Hotelling T^2 control chart based on minimum vector variance for monitoring high dimensional correlated multivariate process

Odunayo Adiat Oyegoke[1], Kayode Samuel Adekeye[2], John Olutunji Olaomi[3], and Jean-Claude Malela-Majika[4*]

[1]Department of Statistics, Osun State Polytechnic, Iree, Nigeria
[2]Department of Mathematics and Statistics, Redeemer's University, Ede, Nigeria
[3]Department of Statistics, University of South Africa, Pretoria, South Africa
[4]Department of Statistics, University of Pretoria, Hatfield, Pretoria 0028, South Africa.
[*]Corresponding author: malela.mjc@up.ac.za

ABSTRACT

Multivariate control charts are practical tools that simultaneously monitor several correlated quality characteristics in a process. Monitoring high dimensional data structures is challenging because in most cases, the process sample size for monitoring parameters is greater than the number of process characteristics. Many researchers have used the multivariate Hotelling's T^2 chart to monitor high dimensional data using the maximum-likelihood methods (MLM) to estimate the covariance matrices. However, the multivariate Hotelling's T^2chart based on MLM suffers from low statistical performance. In this paper, we proposed a multivariate Hotelling's T^2 chart based on the minimum vector variance (MVV) and some regularized methods for monitoring high dimensional data structures. The performance of the proposed chart is evaluated in terms of the average run length (*ARL*). The results reveal the superiority of the proposed MVV Hotelling's T^2 chart over the existing Hotelling's T^2 charts for high dimensional correlated processes.

Keywords: High dimensional data; Hotelling T^2 control chart; correlated multivariate process; covariance matrix; minimum vector variance; regularized methods

1.0 INTRODUCTION

Most real-world challenges in a process are multivariate in nature and often laced with either natural or unnatural variation. Monitoring unnatural variations are an increasingly important issue in many real-life applications [1]. Unnatural variations are often based on either process mean and variability, and thus it is essential to determine the cause of unnatural variation, which may be due to the 5Ms (i.e., man, machine, materials, methods and measurement); see, for example, [2]. In a multivariate process, the covariance matrix can be monitored to determine whether the process is in a state of statistical control [3]. There are many multivariate statistical process control (MSPC) tools that provide efficient monitoring structures by using the relationships that exist among variables to identify unwanted variation in the quality of the

products or services. Recently, MSPC tools have been used in different scientific and technological fields characterized by high-dimensional data [25]. Many studies on multivariate extensions of the univariate monitoring schemes have been reported in the literature; see, for example, [4], [5], [6], and [7]. Most of these extensions suffer from methodological problems when the parameters of the distributions are unknown, particularly for high-dimensional process monitoring since they are based on the assumption of known process parameters. Due to the high degree of complexity of high-dimensional structure, estimating the covariance matrix becomes more challenging (see, [8], [9] and [10]). Therefore, using the classical maximum likelihood estimator (MLE) has been established to perform poorly for high dimensional structures because of its singular and ranks deficiency for high dimensional structure ([11] and [12]). Thus, the monitoring structures are based on a linear combination of the process variables, making interpreting results more challenging. The shortcoming necessitates the development of a chart involving shrinkage and other estimators that can work for high dimensional covariance matrices. To overcome the challenges posed by high dimensionality in monitoring multivariate processes, some regularized methods have been proposed by [13] and [14]. Therefore, this study focuses on monitoring shifts in the mean vector ($\boldsymbol{\mu}_0$) using Hotelling's T^2 chart with shrinkage and minimum vector variance estimators for individual high-dimensional data structure. The shrinkage estimators considered are the ridge quadratic variance and ridge quadratic null regularizations.

The remainder of this paper is organised as follows. Section 2 presents the materials and methodological background of this study. In addition, Section 2 introduces the new multivariate Hotelling's T^2 charts based on the VVM and regularized methods. The simulation study is presented in Section 3. The simulation results, comparative analysis and illustrative example are given in Section 4. Section 5 presents the overall summary and future research works.

2.0 *Materials and Methods*

This section presents a procedure for multivariate process control of high-dimensional process for phase I and phase II process monitoring.

2.1 *High-Dimensional Multivariate Structure*

Suppose *Y* is a multivariate *p*-dimensional vector from a multivariate normal distribution with mean vector, **μ**, and covariance matrix, $\sum$. If the mean vector and covariance matrix are known, then **μ** = **μ₀** and $\sum = \sum_{\mathbf{0}}$ represent the known mean vector and covariance matrix, respectively.

In High dimensional structure, the observation, known mean and known covariance are represented as:

$$Y = [Y_1, Y_2, Y_3, \cdots Y_p]', \tag{1}$$

$$\mu = [\mu_1, \mu_2, \mu_3, \cdots \mu_p]' \tag{2}$$

and

$$\sum = \begin{bmatrix} \sigma_1^2 & \sigma_{12} & \cdots & \sigma_{1p} \\ \sigma_{21} & \sigma_2^2 & \cdots & \sigma_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{p1} & \sigma_{p2} & \cdots & \sigma_p^2 \end{bmatrix}. \tag{3}$$

Now, assuming Y_t= [$y_{i1t,}\, y_{i2t,}\, y_{i3t,}\ldots,\, y_{ipt}$] is the t^{th} sample matrix which contains y_{ijt} for i^{th}observation of the j^{th} quality characteristics on the t^{th}sample, i =1, 2, 3, …, *n; j* =1, 2, 3, …, p and t = 1, 2, 3, …, *m).* If $\bar{Y}_t$is the t^{th}*p*-variate mean vector, and $\boldsymbol{S}_t$ is the $p \times p$ covariance matrix of the t^{th} sample or profile, then, we have the following:

$$\overline{Y}_t \equiv [\bar{Y}_{1t}, \bar{Y}_{2t}, \cdots \bar{Y}_{pt}]' = \begin{bmatrix} \bar{Y}_{1t} \\ \bar{Y}_{2t} \\ \vdots \\ \bar{Y}_{pt} \end{bmatrix}, \text{ where } \bar{Y}_{kt} = \frac{1}{n}\sum_{i=1}^{n} Y_{ikt}\,,\ k = 1,2, \ldots p$$

and

$$\boldsymbol{S}_t = \ [S_{kkt}], \text{where } S_{kkt} = S_{kt}^2 = \frac{1}{n-1}\sum_{i=1}^{n}(y_{ikt} - \bar{Y}_{kt})(y_{ikt} - \bar{Y}_{kt})\,.$$

When both $\boldsymbol{\mu}_0$ and $\sum_0$ *a*re unknown which is often the case in real-life application, they are usually estimated using the classical estimators from the Phase I *m-in-control* sample.

The estimator for the unknown mean, μ, and covariance matrix, $\sum$, for the Phase I *m-in-control* sample are given by:

$$\bar{\bar{Y}} = \frac{\sum_{t=1}^{m} \bar{Y}_t}{m} \text{ and } \bar{S} = \begin{bmatrix} \bar{s}_1^2 & \bar{s}_{12} & \cdots & \bar{s}_{1p} \\ \bar{s}_{21} & \bar{s}_2^2 & \cdots & \bar{s}_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \bar{s}_{p1} & \bar{s}_{p2} & \cdots & \bar{s}_p^2 \end{bmatrix},$$

where

$\bar{s}_k^2 = \frac{\sum_{i=1}^{m} s_{kt}^2}{mc_{4,m}} and \bar{s}_{jk} = \frac{\sum_{i=1}^{m} s_{ijk}}{mc_{4,m}} forall j \neq k$ with $c_{4,m} = \frac{\sqrt{2}\Gamma\left(\frac{m(n-1)+1}{2}\right)}{\sqrt{m(n-1)}\Gamma\left(\frac{m(n-1)}{2}\right)}$.

2.2*The Hotelling's T^2 Control Chart for High Dimensional Structure*

The Hotelling's T^2control chart is one of the most used multivariate process control charts by practitioners. The chart statistics is given by:

$$T_t^2 = n(\bar{Y} - \bar{\bar{Y}})'\bar{S}^{-1}(\bar{Y} - \bar{\bar{Y}}) \quad (4)$$

To monitor individual sample observation, that is for n = 1, Equation (4) becomes:

$$T_t^2 = (Y - \bar{Y})'\bar{S}^{-1}(Y - \bar{Y}). \quad (5)$$

The Hotelling's T^2 control chart gives a signal if

$$T_t^2 >= UCL \quad (6)$$

where UCL = $\frac{P(m-1)^2}{m}\beta(\alpha)_{\frac{p}{2},\frac{m-p-1}{2}}$.

Since it's a one-sided chart, $\beta(\alpha)_{\frac{p}{2},\frac{m-p-1}{2}}$ is the upper α percentage point of a beta distribution with $p/2$ and $(m-p-2)/2$ degrees of freedom and **α** (0.05) is the probability of a type I error. If the statistics in Equation (4) or Equation (5) is greater than the UCL, then the process under investigation is adjudged to be *out-of-control*. It should be noted that the above is the Phase I operation.

In the Phase II operation, the estimated parameters will then be used to construct different charts based on the different estimation methods and will be used to monitor the new process observations. The different estimation methods namely, the Minimum Vector Variance (MVV) and the regularized methods are presented in the next section.

