2.1. Proposed methodology

In this section, we discussed the estimation procedure of our proposed shrinkage estimator of the covariance matrix. Tong et al. (2014) reviewed different procedures for estimating the covariance matrix. They categorized all those procedures as Shrinkage estimators, Bayesian estimators, Regression estimators, Stein-type estimators, Sparse estimators, and Ridge estimators. The rationale behind Stein-type estimators is the shrinkage of eigenvalues of the covariance matrix. For $p \geq m$ scenario eigenvalues of the sample covariance matrix are usually highly unstable and variable. Therefore, shrinkage of eigenvalues towards a target structure will provide a well-conditioned and stable estimator.

Li et al. (2023) proposed a new shrinkage estimator for high frequency and high dimensional covariance matrix. Their proposed estimator is based on the shrinkage of eigenvalues towards a well-conditioned target estimator. They combined the realized covariance estimator that removes the effect of noise and heavy tails from a microstructure data of financial assets return in high-frequency trading with the shrinkage estimator of the high-dimensional covariance matrix to obtain a newly realized covariance shrinkage estimator. Using the idea of the shrinkage estimator of Li et al. (2023), we proposed the shrinkage estimator of the covariance matrix as

$$\hat{\Sigma}_{sh} = (\mathbb{S} + \psi I)^{1-\theta}\, T_g{}^{\theta}, \tag{4}$$

where $\mathbb{S}$ is the sample covariance matrix, I is the identity matrix, ψ is the perturbation parameter, $\theta \in [0,1]$ is the shrinkage intensity, and T_g is the target estimator. Regarding T_g we made two choices. Firstly, we chose T_g as the diagonal matrix of the geometric mean of sample variances of p quality characteristics and then we used the diagonal matrix of the arithmetic mean of the sample variances of p quality characteristics. Therefore, we denote both target estimators as $T_{g1} = \text{diag}\left(\left(\prod_{j=1}^{p} s_j^2\right)^{1/p}\right)$ and $T_{g2} = \text{diag}\left(p^{-1}\sum_{j=1}^{p} s_j^2\right)$, where $j = 1,2,\ldots,p$, $s_j^2 = (m-1)^{-1}\sum_{i=1}^{m}\left(y_{ij} - \bar{y}_j\right)^2$, y_j is j^{th} quality characteristic, y_{ij} is i^{th} measurement on y_j, and $\bar{y}_j = m^{-1}\sum_{i=1}^{m} y_{ij}$ is the sample mean of j^{th} quality characteristic. The rationale behind the choice of target estimator is provided in subsection 2.2. For our ease, we now rewrite Eq. (4) as

$$\hat{\Sigma}_{sh1} = (\mathbb{S} + \psi I)^{1-\theta_1} T_{g1}^{\theta_1}, \tag{5}$$

and

$$\hat{\Sigma}_{sh2} = (\mathbb{S} + \psi I)^{1-\theta_2}\, T_{g2}^{\theta_2}. \tag{6}$$

The unknown parameter θ_1 denotes the shrinkage intensity when the target estimator is $T_{g1} = \text{diag}\left(\left(\prod_{j=1}^{p} s_j^2\right)^{1/p}\right)$, and θ_2 denotes the shrinkage intensity when the target estimator is $T_{g2} = \text{diag}(p^{-1}\sum_{j=1}^{p} s_j^2)$. To estimate θ_1 we first use the spectral decomposition of the sample covariance matrix as

$$\mathbb{S} = U^t \Lambda_s U, \tag{7}$$

where U is the matrix whose column represents orthogonal and normalized eigenvectors of $\mathbb{S}$ and it further satisfies $UU^t = I$. While Λ_s is the diagonal matrix of the eigenvalues of $\mathbb{S}$ and is written as $\Lambda_s = \text{diag}(\lambda_1, \lambda_2, \ldots, \lambda_p)$. The Eq. (7) is further rewritten as

$$\mathbb{S} = U^t \begin{bmatrix} \lambda_1 & 0 & \ldots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & \lambda_p \end{bmatrix} U,$$

where $\lambda_1, \lambda_2, \ldots, \lambda_p$ are the eigenvalues of $\mathbb{S}$. This enables us to write

Schott: Teorema 10.16. Pág. 447

$$(\mathbb{S} + \psi I)^{1-\theta_1} = U^t \begin{bmatrix} (\lambda_1 + \psi)^{1-\theta_1} & 0 & \ldots & 0 \\ 0 & (\lambda_2 + \psi)^{1-\theta_1} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & (\lambda_p + \psi)^{1-\theta_1} \end{bmatrix} U.$$

