

Dobbin, K.K., and Simon, R.M., (2007). Sample Size Planning for Developing Classifiers Using High-Dimensional DNA Microarray Data. *Biostatistics*, 8(1), 101-117.

Jaggannathan, R., and Ma, T., (2003). Risk Reduction in Large Portfolios: Why Imposing the Wrong Constraints Helps. *Journal of Finance*, 58(4), 1651-1683.

Kriegeskorte, N., Goebel, R., and Bandettini, P. (2006) Information-Based Functional Brain Mapping. *PNAS*, 103(10), 3863-3868.

Kwan, C.C.Y., (2008). Estimation Error in the Average Correlation of Security Returns and Shrinkage Estimation of Covariance and Correlation Matrices. *Finance Research Letters*, 5, 236-244.

Kwan, C.C.Y., (2009). Estimation Error in the Correlation of Two Random Variables: A Spreadsheet-Based Exposition. *Spreadsheets in Education*, 3(2), Article 2.

Kwan, C.C.Y., (2010). The Requirement of a Positive Definite Covariance Matrix of Security Returns for Mean-Variance Portfolio Analysis: A Pedagogic Illustration. *Spreadsheets in Education*, 4(1), Article 4.

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, 603-621.

Ledoit, O., and Wolf, M., (2004a). A Well-Conditioned Estimator for Large-Dimensional Covariance Matrices. *Journal of Multivariate Analysis*, 88, 365-411.

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

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

Opgen-Rhein, R., and Strimmer, K., (2006a). Inferring Gene Dependency Networks from Genomic Longitudinal Data: A Functional Data Approach. *REVSTAT,* 4(1), 53-65.

Opgen-Rhein, R., and Strimmer, K., (2006b). Using Regularized Dynamic Correlation to Infer Gene Dependency Networks from Time-Series Microarray Data. *Proceedings of the 4th International Workshop on Computational Systems Biology*, 73-76.

Opgen-Rhein, R., and Strimmer, K., (2007a). Accurate Ranking of Differentially Expressed Genes by a Distribution-Free Shrinkage Approach. *Statistical Applications in Genetics and Molecular Biology*, 6(1), Article 9.

Opgen-Rhein, R., and Strimmer, K., (2007b). Learning Causal Networks from Systems Biology Time Course Data: An Effective Model Selection Procedure for the Vector Autoregressive Process. *BMC Bioinformatics,* 8, Supplement 2, S3.

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.

Silvennoinen, A., and Teräsvirta, T., (2009). Multivariate GARCH Models, in *Handbook of Financial Time Series*, Anderson, T.G., Davis, R.A., Kreiss, J.-P., and Mikosch, T., editors, Springer, 201-232.

Tenenhaus, A., Guillemot, V., Gidrol, X., and Frouin, V., (2010). Gene Association Networks from Microarray Data Using a Regularized Estimation of Partial Correlation Based on PLS Regression. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 7(2), 251-262.

Werhli, A.V., Grzegorczyk, M., and Husmeier, D., (2006). Comparative Evaluation of Reverse Engineering Gene Regulatory Networks with Relevance Networks, Graphical Gaussian Models and Bayesian Networks. *Bioinformatics*, 22, 2523-2531.

Yao, J., Chang, C., Salmi, M.L., Hung, Y.S., Loraine, A., and Roux, S.J., (2008). Genome-scale Cluster Analysis of Replicated Microarrays Using Shrinkage Correlation Coefficient. *BMC Bioinformatics*, 9:288.

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

Appendix A

Let $\boldsymbol{y}$ be an $n \times T$ matrix with each element (i, t) being

$$y_{it} = \frac{1}{\sqrt{T-1}} \left(R_{it} - \overline{R}_i \right), \quad \text{for } i = 1, 2, \ldots, n \text{ and } t = 1, 2, \ldots, T. \tag{A1}$$

As

$$s_{ij} = \sum\nolimits_{t=1}^{T} y_{it} y_{jt}, \quad \text{for } i = 1, 2, \ldots, n \text{ and } j = 1, 2, \ldots, n, \tag{A2}$$

which is the same as the product of row i of $\boldsymbol{y}$ and column j of $\boldsymbol{y}'$, we can write

$$\widehat{\boldsymbol{V}} = \boldsymbol{y}\,\boldsymbol{y}'. \tag{A3}$$

Accordingly, we have $\boldsymbol{x}'\widehat{\boldsymbol{V}}\boldsymbol{x} = \boldsymbol{x}'(\boldsymbol{y}\boldsymbol{y}')\,\boldsymbol{x} = (\boldsymbol{y}'\boldsymbol{x})'(\boldsymbol{y}'\boldsymbol{x})$, which is the product of a T-element row vector that $(\boldsymbol{y}'\boldsymbol{x})'$ represents and its transpose that $\boldsymbol{y}'\boldsymbol{x}$ represents. With v_t being element t of this row vector, it follows that $\boldsymbol{x}'\widehat{\boldsymbol{V}}\boldsymbol{x} = \sum_{t=1}^{T} v_t^2 \geq 0$. Thus, $\widehat{\boldsymbol{V}}$ is always positive semidefinite.