2.3 *Estimation of Covariance Matrix*

In this study, the choice of the covariance matrix of the Hotelling T^2 statistic in Equation (4) is very crucial and stands as the nerve of this study. The determination of the covariance matrix plays a very important role on the robustness of the Hotelling T^2 control chart in the determination of *out-of-control* signal and the Average run length (ARL). The two methods considered in this study for the estimation of the covariance matrix are the minimum vector variance and the shrinkage methods.

2.3.1 *Minimum Vector Variance*

The MVV is one of the multivariate robust estimators that is effective in detecting outliers and controlling the type I error [21]. Furthermore, the robustness of the MVV estimator has been found to be equivalent to that of the minimum covariance determinant (MCD) and has greater computational efficiency than the MCD; see for example,[19] and [20]. The major techniques used in the estimation of MVV is known as the Mahalanobis squared distances (MSDs). Let $Y = \{y_1, y_2, \cdots, y_n\}$ be a data set of p-variate observations. Let the MVV estimators for the location parameter and scatter be M_{MVV},and S_{MVV}, respectively. Then, let $H \subseteq Y$, the M_{MVV} and S_{MVV} are obtained based on the data subset H of $h = \left\lfloor \frac{n+p+1}{2} \right\rfloor$ data points, their indices are the are index set H,

The location (Mmvv) and the spread (Smvv) of MVV estimator for a random variable Y are respectively.

$$\mathrm{M_{mvv}} = \frac{1}{h}\sum_{i\in H} Y_i \tag{7}$$

and

$$S_{\mathrm{mvv}} = \frac{1}{h}\sum_{i\in H}(Y_i - M_{MVV})\,(Y_i - M_{MVV})^t, \tag{8}$$

where y_i are the observable value of the random variable Y.

2.3.2 *Ridge Estimator*

The ridge estimator is a regularization method that shrinks the estimated elements of the concentration matrix proportionately together with some form of post hoc element selection.

The idea of using regularization approaches to address the large p small n problem has been suggested by [23]. In the case of covariance matrix estimation, the goal of the penalized likelihood approach is to constrain the MLE to a symmetric positive definite and thus invertible matrix with a simpler structure and improved statistical properties (e.g., correcting for the over

dispersion of the eigenvalues or reducing the quadratic loss). This is achieved by adding a penalty term to the likelihood that favors solutions with the required structure and by finding the optimal solution that maximizes the penalized likelihood.

In this paper, two ridge estimators were considered; namely, the ridge quadratic variance and ridge quadratic null estimators.

(a) *Ridge quadratic variance estimator*

The ridge quadratic variance estimator, denoted as *Rqv* hence forth in this paper, was presented by [12] as a convex combination of the variance (***S***) and a target matrix (**T**) and it is given by

$$\widehat{\boldsymbol{\Omega}}(\lambda) = \frac{1}{\lambda}[\boldsymbol{S} - (\boldsymbol{S} - \lambda \boldsymbol{T})], \tag{9}$$

where $\boldsymbol{\lambda} \in (0, \infty)$ is a penalty parameter. The implication here is that $\boldsymbol{\lambda}$ can take values between 0 and ∞. If $\boldsymbol{\lambda}$= 0, then the expression becomes Least Square Estimate (LSE). If $\boldsymbol{\lambda}$ approaches ∞, the expression converges to zero.

(b) *Ridge quadratic null estimator*

The Ridge quadratic null estimator (*Rqnull*) homogeneously reduces all terms of the target matrix $\boldsymbol{T}$ to zero. It should be noted that the *Rqnull* are rotation-invariant, meaning that they are robust to data orientation [22]. The computation formula for *Rqnull* is given by:

$$\widehat{\boldsymbol{\Omega}}(\lambda) = \left\{\left[\lambda \mathbf{I}_p + \frac{1}{4}(\mathbf{S} - \lambda \mathbf{T})^2\right]^{1/2} + \frac{1}{2}(\mathbf{S} - \lambda \mathbf{T})\right\}^{-1}. \tag{10}$$

When the target matrix **T** choose to be null element of $\hat{S}$ and the regularization parameter $\boldsymbol{\lambda}$ can take any value in (0, ∞), Equation (10) becomes:

$$\widehat{\boldsymbol{\Omega}}(\lambda) = \left\{\left[\lambda \mathbf{I}_p + \frac{1}{4}\boldsymbol{s}^2\right]^{1/2} + \frac{1}{2}\boldsymbol{s}\right\}^{-1}. \tag{11}$$

2.4 *Selection of the Shrinkage parameters*

There are several approaches in the literature that could be considered for selecting an optimal value for the turning parameters λ (see [15], [16] and [17]). To choose λ using cross-validation,

we adopt the Leave-One-Out-Cross-Validation (LOOCV) approximate version of the penalized likelihood for $\widehat{\Omega}(\lambda)$ of a fixed λ given by [18] as:

$$\phi^{n}(\lambda) = -\frac{1}{n}L\left[\widehat{\Omega}(\lambda); S\right] + \frac{1}{2n(n-1)}\sum_{i=1}^{n}\vartheta_1, \qquad (12)$$

where

$$\vartheta_i = \sum_{j1=1}^{p}\sum_{j2=1}^{p}\{[[\Omega(\lambda)]^{-1} - Y_i Y_i'] \circ [\Omega(\lambda)\left(S - Y_i Y_i'\right)\Omega(\lambda)]\}_{j_1,j_2},$$

and ° is the Hadamard product.

3. Simulation Study

The *Hotelling T^2 control chart* based on the ridge-quadratic type estimate is given as:

$$T^2\widehat{\Omega}_{ridge} = (Y_i - \bar{Y})'\widehat{\Omega}(\lambda^*)(Y_i - \bar{Y}), \qquad (13)$$

where $\widehat{\Omega}(\lambda^*)$ is the optimum precision matrix from any ridge-quadratic type estimation.

The chart will give an out - of - control (OOC) signal whenever the statistics $T^2\,\widehat{\Omega}_{ridge}$ in Equation (13) is greater than the upper control limit (UCL), W ($W > 0$), which is chosen to achieve a specified in-control ARL (ARL_0).

In this study, the binary search algorithm proposed by [2] was used to search for the values of the control charts' limits that will produce the desired in-control ARL value. We prespecified the ARL_0 to be 250. To obtain the limits, a Monte-Carlo simulation was conducted with 10,000 repetitions. In each simulation, we generated the Phase I sample from a standard multivariate normal distribution with $\boldsymbol{\mu_0}$ and $\boldsymbol{\Sigma_0}$.

4. Results and Discussion

4.1 Simulation Results and Performance evaluation

In this section, performances of the proposed Hotelling T^2 control chart based on three estimators are investigated in terms of the ARL metric. The run-length variable represents the number of samples plotted on the chart before it gives an OOC signal for the first time. The performances of the proposed charts are computed using Monte Carlo simulation in R and the results are summarized in Tables 1-3. A visual display of the performance of the proposed chart is given in Figure 1- Panels (a)-(f). Figure 1 investigates the performance of the proposed chart in terms of the OOC ARL profile for different values of p and M. From Tables 1-3 and Figure

1, it can be observed that regardless the values of M, as p increases, there is a noticeable difference in the performance of the proposed charts. The difference became much noticeable as the shift size in the process mean increases. For small and moderate values of p and M, the Hotelling T^2 chart based on MVV is more sensitive and perform better as compared to the competing charts based on other estimators. But, for large values of p and M, regardless of the magnitudes of the shift in the process mean, the Hotelling T^2 based on regularized estimators perform similarly and better than other competing charts (see Tables 1-3).

Tables 1-3 present the OOC ARL (ARL_1) values of the proposed charts based on the MLE, MVV, Rqv, and Rqnull for $p = 2$, 5, 10, 15, 20 and 25; M = 50,100,150, 200 and 250 under different shifts ($\boldsymbol{\delta}$) for the auto-regressive covariance structure. The Hotelling T^2 chart based on the MLE estimator performs worse regardless of the value of p and M.

Table 1: ARL profile of the auto-regressive covariance structure when M = 50 and 100 for different p values