Further, the spectral decomposition of the target estimator T_{g1} is written as

$$T_{g1} = U^t \Lambda_{T_{g1}} U,$$

where $\Lambda_{T_{g1}} = \text{diag}\left(\lambda_{T_{g11}}, \lambda_{T_{g12}}, \ldots, \lambda_{T_{g1p}}\right)$. This only holds if the target estimator is assumed to be a diagonal matrix of some scalar quantity. Then we can write $T_{g1}^{\theta_1}$ as

$$T_{g1}^{\theta_1} = U^t \begin{bmatrix} \lambda_{T_{g11}}^{\theta_1} & 0 & \ldots & 0 \\ 0 & \lambda_{T_{g12}}^{\theta_1} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & \lambda_{T_{g1p}}^{\theta_1} \end{bmatrix} U,$$

where $\lambda_{T_{g11}}, \lambda_{T_{g12}}, \ldots, \lambda_{T_{g1p}}$ are eigenvalues of T_{g1}.Then Eq. (5) can be rewritten as

$$\hat{\Sigma}_{sh1} = U^t \begin{bmatrix} (\lambda_1+\psi)^{1-\theta_1} & 0 & \dots & 0 \\ 0 & (\lambda_2+\psi)^{1-\theta_1} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & (\lambda_p+\psi)^{1-\theta_1} \end{bmatrix} \begin{bmatrix} \lambda_{T_{g11}}^{\theta_1} & 0 & \dots & 0 \\ 0 & \lambda_{T_{g12}}^{\theta_1} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \lambda_{T_{g1p}}^{\theta_1} \end{bmatrix} U.$$

$$= U^t \begin{bmatrix} (\lambda_1+\psi)^{1-\theta_1}\lambda_{T_{g11}}^{\theta_1} & 0 & \dots & 0 \\ 0 & (\lambda_2+\psi)^{1-\theta_1}\lambda_{T_{g12}}^{\theta_1} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & (\lambda_p+\psi)^{1-\theta_1}\lambda_{T_{g1p}}^{\theta_1} \end{bmatrix} U.$$

Further, we rewrite the above equation as

$$\hat{\Sigma}_{sh1} = U^t \left\{ (\Lambda_s + \psi I)^{1-\theta_1} \Lambda_{T_{g1}}^{\theta_1} \right\} U. \tag{8}$$

To estimate shrinkage intensity some authors for example James et al. (1961) and Tong and Wang (2007) suggested square and Stein loss functions. However, it is not guaranteed to be convex over [0,1], Li et al. (2022). Xiao et al. (2020) suggested using the squared log Euclidean loss (SLEL) function to obtain a unique optimum value for shrinkage intensity. Therefore, we employ SLEL to estimate a unique optimum value of θ_1 (i.e., θ_{1opt}) and define it as

$$\theta_{1opt} = \max\{0, \min(1, \tilde{\theta}_1)\}, \tag{9}$$

where $\tilde{\theta}_1$ is defined by the SLEL function as

$$\tilde{\theta}_1 = \underset{\theta_1}{\operatorname{argmin}}\, E\left\| \log(\hat{\Sigma}_{sh1}) - \log(\Sigma) \right\|_F^2, \tag{10}$$

where $\|.\|_F^2$ is squared Forbenius norm and Σ is the population covariance matrix and is defined as

$$\Sigma = V^t \Lambda_\Sigma V,$$

where V represents the eigenvector of Σ, $\Lambda_\Sigma = \operatorname{diag}(\lambda_{\Sigma_1}, \lambda_{\Sigma_2}, \dots, \lambda_{\Sigma_p})$ is a diagonal matrix of eigenvalues of Σ, and $\lambda_{\Sigma_1}, \lambda_{\Sigma_2}, \dots, \lambda_{\Sigma_p}$ are eigenvalues of Σ. As we already assumed U is an orthogonal matrix and $\hat{\Sigma}_{sh1}$ does not depend upon U, therefore, we write

$$\log(\hat{\Sigma}_{sh1}) = U^t \log\left\{ (\Lambda_s + \psi I)^{1-\theta_1} \Lambda_{T_{g1}}^{\theta_1} \right\} U.$$