We now show that, if $\widehat{\boldsymbol{V}}$ is estimated with insufficient observations, its determinant is zero. If so, $\widehat{\boldsymbol{V}}$ is not invertible and thus is not positive definite. For this task, let us consider separately cases where $T < n$ and $T = n$. For the case where $T < n$, we can append a block of zeros to the $n \times T$ matrix to make it an $n \times n$ matrix. Specifically, let $\boldsymbol{z} = \begin{bmatrix} \boldsymbol{y} & \boldsymbol{0} \end{bmatrix}$, where $\boldsymbol{0}$ is an $n \times (n-T)$ matrix with all zero elements. With $\boldsymbol{z}\boldsymbol{z}' = \boldsymbol{y}\,\boldsymbol{y}'$, we can also write $\widehat{\boldsymbol{V}} = \boldsymbol{z}\,\boldsymbol{z}'$, a product of two square matrices. The determinant of $\widehat{\boldsymbol{V}}$ is the product of the determinant of $\boldsymbol{z}$ and the determinant of $\boldsymbol{z}'$. As each of the latter two determinants is zero, so is the determinant of $\widehat{\boldsymbol{V}}$.

For the case where $T = n$, $\boldsymbol{y}$ is already a square matrix. With each y_{it} being a mean-removed observation of R_{it}, scaled by the constant $\sqrt{T-1}$, the sum $\sum_{t=1}^{T} y_{it}$ for any i must be zero. Then, for $i = 1, 2, \ldots, n$, each y_{it} can be expressed as the negative of the sum of the remaining $T-1$ terms among $y_{i1}, y_{i2}, \ldots, y_{iT}$. That is, each column of $\boldsymbol{y}$ can be replicated by the negative of the sum of the remaining $T-1$ columns. Accordingly, the determinant of $\boldsymbol{y}$ is zero, and so is the determinant of $\widehat{\boldsymbol{V}}$.

Appendix B

As taking the expected value is like taking a weighted average, we have

$$E\left\{\sum\nolimits_{j>i}[(1-\lambda)s_{ij} - \sigma_{ij}]^2\right\} = \sum\nolimits_{j>i} E\{[(1-\lambda)s_{ij} - \sigma_{ij}]^2\}. \tag{B1}$$

Noting that $Var(\cdot) \equiv E\{[(\cdot) - E(\cdot)]^2\} = E[(\cdot)^2] - [E(\cdot)]^2$, for any random variable $(\cdot)$, we can also write

$$E\{[(1-\lambda)s_{ij} - \sigma_{ij}]^2\} = Var[(1-\lambda)s_{ij} - \sigma_{ij}] + \{E[(1-\lambda)s_{ij} - \sigma_{ij}]\}^2. \tag{B2}$$

With σ_{ij} being a constant, the term $Var[(1-\lambda)s_{ij} - \sigma_{ij}]$ reduces to $(1-\lambda)^2 Var(s_{ij})$. Further, as $E(s_{ij}) = \sigma_{ij}$, the term $E[(1-\lambda)s_{ij} - \sigma_{ij}]$ reduces to $-\lambda\sigma_{ij}$. It follows that

$$\sum\nolimits_{j>i} E\{[(1-\lambda)s_{ij} - \sigma_{ij}]^2\} = (1-\lambda)^2 \sum\nolimits_{j>i} Var(s_{ij}) + \lambda^2 \sum\nolimits_{j>i} \sigma_{ij}^2. \tag{B3}$$

Minimization of the loss function, by setting its first derivative respect to λ equal to zero, leads to

$$\frac{d}{d\lambda}\sum_{j>i} E\{[(1-\lambda)s_{ij}-\sigma_{ij}]^2\} = -2(1-\lambda)\sum_{j>i} Var(s_{ij}) + 2\lambda\sum_{j>i}\sigma_{ij}^2 = 0. \tag{B4}$$

Equation (5) follows directly.

Appendix C

It is well known in matrix algebra that a symmetric positive definite matrix can be written as the product of a triangular matrix with zero elements above its diagonal and the transpose of such a triangular matrix. This is called the Cholesky decomposition. Let $\boldsymbol{V}$ be an $n \times n$ covariance matrix and $\boldsymbol{L}$ be the corresponding triangular matrix, satisfying the condition that $\boldsymbol{LL'} = \boldsymbol{V}$.