P	δ	M=50				M=100			
		MLE	MVV	Rqv	Rqnull	MLE	MVV	Rqv	Rqnull
2	0.0	276.21	251.09	259.18	267.96	258.50	248.68	251.94	249.94
	0.5	161.84	148.06	149.19	156.22	190.23	120.75	157.67	180.46
	1.0	96.51	69.52	75.12	83.64	67.75	40.12	35.77	33.66
	1.5	69.46	46.48	53.61	60.22	34.75	10.63	12.16	23.15
	2.0	45.50	20.32	30.20	34.90	25.60	6.50	10.80	15.89
	2.5	10.06	5.09	7.54	8.73	9.46	3.35	4.98	6.08
	3.0	5.96	2.53	3.02	4.25	4.13	2.11	1.88	1.90
	3.5	3.47	1.78	2.29	3.11	2.58	1.47	1.31	1.41
	4.0	2.20	1.05	1.21	1.77	2.33	1.19	1.56	2.02
5	0.0	258.00	249.05	251.72	255.40	258.96	261.80	251.10	249.00
	0.5	173.13	132.84	140.46	152.83	181.67	150.42	164.96	172.83
	1.0	82.85	59.27	65.37	70.74	74.48	33.82	55.16	63.74
	1.5	40.49	10.06	19.35	25.20	54.81	15.66	26.53	33.20
	2.0	15.73	5.09	9.11	11.25	17.01	7.90	8.83	10.25
	2.5	8.19	3.02	5.85	6.37	13.78	4.54	6.94	8.37
	3.0	4.99	2.03	3.08	3.96	8.31	2.43	3.06	4.06
	3.5	3.51	1.53	1.35	2.41	4.59	2.03	2.61	3.41
	4.0	2.23	1.03	1.30	2.04	2.58	1.09	1.42	1.14
10	**0.0**	266.12	253.10	249.52	254.38	261.45	248.53	257.54	252.50
	0.5	163.72	140.75	170.42	182.24	180.36	160.94	154.34	170.29
	1.0	90.26	52.88	64.19	83.25	76.65	46.68	51.62	62.85
	1.5	45.86	28.14	30.17	36.23	50.82	21.26	35.88	47.86
	2.0	30.69	12.59	21.23	27.29	40.67	18.22	26.25	30.16
	2.5	15.37	6.19	10.30	11.11	20.85	11.26	15.04	14.54
	3.0	10.77	4.08	6.53	7.58	15.81	4.70	7.98	10.62
	3.5	4.76	2.41	3.04	4.21	6.44	2.22	3.77	4.29
	4.0	2.40	1.36	2.01	2.07	3.20	1.87	2.15	2.07
15	**0.0**	255.13	251.84	256.04	249.57	258.27	245.36	252.45	260.29
	0.5	184.97	124.11	121.50	174.43	167.30	126.71	135.65	150.65
	1.0	73.87	31.37	59.06	63.45	92.40	45.47	51.63	71.05
	1.5	42.98	15.05	25.20	32.65	67.81	22.25	39.84	46.96
	2.0	21.20	8.11	11.90	19.96	34.87	14.78	19.18	24.34
	2.5	9.72	8.21	7.32	10.53	13.21	7.08	8.09	10.86
	3.0	5.80	3.67	4.06	5.25	2.23	5.07	6.41	7.99
	3.5	4.69	2.20	3.75	4.06	3.63	1.68	2.56	3.45
	4.0	2.98	1.00	1.68	2.07	2.53	1.33	1.92	2.16
20	**0.0**	256.00	251.17	247.10	249.78	257.49	251.24	262.32	256.68
	0.5	172.00	157.58	146.68	161.79	186.71	153.75	130.09	145.98
	1.0	91.78	66.12	45.37	51.58	75.69	55.08	25.04	35.07
	1.5	48.60	30.77	19.28	21.42	34.61	28.47	16.47	20.11
	2.0	22.40	17.59	10.44	15.25	23.48	15.72	9.35	10.10
	2.5	13.26	8.31	5.41	11.13	10.62	8.34	6.43	6.93
	3.0	5.14	4.43	3.74	7.82	5.97	5.35	4.07	4.76
	3.5	4.08	3.90	2.37	4.97	3.99	3.52	2.23	3.02
	4.0	3.83	2.44	1.57	3.02	2.53	1.44	1.00	1.09
	0.0	251.25	259.88	248.81	250.91	254.97	262.48	249.15	250.43

25	**0.5**	202.50	148.06	119.82	134.94	198.51	130.75	111.49	117.95
	1.0	90.23	80.52	70.23	77.46	130.26	108.12	40.53	60.92
	1.5	37.75	33.48	21.96	26.66	84.69	63.63	13.72	21.86
	2.0	17.75	14.32	8.79	10.15	38.85	26.50	10.71	12.91
	2.5	8.26	6.09	3.09	5.19	18.56	13.35	5.37	6.58
	3.0	4.46	3.83	2.14	2.08	7.45	5.11	2.40	3.63
	3.5	3.13	2.78	1.06	1.90	3.49	2.47	1.41	2.31
	4.0	2.58	1.35	1.00	1.00	2.33	2.19	1.00	1.84

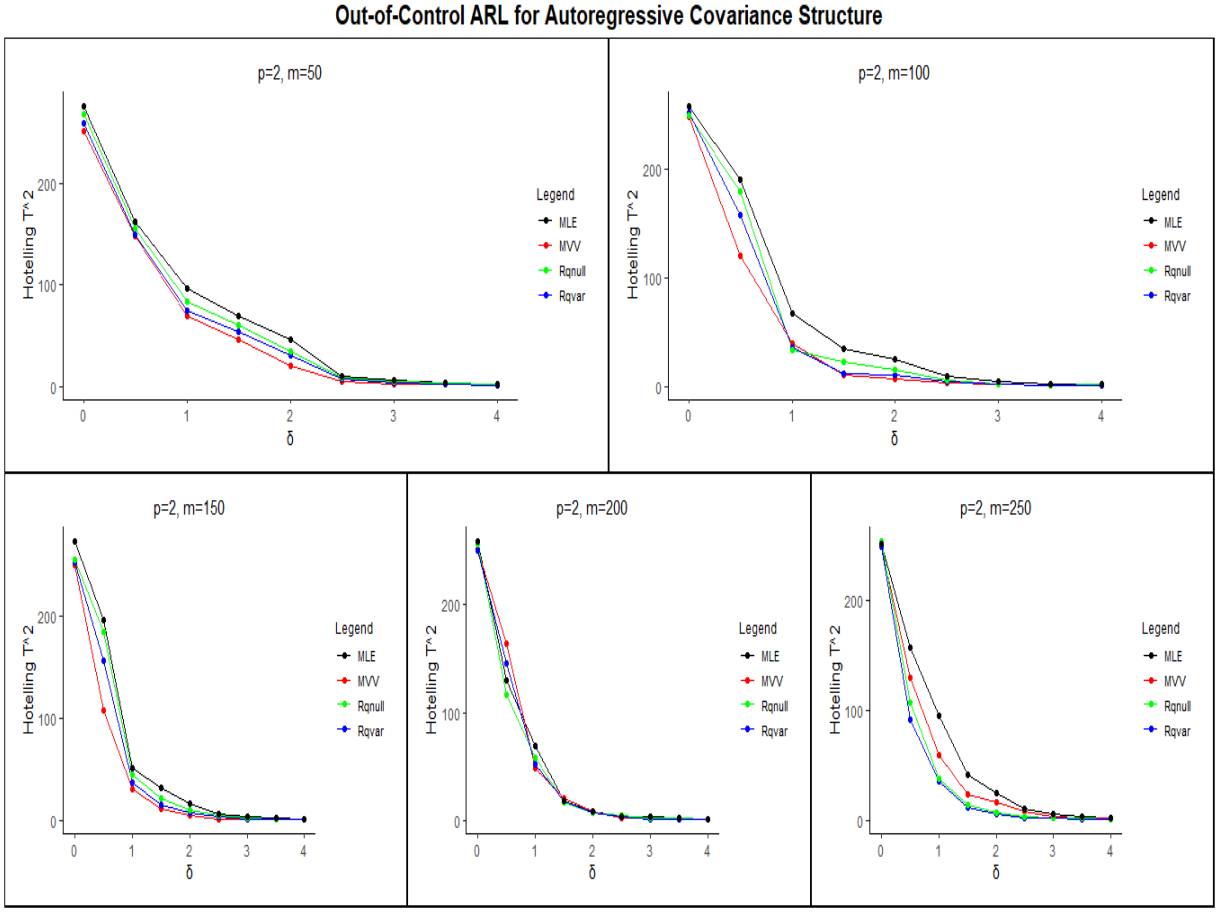
Table 2: ARL profile of the auto-regressive covariance structure when M = 150 and 200 for different p values