By a similar argument, we can write $\log(\Sigma) = V^t \log(\Lambda_\Sigma) V$. Further, Eq. (10) is rewritten as

$$\tilde{\theta}_1 = \underset{\theta_1}{\operatorname{argmin}}\, E \left\| U^t \log\left\{ (\Lambda_s + \psi I)^{1-\theta_1} \Lambda_{T_{g1}}^{\theta_1} \right\} U - V^t \log(\Lambda_\Sigma) V \right\|_F^2,$$

further simplification provides

$$\tilde{\theta}_1 = \underset{\theta_1}{\operatorname{argmin}}\, E \left\| (1-\theta_1)U^t \log(\Lambda_s + \psi I)U + \theta_1 U^t \log(\Lambda_{T_{g_1}})U - V^t \log(\Lambda_\Sigma)V \right\|_F^2. \quad (11)$$

The θ_{1opt} can be estimated by minimizing Eq. (11) for θ_1 only if the structure of Σ is known (i.e., V and Λ_Σ are known). The values U, $\Lambda_s + \psi I$, and $\Lambda_{T_{g_1}}$ are calculated using sample data. However, in practical scenarios V and Λ_Σ are mostly unknown. Therefore, we cannot estimate θ_{1opt} algebraically using Eq. (11). Another solution for such a scenario is an iterative method of estimation provided by Chen et al. (2010) and later adopted by Ming et al. (2024) to estimate an OAS estimator of optimal shrinkage intensity. Following the same idea, the following steps are used to estimate θ_{1opt}.

i. Set Σ as the identity matrix and compute V and Λ_Σ.

ii. Using the computed value of V and Λ_Σ and the sample-based values i.e., U, $\Lambda_s + \psi I$, and $\Lambda_{T_{g_1}}$, compute the following function for different values of $\theta_1 \in [0,1]$.

$$H(\theta_1) = \underset{\theta_1}{\operatorname{argmin}} \frac{1}{M} \sum_{\ell=1}^{M} \left\| (1-\theta_1)U_\ell^t \log(\Lambda_{s\ell} + \psi I)\, U_\ell + \theta_1 U_\ell^t \log\left(\Lambda_{T_{g1\ell}}\right) U_\ell - V^t \log(\Lambda_\Sigma)V \right\|_F^2,$$

where M denotes the number of bootstrap samples. $H(\theta_1)$ is computed for $M = 100$ bootstrap samples for $\theta_1 = \{0.01, 0.02, \ldots, 1\}$. The value of θ_1 which provides the minimum value of $H(\theta_1)$ is chosen as the optimum value of θ_1 at this stage. We denote this value as $\theta_{1opt}^{(1)}$.

iii. Use $\theta_{1opt}^{(1)}$ to compute population covariance matrix Σ as

$$\Sigma_0 = (\mathbb{S} + \psi I)^{1-\theta_{1opt}^{(1)}} T_{g1}^{\theta_{1opt}^{(1)}}. \quad (12)$$

Now use Eq. (12) to compute V and Λ_Σ and repeat step (ii) to obtain $\theta_{1opt}^{(2)}$. Now using $\theta_{1opt}^{(2)}$, recompute Eq. (12), and repeat the procedure to obtain $\theta_{1opt}^{(3)}$. The process is repeated until the condition of convergence $\frac{\theta_{1opt}^{(k)} - \theta_{1opt}^{(k-1)}}{\theta_{1opt}^{(k-1)}} \leq 0.005$ is not satisfied, where the notation k denotes any k^{th} iteration. The value of $\hat{\theta}_{1opt}$ is obtained as the value that satisfies the above condition of convergence during this iterative process. A detailed iterative algorithm based on the above steps

is provided in Appendix A. The same algorithm is adapted to compute the optimum value of θ_2 (i.e., $\hat{\theta}_{2opt}$). The code using the above algorithm is written in software *R* and is available upon request. The one major limitation of this code is that it is computationally expensive and needs a powerful computer to run. After obtaining $\hat{\theta}_{1opt}$ and $\hat{\theta}_{2opt}$ we can easily obtain $\hat{\Sigma}_{sh1}$ and $\hat{\Sigma}_{sh2}$ using Eq. (5) and Eq. (6), respectively.

