International Journal of Applied Engineering Research ISSN 0973-4562 Volume 12, Number 16 (2017) pp. 6241-6249

New Statistical Test for Quality Control in High Dimension Data Set

Shamshuritawati Sharif [1], Suzilah Ismail and Zurni Omar

School of Quantitative Sciences, UUM-College of Arts and Sciences,
Universiti Utara Malaysia, 06010 UUM Sintok, Kedah, Malaysia.

[1]0000-0002-3305-9661

Abstract

The existing statistical test for quality control such as Box's M test caters for number of sample size *n* larger than number of variables or dimensions *p* (*n*>*p*). However, in real world application of Small Medium Enterprise (SME), the number of variables or dimensions *p* can be larger than sample size *n* (*p*>*n*) which is known as high dimension data set. This is due to small number of daily productions which lead to small number of sample size (*n*) but high dimensions products (*p*). One of the examples is rubber gloves which rely on machine capacity or latex supply that limiting the daily productions (*n*) but involves many dimensions (*p*) of measurement such as the size of five different fingers, the width of the palm and wrist; the strength, number of holes and etc. Another drawback, once the samples are tested for quality control, they are discarded which is very wasteful if uses large sample size. Therefore, in this study we have developed a new statistical test known as S* test to accommodate high dimension data set (*p*>*n*). A simulation study was conducted in comparing the performance of Box's M test and S* test using power of test. Based on 10,000 replications and 5% significance level, the power of test indicated that S* test outperformed the Box's M test. Interestingly, when *n* is smaller than *p*, S* test still can be computed which proven it can be used for quality testing in high dimension data set.

INTRODUCTION

Quality control is a crucial approach in driving an industry such as Small Medium Enterprise (SME) to be more effective and competitive. The existing statistical test for quality control such as Box's M statistics focuses on number of sample size *n* larger than number of variables or dimensions *p* ($n \geq p$) (Djauhari, 2005; Djauhari, 2009; Wan Yusoff, 2013; Sharif, 2013). However, in real world application of Small Medium Enterprise (SME), the number of variables or dimensions *p* can be larger than sample size *n* (*p*>*n*) which is known as high dimension data set. The main reason is because of small number of daily productions which lead to selecting small sample size (*n*) for quality testing but involves high dimensions product (*p*).

An example is the quality testing of rubber gloves which requires many dimensions (*p*) of measurement such as the size of five different fingers, the width of the palm and wrist; the strength, number of holes and etc.; but limited number of daily productions in SME perhaps due to machine capacity or the shortage of latex supply. Thus, lead to selecting small sample size (*n*) for quality testing. Another drawback, once the samples are tested for quality control, they are not saleable and discarded which is very wasteful if uses large sample size.

Therefore, in this study we have developed a new statistical test known as S* test to accommodate high dimension data set (*p*>*n*). The following sections outline the development of the test, validation using simulation study based on power of test and conclusions.

DEVELOPMENT OF NEW STATISTICAL TEST (S* Test)

The asymptotic distribution of S is investigated for $p \geq 2$ using the proposition derived by Kollo and Rosen (2005) and theorem established by Anderson (2003).

Now, let $X_1, X_2, \ldots, X_n$ be a random sample of size n drawn from a p-variate normal distribution with covariance matrix Σ. The sample mean vector and sample covariance matrix are given by $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ and $S = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})(X_i - \bar{X})^t$ respectively.

Proposition 1 (Kollo and Rosen, 2005). $\sqrt{n-1}\{vec\,(S) - vec(\Sigma)\} \xrightarrow{d} N_{p^2}\,(0, \Pi)$

where

$\Pi = \left(I_{p^2} + \mathrm{K}\right)(\Sigma \otimes \Sigma)$,

$K = \sum_{i=1}^{p}\sum_{j=1}^{p} \mathrm{H}_{ij} \otimes \mathrm{H}_{ij}^{t}$ is the commutation matrix of size $(p^2 \times p^2)$ and H_{ij} is a matrix of size $(p \times p)$ where its (i, j)-th element is equal to 1 and 0 elsewhere.