P	δ	M = 150				M = 200			
		MLE	**MVV**	**Rqv**	**Rqnull**	**MLE**	**MVV**	**Rqv**	**Rqnull**
2	0.0	272.85	248.97	251.76	254.13	258.51	251.19	249.12	256.01
	0.5	195.05	107.74	156.07	184.75	130.26	163.70	145.97	116.15
	1.0	51.81	31.36	36.79	45.04	69.69	48.19	52.78	58.76
	1.5	32.36	11.34	15.70	21.92	18.85	21.33	19.02	16.80
	2.0	16.90	5.17	7.56	10.38	8.56	8.79	7.84	7.63
	2.5	7.18	2.13	3.58	5.83	4.45	3.12	4.04	5.27
	3.0	3.99	1.88	2.17	2.77	3.69	2.48	2.21	3.22
	3.5	2.55	1.23	1.72	1.99	2.48	1.39	1.71	2.70
	4.0	2.00	1.00	1.25	1.07	2.01	1.11	1.57	1.87
5	0.0	257.07	252.12	251.35	258.85	254.03	256.47	250.77	256.64
	0.5	193.49	107.08	167.65	183.37	186.20	100.08	159.24	177.70
	1.0	56.01	30.19	34.95	49.94	65.53	40.12	48.33	55.60
	1.5	23.96	10.77	14.06	20.36	26.63	16.04	19.65	22.61
	2.0	11.10	5.15	7.38	9.90	15.94	7.35	10.12	12.98
	2.5	7.79	3.49	3.07	6.16	8.41	4.81	5.18	6.94
	3.0	3.91	2.05	2.94	3.95	3.55	2.88	3.23	3.68
	3.5	2.10	1.26	1.89	2.67	2.72	1.10	1.88	2.03
	4.0	1.54	1.00	1.15	1.37	2.04	1.07	1.21	1.89
10	**0.0**	263.81	255.79	262.17	256.90	251.48	257.65	250.00	272.65
	0.5	197.64	185.79	195.42	199.40	191.04	181.02	172.16	163.34
	1.0	43.97	42.72	38.10	39.21	163.50	158.17	151.87	138.79
	1.5	28.31	19.52	17.40	25.24	32.66	29.81	27.47	20.20
	2.0	16.31	9.41	8.39	14.54	18.72	15.77	14.06	10.45
	2.5	9.49	4.89	4.36	8.46	10.09	5.52	8.59	7.43
	3.0	5.66	2.94	2.62	5.05	6.66	3.04	4.05	4.86
	3.5	3.27	2.00	1.78	2.92	4.33	2.05	2.14	2.97
	4.0	2.95	1.03	1.78	2.01	2.73	1.52	1.71	1.95
15	**0.0**	238.25	268.16	260.75	254.14	262.97	273.62	265.64	256.15
	0.5	186.65	156.09	150.14	132.68	57.54	62.91	56.09	51.31
	1.0	91.74	81.30	65.23	55.52	39.35	40.60	36.20	35.08
	1.5	42.98	24.16	22.63	12.49	23.38	24.48	21.83	20.85
	2.0	15.13	13.33	7.54	8.38	12.00	13.91	12.40	10.70
	2.5	9.53	3.10	4.15	7.00	7.46	8.42	7.51	6.65
	3.0	6.59	2.47	3.20	5.04	4.45	4.78	4.26	3.97
	3.5	4.23	1.79	2.39	3.77	2.77	2.80	2.50	2.47
	4.0	2.87	1.04	1.80	2.56	1.93	1.90	1.71	1.72
20	**0.0**	254.98	245.90	249.95	252.27	248.70	251.10	260.65	253.43
	0.5	155.20	141.74	122.29	137.47	146.14	138.90	134.68	141.15
	1.0	81.01	71.72	41.72	60.84	53.61	42.77	21.76	29.97
	1.5	50.50	46.91	25.08	40.82	23.86	21.05	16.10	18.28
	2.0	21.49	15.70	8.65	10.24	15.77	14.60	8.35	10.06
	2.5	7.58	6.94	3.30	4.76	10.23	8.68	5.95	6.23
	3.0	5.36	4.95	3.04	4.08	7.81	6.18	3.73	5.18
	3.5	3.61	3.08	2.06	2.91	4.78	3.77	2.07	2.87
	4.0	2.76	2.43	1.24	2.09	2.56	2.05	1.03	1.98
25	**0.0**	257.50	253.41	270.11	255.61	249.78	267.66	249.58	252.21
	0.5	157.47	117.43	124.46	115.58	183.00	167.18	123.15	150.51
	1.0	95.43	72.80	56.33	63.76	100.57	87.23	49.28	59.45
	1.5	29.74	18.66	10.18	13.47	18.17	16.20	10.23	13.52
	2.0	9.06	8.94	5.97	6.08	12.05	11.61	6.24	8.96
	2.5	6.65	5.21	2.64	4.93	7.99	6.91	3.05	3.96
	3.0	4.68	4.13	2.07	3.17	3.94	3.16	2.60	3.10
	3.5	3.37	2.49	1.82	2.10	3.55	3.07	2.01	2.72
	4.0	2.98	2.18	1.07	2.01	2.73	2.45	1.18	2.03

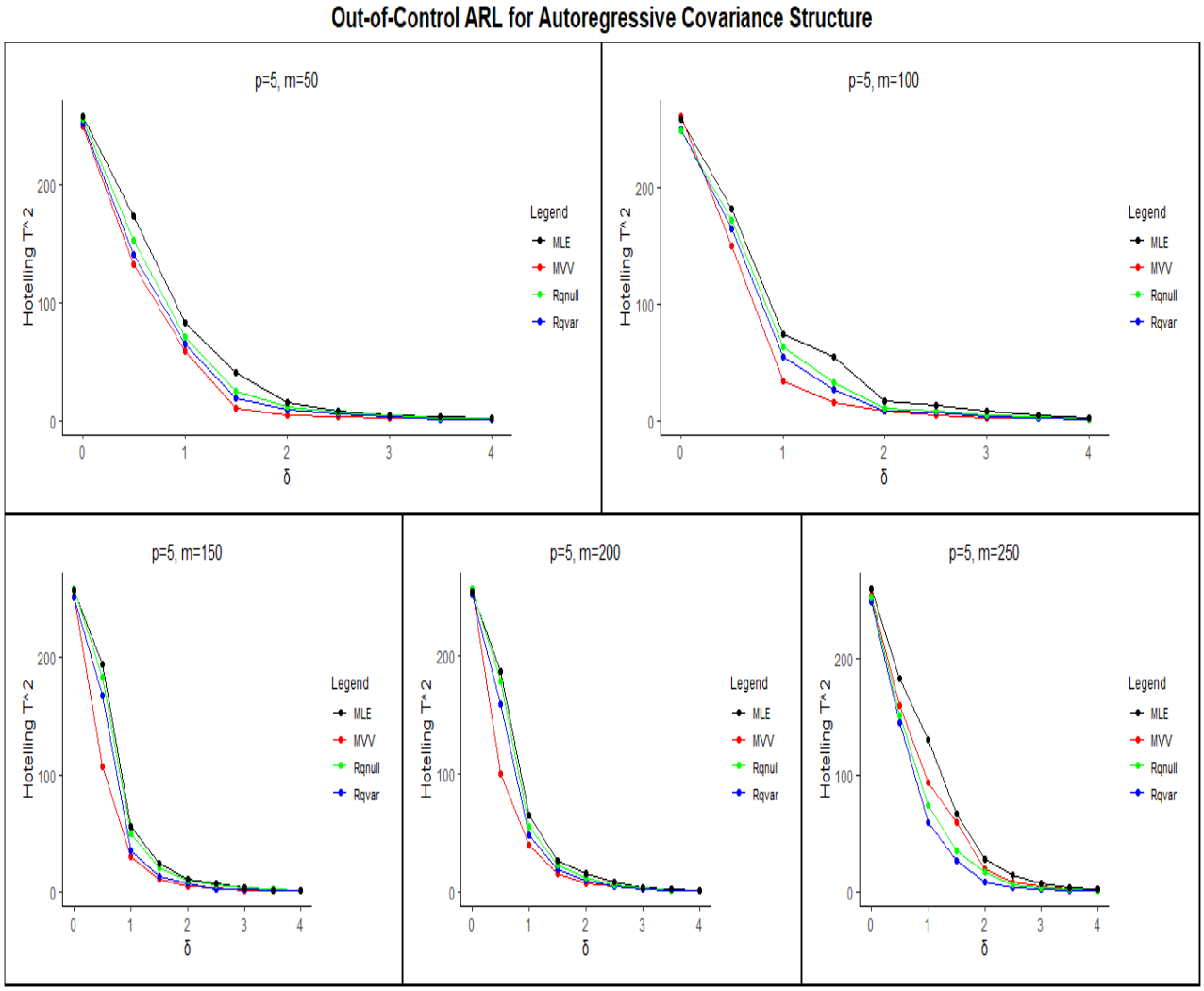
Table 3: ARL profile of the auto-regressive covariance structure when M = 250 for different p values

P	δ	M = 250			
		MLE	MVV	Rqv	Rqnull
2	0.0	251.94	249.36	249.46	254.41
	0.5	157.67	129.94	91.67	107.92
	1.0	95.77	59.91	35.77	38.25
	1.5	42.16	23.58	12.16	14.94
	2.0	25.80	16.49	5.80	7.69
	2.5	10.98	8.34	2.98	4.50
	3.0	6.88	4.11	2.18	3.16
	3.5	3.91	2.97	1.31	2.60
	4.0	2.86	2.19	1.06	1.58
5	0.0	259.96	253.39	248.10	251.82
	0.5	182.96	160.14	144.96	150.75
	1.0	130.16	93.65	60.16	74.93
	1.5	67.53	59.57	27.53	35.00
	2.0	28.83	19.87	8.83	17.10
	2.5	14.94	8.52	3.94	5.83
	3.0	8.06	5.42	3.06	4.34
	3.5	3.96	3.11	1.76	2.71
	4.0	2.72	2.09	1.02	1.89
10	**0.0**	250.14	248.42	252.28	253.13
	0.5	159.28	142.68	109.28	123.70
	1.0	100.18	90.19	70.18	83.15
	1.5	80.07	59.03	38.07	40.62
	2.0	45.53	25.07	14.53	17.80
	2.5	20.69	13.01	9.69	11.23
	3.0	11.81	9.02	4.81	6.02
	3.5	6.37	4.53	2.37	2.53
	4.0	3.09	2.93	1.09	2.05
15	**0.0**	250.96	248.39	250.10	253.82
	0.5	184.96	160.14	144.96	152.75
	1.0	130.16	103.65	85.16	70.93
	1.5	90.53	70.57	47.53	35.00
	2.0	68.83	49.87	20.83	17.10
	2.5	44.94	25.52	9.94	7.83
	3.0	13.06	8.42	3.06	2.34
	3.5	5.06	3.31	2.06	1.61
	4.0	3.42	2.59	1.42	1.10
20	**0.0**	249.98	248.29	249.89	252.67
	0.5	175.42	150.60	127.42	120.90
	1.0	92.19	74.77	49.19	39.38
	1.5	46.17	38.06	16.17	10.94
	2.0	21.23	12.55	5.12	3.74
	2.5	9.30	6.16	3.30	2.40
	3.0	4.53	3.77	2.93	1.80
	3.5	3.04	3.00	2.04	1.38
	4.0	2.31	1.96	1.00	1.01
25	**0.0**	250.80	248.07	250.52	253.35
	0.5	164.34	110.59	94.34	88.74
	1.0	121.62	46.42	41.62	36.87
	1.5	77.88	31.11	27.88	20.93
	2.0	46.25	18.15	15.25	11.73
	2.5	20.04	11.22	7.04	5.89
	3.0	15.98	6.69	3.98	3.84
	3.5	4.77	4.21	2.77	1.46
	4.0	3.55	2.86	1.55	1.22