Generally, the perturbation parameter ψ is used to improve the robustness of the shrinkage estimator towards outliers and stabilize the results. By using the appropriate value of ψ the influence of outliers can be mitigated without compromising the benefits of the shrinkage process. Li et al. (2023) suggested that using ψ will adjust singularity issues in the estimation process of the covariance matrix. They suggested that the value of the perturbation parameter should be small and based on a simulation study, they recommended that $\psi = 0.0001$ provides minimum estimation error. Therefore, we also prefer to set $\psi = 0.0001$.

We discussed earlier in section 1 that for $p \geq m$ settings, Eq. (2) cannot be employed for monitoring the mean vector of a multivariate process because of the dimensionality issues (i.e., the singularity) of the covariance matrix. Similarly, when $m > p$, the covariance matrix is ill-conditioned if p is high and m is not large enough accordingly. However, our proposed shrinkage estimator of the covariance matrix provides an invertible and well-conditioned covariance matrix. This further enables us to develop a $NSBT^2$ control chart for improved monitoring of the process mean vector in $p \geq m$ scenario. Therefore, our proposed $NSBT^2$ control statistic for $p \geq m$ scenario is given as

$$T_i^2 = (Y_i - \bar{Y})^t \left(\hat{\Sigma}_{sh}\right)^{-1} (Y_i - \bar{Y}). \tag{13}$$

Further, the values of control statistic $T_1^2, T_2^2, \ldots, T_m^2$ for a given combination of p and m can be computed using Eq. (13). These values can be further plotted against the UCL to construct a control chart and compute the average run length (ARL). To compute the UCL, we used the concept of approximated chi-square distribution (Montgomery, 2005; Omolofe et al., 2022), which provides UCL = $\chi^2_{\alpha,p}$, where p is the degrees of freedom (d.f) and α is the level of significance. It is a well-established fact that using estimated values in the process adds extra variation to the process and it affects the in-control and out-of-control run lengths (Mahmoud & Maravelakis, 2010).

Therefore, many researchers, for example, Adegoke et al. (2021), Champ et al. (2005), and Mahmoud and Maravelakis (2010) studied the design and performance of MCC based on fixed in-control average run length (ARL_0) of 200. The ARL_0 is defined as the expected number of samples plotted on a control chart before the control chart detects an out-of-control point for an in-control process. In our study, we also computed the UCL at a fixed ARL_0 of 200.

To obtain a UCL, which provides a desired ARL_0 of 200 for a given combination of p and m, we initially utilize a prespecified value of α in UCL = $\chi^2_{\alpha,p}$, and determined a quantile point. Then using 5000 simulations we computed ARL_0. If the ARL_0 is 200, our search is over, and the quantile point of $\chi^2_{\alpha,p}$ is considered as the optimum value of UCL and is denoted as h_{opt} and the value of α which provides this h_{opt} is referred as α_{opt}. If ARL_0 is not equal to 200, we keep adjusting α and re-run the whole simulation unless ARL_0 of 200 is achieved. The simulation plan for computing UCL is provided in subsection 3.1.

2.2. Selection of the proper target estimator

The selection of a target estimator is not a random practice. Intuitively, the shrinkage estimation shrunk the estimator under study towards a well-structured estimator (i.e., target estimator). The target estimator must possess the two properties at the same time (1) It should be highly structured, that is, it must involve a small number of free parameters; (2) It must involve the characteristic of the unknown quantity being estimated (Ledoit & Wolf, 2003).

Therefore, the selection of a proper target estimator requires some diligence. In general, the target estimator is selected by presuming the lower dimensional (less number of estimated parameters) structure in the data set, as this improves the efficiency of shrinkage estimation over empirical estimation (Schäfer & Strimmer, 2005). They also suggested that employing any target structure or target estimator will result in a reduced mean square error (MSE) except in those minor cases where one has completely misspecified the target structure.

The most commonly used target estimator is the identity matrix and its scalar multiples. Schäfer and Strimmer (2005) provided a list of six types of target estimators commonly found in the literature. These six types are Type (i) diagonal, unit variance target estimator. It requires zero estimated parameters. Type (ii) diagonal, common variance (i.e., average of sample variance as diagonal entries). It requires one estimated parameter. Type (iii) common (co)variance that is

diagonal entries will have an average of sample variances and off-diagonal entries will have an average of covariances. It requires 2 estimated parameters. Type (iv) diagonal unequal variance and it requires p estimated parameters. Type (v) perfect positive correlation model and it also requires p estimated parameters. Type (vi) constant correlation model and requires $p+1$ estimated parameters. For details see Schäfer and Strimmer (2005).