Proof : $\therefore V(S) = \frac{1}{n-1}\left(I_{p^2} + K\right)(\Sigma \otimes \Sigma)$

For $p = 2$, let

$$\Sigma = \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{bmatrix}, S = \begin{bmatrix} s_{11} & s_{12} \\ s_{21} & s_{22} \end{bmatrix},$$

$$vec(S) = [s_{11} \quad s_{21} \quad s_{12} \quad s_{22}]^t,$$

$$V(S) = Var(vec(S)) = \begin{bmatrix} s_{11}s_{11} & s_{11}s_{21} & s_{11}s_{12} & s_{11}s_{22} \\ s_{21}s_{11} & s_{21}s_{21} & s_{21}s_{12} & s_{21}s_{22} \\ s_{12}s_{11} & s_{12}s_{21} & s_{12}s_{12} & s_{12}s_{22} \\ s_{22}s_{11} & s_{22}s_{21} & s_{22}s_{12} & s_{22}s_{22} \end{bmatrix},$$

$K = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$ with the size of $p^2 \times p^2$,

and $I_{p^2} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$.

Then,

$$\Pi = (I_{p^2} + K)\Sigma \otimes \Sigma$$

$$= \begin{bmatrix} 2 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 2 \end{bmatrix} \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{bmatrix} \otimes \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{bmatrix}$$

$$= \begin{bmatrix} 2 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 2 \end{bmatrix} \begin{bmatrix} \sigma_{11} \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{bmatrix} & \sigma_{12} \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{bmatrix} \\ \sigma_{21} \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{bmatrix} & \sigma_{22} \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{bmatrix} \end{bmatrix}$$

$$= \begin{bmatrix} 2 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 2 \end{bmatrix} \begin{bmatrix} \sigma_{11}\sigma_{11} & \sigma_{11}\sigma_{12} & \sigma_{12}\sigma_{11} & \sigma_{12}\sigma_{12} \\ \sigma_{11}\sigma_{21} & \sigma_{11}\sigma_{22} & \sigma_{12}\sigma_{21} & \sigma_{12}\sigma_{22} \\ \sigma_{21}\sigma_{11} & \sigma_{21}\sigma_{12} & \sigma_{22}\sigma_{11} & \sigma_{22}\sigma_{12} \\ \sigma_{21}\sigma_{21} & \sigma_{21}\sigma_{22} & \sigma_{22}\sigma_{21} & \sigma_{22}\sigma_{22} \end{bmatrix}$$

$$= \begin{bmatrix} \sigma_{11}\sigma_{11} + \sigma_{11}\sigma_{11} & \sigma_{11}\sigma_{12} + \sigma_{12}\sigma_{11} & \sigma_{11}\sigma_{12} + \sigma_{12}\sigma_{11} & \sigma_{12}\sigma_{12} + \sigma_{12}\sigma_{12} \\ \sigma_{11}\sigma_{21} + \sigma_{11}\sigma_{21} & \sigma_{11}\sigma_{22} + \sigma_{12}\sigma_{21} & \sigma_{11}\sigma_{22} + \sigma_{12}\sigma_{21} & \sigma_{12}\sigma_{22} + \sigma_{12}\sigma_{22} \\ \sigma_{21}\sigma_{11} + \sigma_{21}\sigma_{11} & \sigma_{21}\sigma_{12} + \sigma_{22}\sigma_{11} & \sigma_{21}\sigma_{12} + \sigma_{22}\sigma_{11} & \sigma_{22}\sigma_{12} + \sigma_{22}\sigma_{12} \\ \sigma_{21}\sigma_{21} + \sigma_{21}\sigma_{21} & \sigma_{21}\sigma_{22} + \sigma_{22}\sigma_{21} & \sigma_{21}\sigma_{22} + \sigma_{22}\sigma_{21} & \sigma_{22}\sigma_{22} + \sigma_{22}\sigma_{22} \end{bmatrix}$$