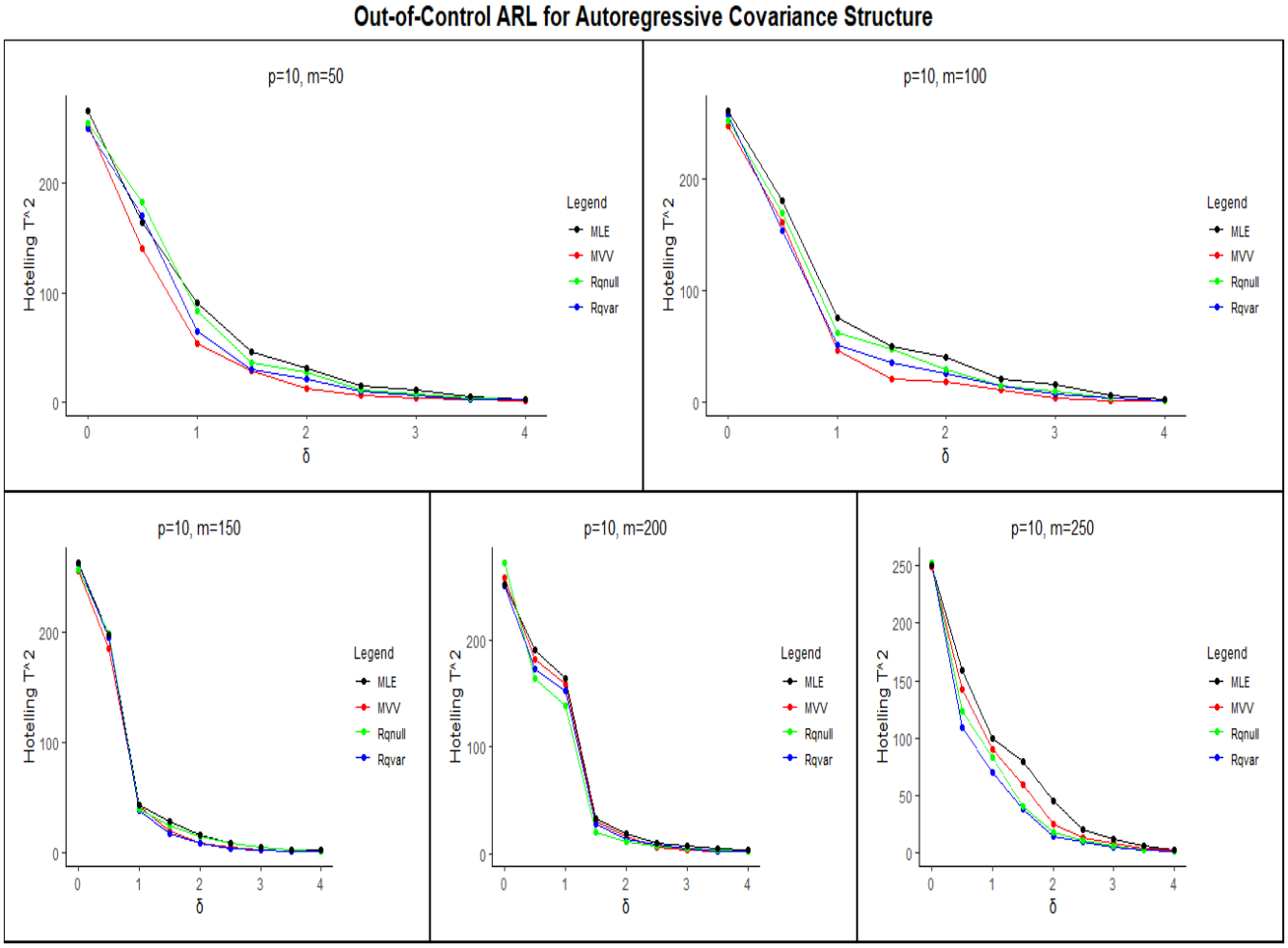
To better appreciate the results in Tables 1-3, the visual display of the ARL profile for the different estimates of the covariance matrix are presented in Figure 1-Panels (a)-(f).



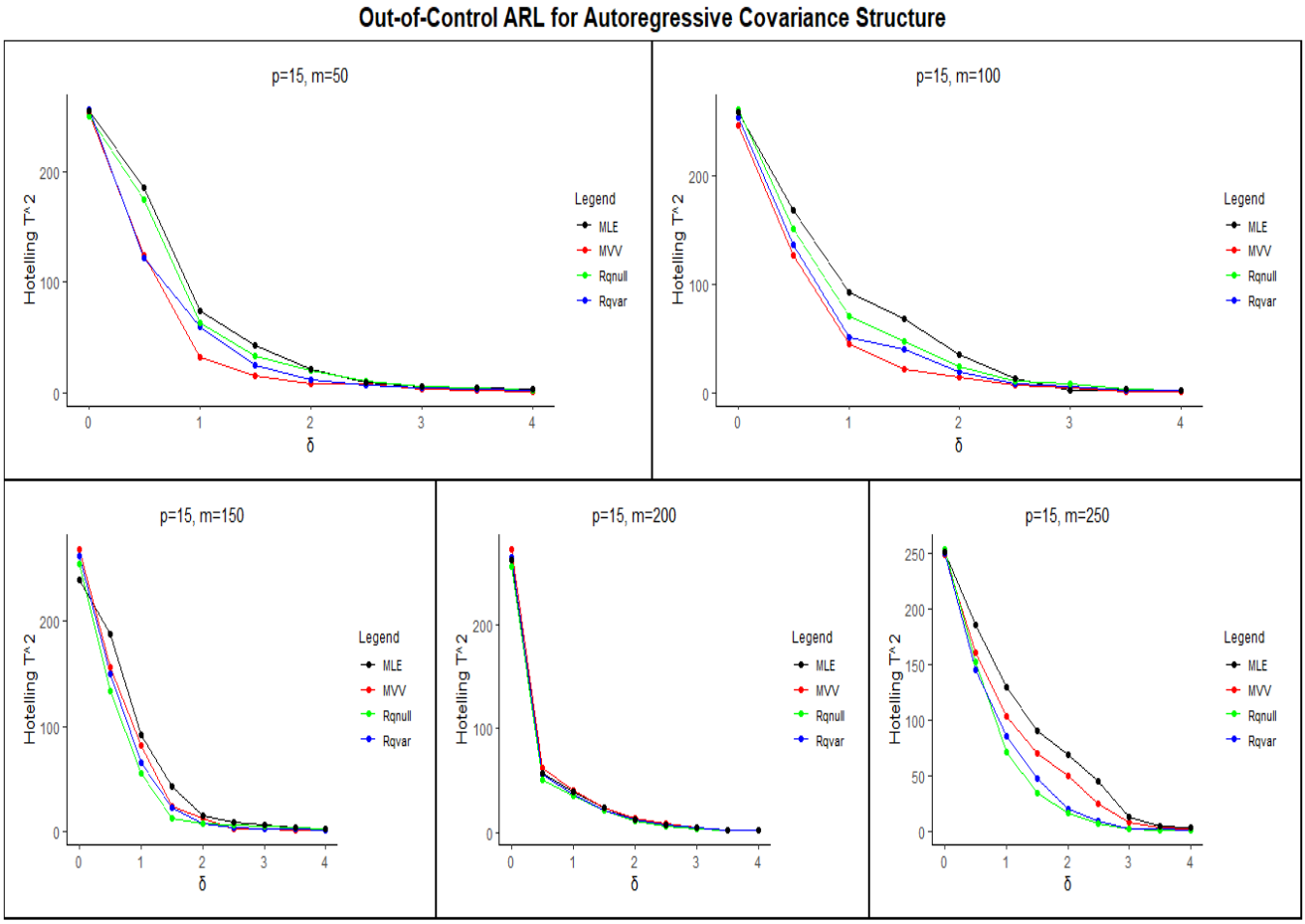
Panel (a): $p = 2$


Out-of-Control ARL for Autoregressive Covariance Structure
p=5, m=50
p=5, m=100
p=5, m=150
p=5, m=200
p=5, m=250
Hotelling T^2
δ
Legend
MLE
MVV
Rqnull
Rqvar