The first three methods are low dimensional (i.e., 0 to 2 free parameters). Therefore, they possess a strong structure and require little data to fit. They have the ability to shrink both the diagonal and off-diagonal elements of the covariance matrix. Several authors frequently utilized the first two types of target estimators for example Hastie (2001) employed Type (i), similarly Dobra et al. (2004), Friedman (1989), Ledoit and Wolf (2004), and Leung and Chan (1998), etc employed Type (ii). The Type (iii) target estimator is not widely used. Type (iv)−Type (vi) also have limited applications because they are parameter-rich and shrink only the off-diagonal element of the covariance matrix.

Table 1: **A list of proposed target estimators (T_g).**

Diagonal matrix of the geometric mean of sample variances. i.e., $T_{g1} = \text{diag}\left(\left(\prod_{j=1}^{p} s_j^2\right)^{1/p}\right)$.	Diagonal matrix of the arithmetic mean of sample variances i.e., $T_{g2} = \text{diag}\left(p^{-1}\sum_{j=1}^{p} s_j^2\right)$.
Diagonal matrix of unit variance or an Identity matrix (I) i.e., $T_{g3} = I$ or any multiple of I i.e., $T_{g3} = kI$, where k is any constant.	Diagonal matrix of the median of sample variances i.e., $T_{g4} = \text{diag}(Med)$, where Med is the median of sample variances.
Diagonal matrix of the arithmetic mean of sample covariances. i.e., $T_{g5} = \text{diag}(A_{cov})$, where A_{cov} is the arithmetic mean of sample covariances.	Diagonal matrix of the geometric mean of the sample covariances i.e., $T_{g6} = \text{diag}(G_{cov})$, where G_{cov} is the geometric mean of sample covariances.

As our study is based on the shrinkage of eigenvalues of the covariance matrix, therefore, we use the spectral decomposition of the target estimator which is given as

$$T_g = U^t \Lambda_{T_g} U.$$

This can be ensured only if the target estimator is some scalar matrix, for example, a diagonal unit variance or a diagonal common variance (i.e., Type (i) or Type (ii) target estimators). This

further provides us with a flexible choice regarding the selection of a target structure. Table 1 provides a list of some proposed target estimators, that can be utilized with the proposed methodology of estimating covariance matrix.

However, we only used $T_{g1} = \text{diag}\left(\left(\prod_{j=1}^{p} s_j^2\right)^{1/p}\right)$ and $T_{g2} = \text{diag}\left(p^{-1}\sum_{j=1}^{p} s_j^2\right)$ as target estimators in our study as both choices ensure high structure (i.e., only one estimated parameter for each target estimator) and they reflect important characteristics regarding the estimated covariance matrix. Nevertheless, one can choose any target structure from the Table 1.

3. Simulation study

A simulation study is conducted to employ the proposed $NSBT^2$ control chart and evaluate its practical significance. All the simulations are conducted in R statistical software. We draw m Phase I samples of size $n = 1$ from p-variate standard normal distribution having identity covariance matrix I_p and mean vector zero $\left(\text{i.e.}, N_p\left(0, I_p\right)\right)$ and computed all Phase I estimates. To compute the optimum shrinkage intensity we employed the algorithm provided in Appendix A. Further, to evaluate the detectability of the shifts, we introduced different percentages of contaminated observations in the initial Phase I in-control data and control charts were constructed. The contaminated observations are Phase II observations and are drawn from p-variate normal distribution having a shift in the mean vector $\left(\text{i.e.}, N_p\left(\mu_A, I_p\right)\right)$, where μ_A represents shifted mean vector. To evaluate the ARL_1 performance of the proposed $NSBT^2$ control chart in Phase II, we also considered both uncorrelated and correlated data sets. The details are provided in the following subsequent subsections.

3.1. Construction of proposed $NSBT^2$ control chart

We draw m in-control samples of size $n = 1$ from in-control p-variate normal distribution $N_p\left(0, I_p\right)$ to compute control chart estimates in Phase I. Firstly, we considered T_{g1} as the target estimator and computed $\hat{\theta}_{1opt}$ and $\hat{\Sigma}_{sh1}$ using the algorithm provided in Appendix A and Eq. (5), respectively. As we discussed earlier the computation of $\hat{\theta}_{1opt}$ is a time-consuming process. Therefore, for our ease, we computed the average values of $\hat{\theta}_{1opt}$ for different combinations p