$$= \begin{bmatrix} V(s_{11}s_{11}) & Cov(s_{11}s_{21}) & Cov(s_{11}s_{12}) & Cov(s_{11}s_{22}) \\ Cov(s_{21}s_{11}) & V(s_{21}s_{21}) & Cov(s_{21}s_{12}) & Cov(s_{21}s_{22}) \\ Cov(s_{12}s_{11}) & Cov(s_{12}s_{21}) & V(s_{12}s_{12}) & Cov(s_{12}s_{22}) \\ Cov(s_{22}s_{11}) & Cov(s_{22}s_{21}) & Cov(s_{22}s_{12}) & V(s_{22}s_{22}) \end{bmatrix}$$

$$\therefore V(S) = \frac{1}{n-1}(I_{p^2} + K)(\Sigma \otimes \Sigma)$$

According to asymptotic distribution of $vec(S)$ in Proposition 1 and the following theorem,

Theorem 1 Anderson (2003, Theorem 4.2.3., p. 132) **:**

Let $\{U(n)\}$ be a sequence of p-component random vectors and b a fixed vector such that $\sqrt{n}\,[U(n) - b] \xrightarrow{d} N(0, \gamma)$ as $n \to \infty$. Let $f(u)$ be a vector-valued function of u such that each component $f_j(u)$ satisfies $\frac{\partial f_j(u)}{\partial u_i}|_{u=b} \neq 0$. If $\frac{\partial f_j(u)}{\partial u_i}|_{u=b}$ is the (i, j)–th component of Φ. Then $\sqrt{n}\,[f(U(n)) - f(b)] \xrightarrow{d} (0, \Phi^t \gamma\, \Phi)$.

The new proposition when Σ_0 is unknown is formulated on the basis of removing the duplication matrix from $vec(S)$ to $vec(S_L)$. Covariance matrix Σ is a $p \times p$ symmetric matrix where contains duplication or redundant elements since $\sigma_{ij} = \sigma_{ji}$, for all $(i \neq j)$. Let $vec(\Sigma_L)$ defined as a vector of size $r = \frac{1}{2}p(p+1)$ representing the lower triangular elements of Σ. The duplication matrix T_p is the matrix of size $p^2 \times \frac{1}{2}p(p+1)$ will be used to transform $vec(\Sigma)$ into $vec(\Sigma_L)$.

Next, we define $u(vec(S)) = T_p \times (vec(S)) = vec(S_L)$ to derived the asymptotic distribution of $vec(S_L)$ .Thus, the new proposition is,

Proposition 2. $\sqrt{n-1}\;(vec(S_L) - vec(\Sigma_L)) \xrightarrow{d} N_{p^2}(0, \phi)$

where

$$\phi = T_p{}^t(I_{p^2} + K)(\Sigma \otimes \Sigma)T_m, \text{ and}$$

$T_p = (a_{ij})$ is a matrix zero-one with p blocks as derived in Appendix.

To put into practice, using Theorem 3.4.2 in Mardia *et al.* (1979) and the asymptotic distribution of $vec(S_L)$ given in Proposition 2, we present two principal results of derivation into Proposition 3 and Proposition 4. These proposition are actually the Mahalanobis square distance between $v(S_L)$ and $v(\Sigma_L)$.

Proposition 3. $[vec(S_L) - vec(\Sigma_L)]^t \;\; \phi^{-1}[vec(S_L) - vec(\Sigma_L)] \xrightarrow{d} \chi_r^2$ where the degrees of freedom $r = \frac{1}{2}p(p+1)$.