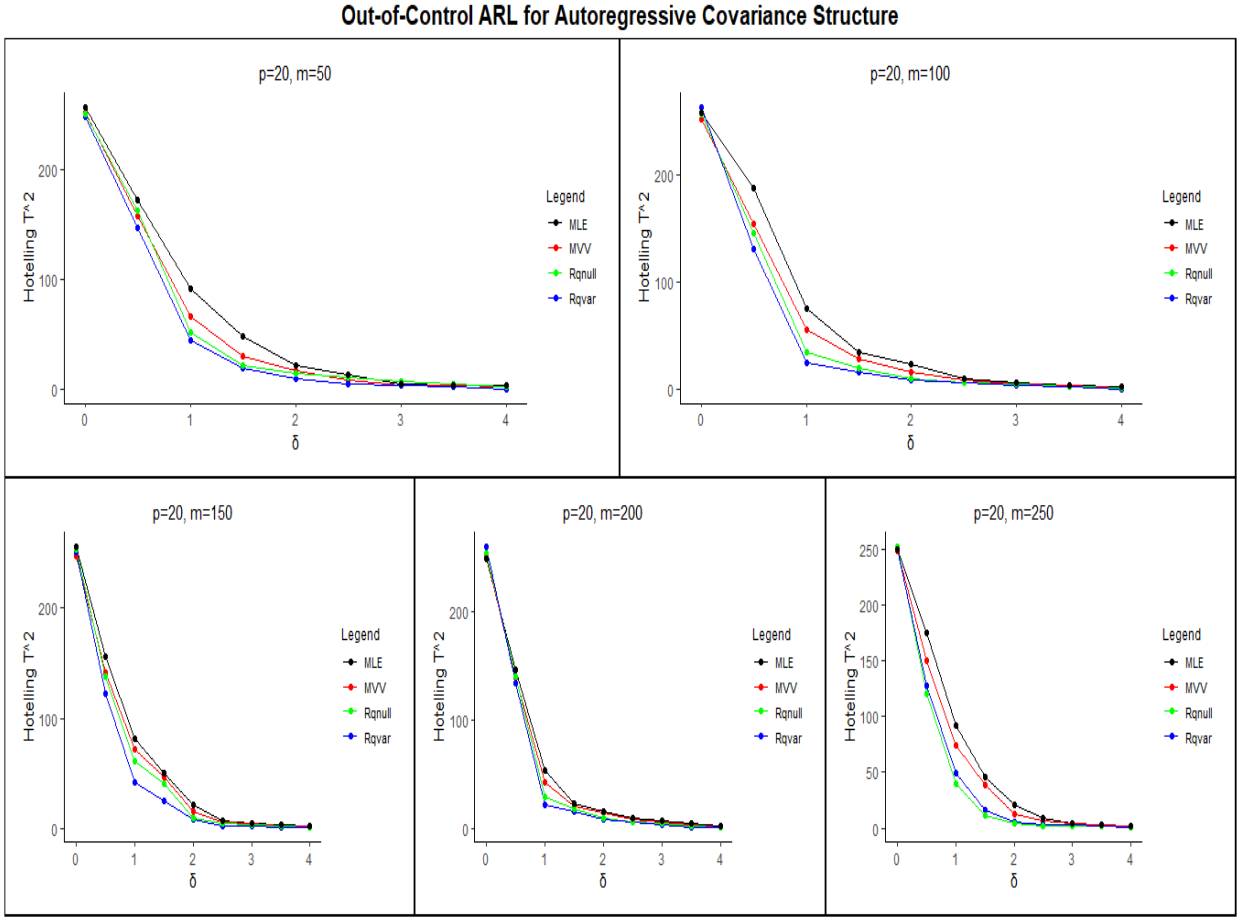
Panel (b): $p = 5$


Out-of-Control ARL for Autoregressive Covariance Structure
p=10, m=50
p=10, m=100
p=10, m=150
p=10, m=200
p=10, m=250
Hotelling T^2
δ
Legend
MLE
MVV
Rqnull
Rqvar

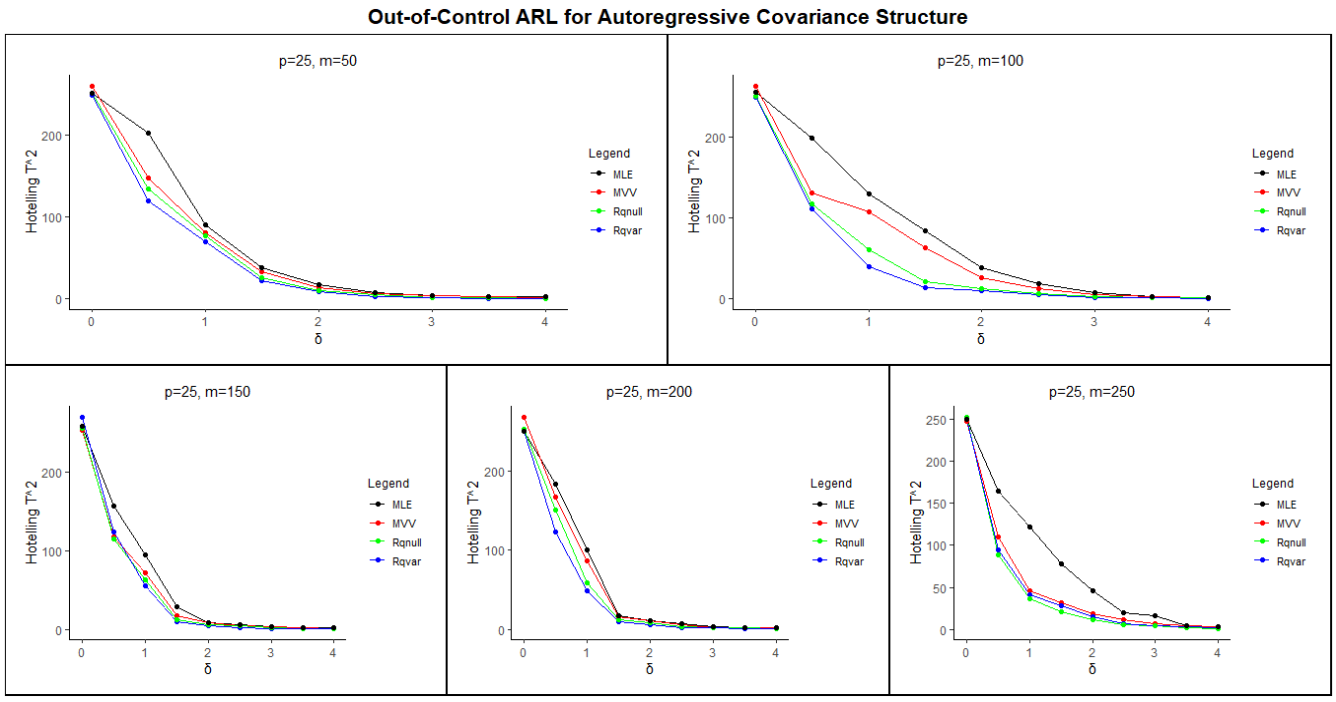
Panel (c): $p = 10$


Out-of-Control ARL for Autoregressive Covariance Structure
p=15, m=50
p=15, m=100
p=15, m=150
p=15, m=200
p=15, m=250
Hotelling T^2
δ
Legend
MLE
MVV
Rqnull
Rqvar

Panel (d): $p = 15$



Panel (e): $p = 20$



Panel (f): $p = 25$

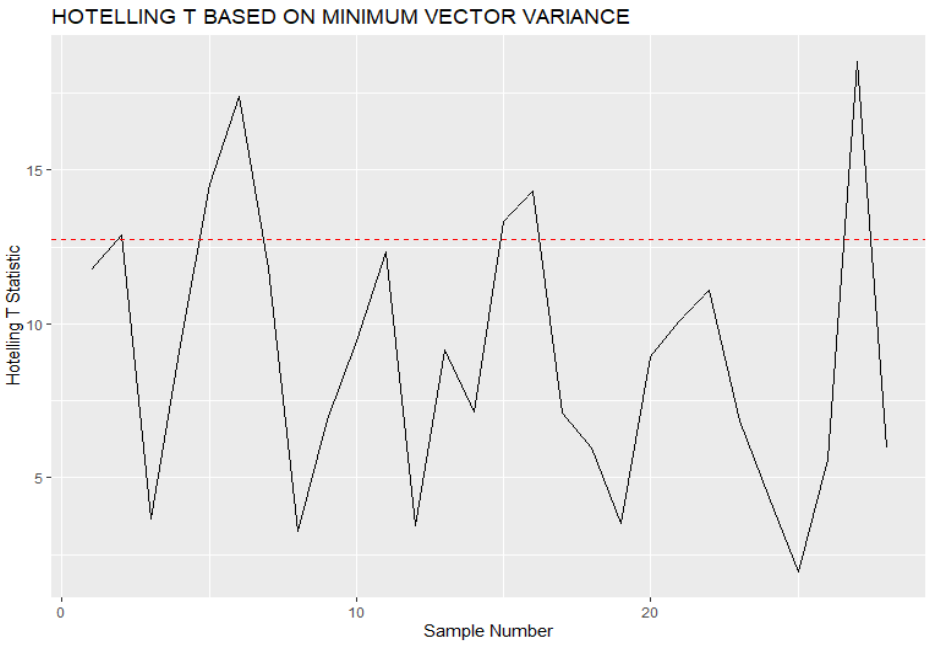
Figure 1: Sensitivity of the proposed charts in terms ARL profile for different values of p and M for a prespecified ARL_0 of 250