To find $\boldsymbol{L}$, let us label the elements in its lower triangle as L_{ij}, for all $j \leq i$. Implicitly, we have $L_{ij} = 0$, for all $j > i$. Each L_{ij} in the lower triangle of $\boldsymbol{L}$ can be determined iteratively as follows:

$$L_{11} = \sqrt{\sigma_{11}}; \tag{C1}$$

$$L_{i1} = \sigma_{i1}/L_{11}, \text{ for } i = 2, 3, \ldots, n; \tag{C2}$$

$$L_{ii} = \sqrt{\sigma_{ii} - \sum_{k=1}^{i-1} L_{ik}^2}, \text{ for } i = 2, 3, \ldots, n; \tag{C3}$$

$$L_{ij} = \frac{1}{L_{jj}}\left(\sigma_{ij} - \sum_{k=1}^{j-1} L_{ik}L_{jk}\right), \text{ for } i = 3, 4, \ldots, n \text{ and } j = 2, 3, \ldots, i-1. \tag{C4}$$

Now, consider the standardized normal distribution, which is a normal distribution with a zero mean and a unit standard deviation. Let us take nT random draws from this univariate distribution and label them as u_{it}, for $i = 1, 2, \ldots, , n$ and $t = 1, 2, \ldots, T$. Let $\boldsymbol{U}$ be the $n \times T$ matrix consisting of the nT random draws of u_{it}.

As each u_{it} is a random draw, the sample mean, $\overline{u}_i = \sum_{t=1}^T u_{it}/T$, approaches zero as T approaches infinity. The sample variance, $\sum_{t=1}^T (u_{it} - \overline{u}_i)^2/(T-1)$, which approaches one as T approaches infinity, can be approximated as $\sum_{t=1}^T u_{it}^2/(T-1)$. The sample covariance, $\sum_{t=1}^T (u_{it} - \overline{u}_i)(u_{jt} - \overline{u}_j)/(T-1)$, which approaches zero as T approaches infinity, can be approximated as $\sum_{t=1}^T u_{it}u_{jt}/(T-1)$, for all $i \neq j$.

Accordingly, $\boldsymbol{UU'}/(T-1)$ approaches an $n \times n$ identity matrix as T approaches infinity. Let $\boldsymbol{W} = \boldsymbol{LU}$. It follows that, with $\boldsymbol{WW'} = \boldsymbol{LU}(\boldsymbol{LU})' = \boldsymbol{L}(\boldsymbol{UU'})\boldsymbol{L}'$, $\boldsymbol{WW'}/(T-1)$ approaches $\boldsymbol{LL'} = \boldsymbol{V}$, as T approaches infinity. The $n \times T$ matrix $\boldsymbol{W}$ can be viewed as a collection of T random draws from an n-variate distribution, with each row being the result of a random draw. To generate $\boldsymbol{W} = \boldsymbol{LU}$ requires the $n \times T$ matrix $\boldsymbol{U}$. By using Excel, we can generate each element of $\boldsymbol{U}$ with the cell formula =NORMSINV(RAND()).

Appendix D

The codes in Visual Basic for Applications (VBA) of a user-defined **function** procedure and a **Sub** procedure, for use in the Excel example, are shown below. The same codes can also be accessed from the supplementary Excel file (shrink.xls) of this paper, by opening the Visual Basic window under the Developer tab. Various examples pertaining to the syntax of the programming language can be found in Excel Developer Reference, which is provided under the Help tab there.

```
Option Explicit

Function SHRINK(nvo As Range, wijbar1 As Range) As Double

Dim nvar As Integer, nobs As Integer, mrr As Integer
Dim mrc As Integer, wijbar As Integer
Dim nvoc As Integer, nvor As Integer
Dim i As Integer, j As Integer, t As Integer
Dim s1 As Double, sum1 As Double, s2 As Double, sum2 As Double

'nvo: the cells containing all mean-removed observations
'nvar: the number of variables
'nobs: the number of observations
'mrr: the row preceding the mean-removed observations
'mrc: the column preceding the mean-removed observations
'wijbar: the row preceding the square matrix of w ij bar

nvar = nvo.Columns.Count
nobs = nvo.Rows.Count
mrr = nvo.Row - 1
mrc = nvo.Column - 1
wijbar = wijbar1.Row - 1

sum1 = 0
sum2 = 0

For i = 1 To nvar - 1
    For j = i + 1 To nvar
        s1 = Cells(wijbar + i, mrc + j).Value
        sum1 = sum1 + s1 * s1
```

```
        For t = 1 To nobs
            s2 = Cells(mrr + t, mrc + i).Value _
            * Cells(mrr + t, mrc + j).Value _
            - Cells(wijbar + i, mrc + j).Value
            sum2 = sum2 + s2 * s2
        Next t
    Next j
Next i

shrink = sum2 / (sum2 + sum1 * nobs * (nobs - 1))

End Function

Sub ViaFunction()

Dim arg1 As Range, arg2 As Range, out As Range

Set arg1 = Application.InputBox(prompt:= _
    "Select the cells for mean-removed observations", Type:=8)
Set arg2 = Application.InputBox(prompt:= _
    "Select the cells for the matrix of w ij bar", Type:=8)
Set out = Application.InputBox(prompt:= _
    "Select the cell for displaying the output", Type:=8)

Cells(out.Row, out.Column).Value = shrink(arg1, arg2)

End Sub
```