4.3 Example with real life data

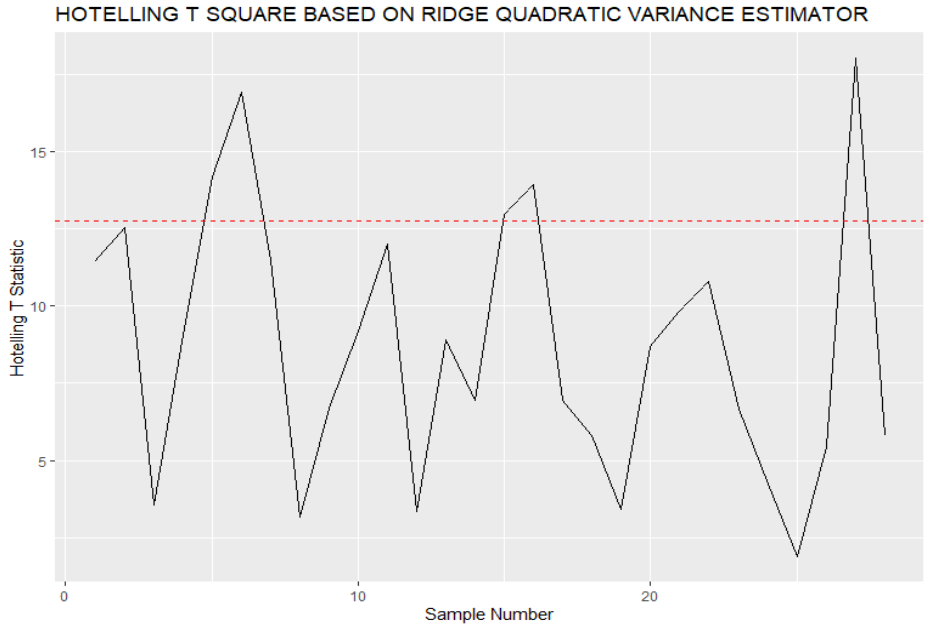
In this section, we consider the example of the bimetal thermostat dataset taken from [26] to compare the performance of Hotelling T^2 chart based on proposed estimators. The dataset consists of five dimensions, which are: the deflection, curvature, resistivity, and hardness respectively of the low and high-expansion sides of the brass and the steel bimetal thermostats. The bimetal thermostat dataset is presented in Appendix A. Phase I data consist of a sample size $M = 28$ with five quality characteristics (i.e., p = 5). The Phase I process is employed in order to study a historical reference sample, which includes establishing the in-control state and assessing the process steadiness to ensure that the reference sample is true representative of the process. The results are displayed in Table 4 below using the four estimators. The plots of MLE, MVV, Rqv and Rqnull Hotelling T^2 based charts are shown in the Figure 2.

Table 4: Hoteling T^2 Value Based on the Four Estimators

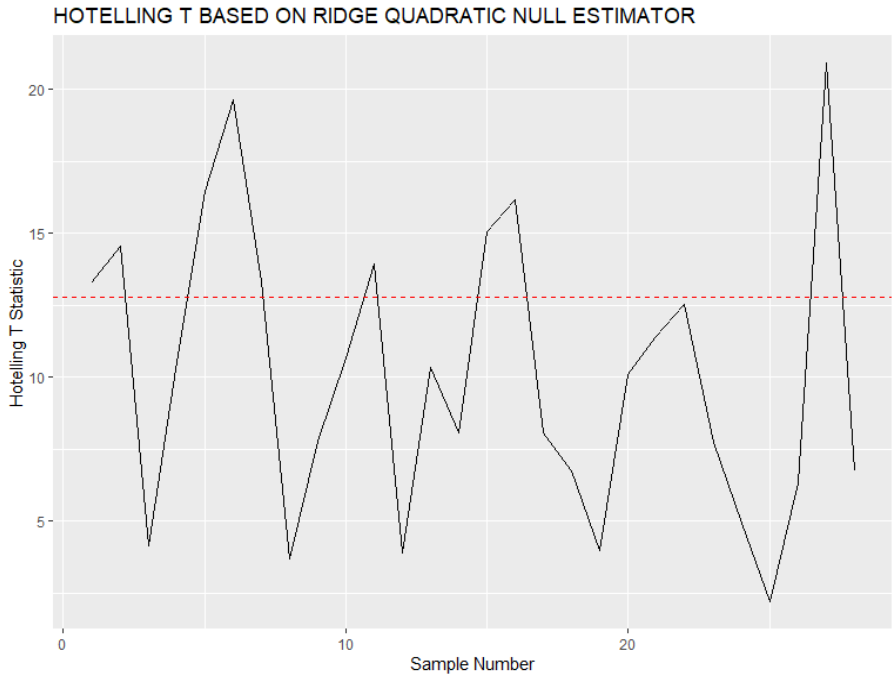
S/N	MLE	MVV	RQV	RQNULL
1	9.2262	11.7400	11.4326	13.2651
2	10.1282	12.8900	12.5502	14.5620
3	2.8779	3.6600	3.5661	4.1377
4	7.3314	9.3300	9.0846	10.5408
5	11.4304	14.5500	14.1639	16.4342
6	13.6479	17.3700	16.9117	19.6226
7	9.2468	11.7700	11.4581	13.2947
8	2.5591	3.2600	3.1711	3.6794
9	5.4038	6.8800	6.6961	7.7694
10	7.4233	9.4500	9.1986	10.6730
11	9.6862	12.3300	12.0025	13.9265
12	2.7061	3.4400	3.3532	3.8907
13	7.1918	9.1500	8.9117	10.3402
14	5.6118	7.1400	6.9538	8.0684
15	10.4568	13.3100	12.9575	15.0345
16	11.2396	14.3100	13.9275	16.1600
17	5.5933	7.1200	6.9309	8.0418
18	4.6989	5.9800	5.8226	6.7560
19	2.7588	3.5100	3.4186	3.9665
20	7.0172	8.9300	8.6953	10.0891
21	7.9341	10.1000	9.8315	11.4074
22	8.7175	11.1000	10.8023	12.5338
23	5.3855	6.8600	6.6734	7.7431
24	3.4130	4.3400	4.2292	4.9071
25	1.5161	1.9300	1.8787	2.1798
26	4.3904	5.5900	5.4403	6.3123
27	14.5533	18.5300	18.0336	20.9243
28	4.6946	5.9800	5.8173	6.7497


HOTELLING T BASED ON MINIMUM VECTOR VARIANCE
Hotelling T Statistic
15
10
5
0
10
20
Sample Number

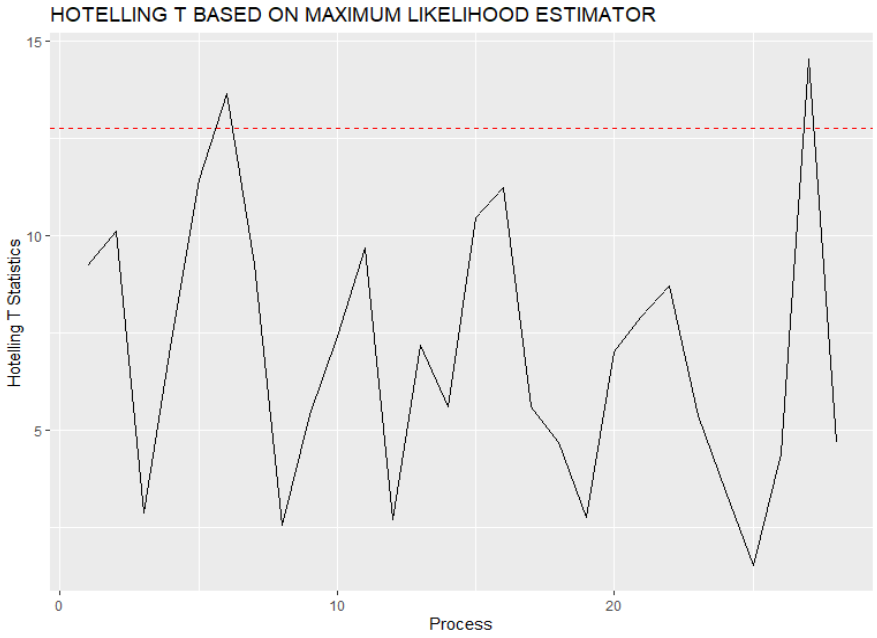
(a) Mvv


HOTELLING T SQUARE BASED ON RIDGE QUADRATIC VARIANCE ESTIMATOR
Hotelling T Statistic
15
10
5
0
10
20
Sample Number

(b) Rqv



(c) Rqnull



(d) MLE

Figure 2: Application of the proposed control charts based on the MLE, MVV, Rqnull and Rqv estimators using real life data

4. 4 Summary of Results

It has been shown from the Tables 1-3 that Hotelling T^2 based on MVV performs best when the value of p = 2 and the values of M =50, 100, 150 and 200 irrespective of the magnitude of the shifts then, follow by Hotelling T^2 based on Rqv and the Rqnull. Also, when the values of p = 5 and 10, Hotelling T^2 based on MVV performs better when M = 50, 100, 150 and 200;

but, when p = 15 Hotelling T^2 based on Rqnull performs best when the values of the shifts are small (0.5, 1.0 and 1.5) and Hotelling T^2 based on MVV performs better when the shift in process mean is very large. For larger p values (i.e., p =20 and 25), irrespective of the values of M and the magnitudes of the shifts, the Hotelling T^2 based on Rqv performs best follows by Rqnull.

Although the MLE is computationally faster, however, it performs poorly in a situation where quality characteristics (P) are more than the sample size (n). The regularization estimators considered in this study (Rqvar and Rqnull) and the MVV performed better in terms of bias-variance trade-off. This is because, the MVV estimator introduces bias, but reduces the variance of a process. Thus, the MVV efficiency is greater than that of the MLE. Furthermore, MLE is prone to overfitting when dealing with high dimensionality while regularization and MVV estimators are both optimal for minimizing the variance and can reduce the problem of singularity observed in MLE for high dimensional processes. It should be noted that the availability of software for analysis and handling of computational complexity has drastically and significantly reduced the computational time and cost. This is evidence in the results obtained in this study.

5. Conclusion

This paper proposes a new Hotelling T^2 control chart based on MVV and regularized estimators (Ridge quadratic variance (Rqvar) and Ridge quadratic null (Rqnull)) for monitoring process mean vectors under the assumption of unknown process parameters. It was observed that the Hotelling T^2 control chart based on MVV performed better in many cases irrespective of the magnitude of the shifts, and the chart based on Rqvar and Rqnull performed better when the sample size is very large. The operators in the industrial and non-industrial organisations are advised to use these newly proposed control charts especially when dealing with high dimensional data sets. The researchers can also consider improving these charts by using adaptive techniques such as the variable sampling techniques and interval, and variable parameters. Furthermore, the MVV and regularized methods considered in this study can also be compared with some other modern techniques in future research.

References

[1] Abbas,N., Riaz, M. & Does, R. J. (2013). Mixed exponentially weighted moving average cumulative sum charts for process monitoring. *Quality and Reliability Engineering International*, 29 (3): 345 – 356.

[2] Olusola T. Omolofe, Nurudeen A. Adegoke, Olatunde Adeoti, Olusoga A. Fasorunbaku, and Saddam A. Abbasi (2021). Multivariate control charts for monitoring process mean vector of individual observations under regularized covariance estimation. *Quality Technology and Quantitative Management* 19(1):1-22

[3] Jinho Kim et al. (2019). Control charts for variability monitoring in high-dimensional processes. *Computers and Industrial Engineering* 130: 309 – 316

[4] Adegoke, N. A, Smith, A. N. H., Anderson M. J., and Pawley, M. D. M. (2019b) Efficient Homogenously Weighted Moving Average Chart. *IEEE*, 7:9586 – 9597.

[5] Pignatiello, Joseph J, and George C Runger (1990). Comparisons of Multivariate CUSUM Charts. *Journal of Quality Technology* 22 (3): 173-186.

[6] Hotelling, H. (1974). *Multivariate quality control-illustrated by air testing of sample bombsights, in Techniques of statistical analysis*. M. W. H. C. Eisenhartnd W. A. Wallis, Eds. New York, NY, USA: McGraw-Hill, 111-184

[7] Abbas Nasir. (2018). Homogenously weighted moving average control chart with an application in substrate manufacturing process. *Journal of Computer and Industrial Engineering*, 120:460 – 470.

[8] Maboudou-Tchao, Edgard M, and Vincent Agboto (2013). Monitoring the Covariance sMatrix with Fewer Observations than Variables. *Computational Statistics and Data Analysis* 64:99 – 12.

[9] Ledoit, Olivier, and Michael Wolf. (2004a). A well-conditioned Estimator for Large-Dimensional Covariance Matrices. *Journal of Multivariate Analysis* 88 (2): 365 – 411.

[10] Fisher, T. J and Sun, X. (2011). Improved stein-type shrinkage estimators for the high dimensional multivariate normal covariance matrix. *Computational Statistics and Data Analysis*55: 1909 – 1918.

[11] Ullah, I., Pawley, M.D.M., Smith, A.N.A., and Jones, B. (2017). Improving the Detection of Unusual Observations in High-Dimensional Settings. *Australian & New Zealand Journal of Statistics* 59 (4):449 – 462.

[12] Van Wieringen W.N., Peeters, C.F.W. (2016). Ridge Estimation of Inverse Covariance Matrices from High-Dimensional Data. *Computational Statistics and Data Analysis* 103:284 – 303.

[13] Li B., Wang K. and Yeh B, A., (2013) Monitoring the covariance matrix via penalized likelihood estimation. *IIE Transactions* 45(2):132 – 146.

[14] Adegoke, N A. Smith, A. N., H., Anderson M. J Abbasi, S.A and Pawley M. D. M. (2018). Shrinkage estimates of Covariance Matrices to Improve the performance of Multivariate Cumulative Sum Control Charts. *Computers and Industrial Engineering* 117: 207 – 216.

[15] Lian Hneg. (2011). Shrinkage Tuning Parameter Selection in Precision Matrices Estimation. *Journal of Statistical Planning and Inference* 141(8): 2839 – 2848.

[16] Yuan, Ming, and Yi Lin. (2007). Model Selection and Estimation in the Gaussian Graphical Model. *Biometrika* 94 (1): 19 – 35.

[17] Clavel, Julien, Leandro Aristide, and Helene Morlon. (2019). A Penalized Likelihood Framework for High-Dimensional Phylogenetic Comparative Methods and an Application to New-World Monkeys Brain Evolution. *Systematic Biology* 68(1): 93 – 116.

[18] Vujacic Ivan, Antonimo Abbruzzo, and Ernst Wit (2015). A Computationally Fast Alternative to Cross-Validation in Penalized Gaussian Graphical Models. *Journal of Statistical Computation and Simulation* 85(18):3628 – 3640.

[19] Hazlina Alli, Sharipah-Soaad Syed-Yahaya, and Omar Zurni(2013). Robust Hotelling T^2 Control Chart with Consistent Minimum Vector Variance. *Mathematical Problems in Engineering* 2013 (4), Article ID 401350. http://dx.doi.org/10.1155/2013/401350.

[20] Hazlina Alli et al. (2015). Enhancing Minimum Vector Variance Estimators Using Reweighted Scheme. *Far East Journal of Mathematical Sciences* 98(7): 819 – 830.

[21] Mohsen Ebadi, Shojaeddin Chenouri, Dennis K. J. Lin, and Stefan H. Steiner (2021). Statistical Monitoring of the Covariance Matrix in Multivariate Processes: A Literature Review.*Journal of Quality Technology*54(3):269–289.

[22] Wessel N. Van Wieringen (2017). On the mean squared error of the ridge estimator of the covariance and precision matrix. *Statistics & Probability Letters* 123: 88 – 92.

[23] Dunn B, Paulish T, Stanbery A, Piotrowski J, Koniges G, Kroll E, Louis EJ, Liti G, Sherlock G, Rosenzweig F (2013). Recurrent rearrangement during adaptive evolution in an interspecific yeast hybrid suggests a model for rapid introgression. *PLoS Genet* 9(3): e1003366

[24] Yáñez S., González, N., and Vargas, J.A. (2010). Hotelling's T^2 control charts based on robust estimators. *Dyna*77 (163): 239 – 247.

[25] Rojas-Preciado, W., Rojas-Campuzano, M., Galindo-Villardón, P., Ruiz-Barzola, O. (2023). Control Chart T2Qv for Statistical Control of Multivariate Processes with Qualitative Variables. *Mathematics* 2023, *11*(12): 2595. https://doi.org/10.3390/ math11122595

[26] Edgar Santos-Ferna´ndez (2012). Multivariate Statistical Quality Control Using R. New-York, NY, USA: Springer.

Appendix A

Bimetal Thermostat Dataset

Sample	Deflection	Curvature	Resistivity	Hardness low expansion side	Hardness high expansion side
1	21.15	40.24	14.95	22.24	26.24
2	21.10	39.99	14.79	21.62	25.92
3	20.95	39.82	14.91	22.04	25.95
4	21.03	40.01	14.89	21.74	26.19
5	21.21	40.03	15.03	22.32	2586
6	21.37	40.31	15.21	22.03	26.08
7	20.70	39.90	14.75	21.67	25.86
8	20.87	39.89	15.04	21.89	26.02
9	21.27	40.14	15.20	22.27	26.23
10	20.97	40.13	14.98	22.11	26.22
11	21.34	40.20	14.91	21.99	25.89
12	20.92	39.87	14.90	21.76	25.93
13	20.83	40.00	15.15	22.20	26.02
14	20.84	39.90	15.06	22.08	26.07
15	20.95	40.16	14.97	22.20	26.25
16	20.75	39.80	14.71	22.01	25.66
17	21.00	40.05	15.10	22.36	26.10
18	21.21	40.26	15.05	22.15	26.17
19	21.03	39.87	14.98	22.05	26.07
20	21.01	39.84	14.97	21.89	26.19
21	21.08	40.00	14.78	22.20	25.90
22	21.08	39.78	14.96	22.02	26.09
23	20.69	39.77	14.92	21.91	25.87
24	20.88	39.85	15.00	21.79	26.00
25	21.01	40.02	15.06	21.92	26.08
26	21.01	39.95	14.78	22.02	25.86
27	21.07	40.08	15.40	22.15	26.06
28	20.97	39.87	14.99	21.77	25.91