Proposition 4. $[vec(S_{k,L}) - vec(\Sigma_{0,L})]^t \;\; \phi_0^{-1}[vec(S_{k,L}) - vec(\Sigma_{0,L})] \xrightarrow{d} \chi_r^2$ where the degrees of freedom $r = \frac{1}{2}p(p+1)$.

Let S_k be the covariance matrix related to the k-th independent sample of size drawn from $N_p(\mu, \Sigma_k)$; $k = 1, 2, \dots, g$. Consider the hypothesis testing $H_0: \Sigma_1 = \Sigma_2 = ... = \Sigma_g (= \Sigma_0$ say). Under H_0, if Σ_0 is known, then

Proposition 5. $\left[vec(S_{k,L}) - vec(\Sigma_{0,L})\right]^t \ \phi_0^{-1}\left[vec(S_{k,L}) - vec(\Sigma_{0,L})\right] \xrightarrow{d} \chi_r^2$

where

$S_{k,L}$ and $\Sigma_{0,L}$ are lower elements of S_k and Σ_0, respectively,

$\phi_0 = T_p{}^t\left(I_{p^2} + K\right)(\Sigma_0 \otimes \Sigma_0)T_p$ where $\Sigma_{0,d}$ is a diagonal matrix of Σ_0, and

$r = \frac{1}{2}p(p+1)$.

Now, we consider the hypothesis $H_0: \Sigma_1 = \Sigma_2 = ... = \Sigma_g (= \Sigma_0$ say) versus $H_1: \Sigma_i \neq \Sigma_j$ for at least one pair (*i*, *j*). Testing that hypothesis is equivalent to the repeated tests (Montgomery, 2005). Therefore the hypothesis testing can be written as, $H_0: \Sigma_k = \Sigma_0$ for all $k = 1, 2, \ldots, g$ versus $H_1: \Sigma_k \neq \Sigma_0$ for at least one k; $k = 1, 2, \ldots, g$. Therefore, proposed statistic, S^* is as follows

$$S^* = [vec(S_L) - vec(\Sigma_L)]^t \phi^{-1}[vec(S_L) - vec(\Sigma_L)] \xrightarrow{d} \chi_r^2$$

where,

S_L and Σ_L are lower elements of S_k and Σ_0, respectively

$\phi = T_p{}^t\left(I_{p^2} + K\right)(\Sigma_0 \otimes \Sigma_0)T_p$

$r = \frac{1}{2}p(p+1)$

The null hypothesis is rejected if $S^* > \chi_r^2$. In real case study, when is Σ unknown $\Sigma = \hat{\Sigma}$. Matrix $\hat{\Sigma}$ is estimated reference sample of covariance matrix, i.e pooled sample covariance matrix.

Next, a simulation study was conducted in comparing the performance of S* test and Box's M test using power of test. The Box's M test is as follows

$$M = N\ln|\bar{S}| - \sum_{i=1}^{g} n_i \ln|S_i| \quad \textbf{(1)}$$

where

$\bar{S} = \frac{1}{N}\sum_{i=1}^{g} n_i S_i$ is the pooled sample variance-covariance matrix,

S_i is the variance-covariance matrix calculated from the sample *i*,

g is the number of subgroup where the stability of matrices is hypothesized, and

$N = n_1 + n_2 + \cdots + n_g$; n_i = *i*-th sample size.

The power of test is used to compare between different statistical testing procedures where the most powerful statistical will contain the higher number of rejection of null hypothesis (Mittelhammer, 1996). According to Yue et al. (2002), the power can be estimated by,

$$\text{Power} = \frac{R_n}{N}$$

where

R_n is the number of experiments that fall in the rejection region and

N is the total number of repetition in simulation experiments.

The simulation study was executed based on 5000 replications to obtain the accuracy of power approximation.

In this study, the simulation is based on 10,000 replications, 5% significance level, several conditions which are different number of variables ($p = 20{,}30$), different number of sample sizes($n = 10, 20, 30, 40{,}50$),different covariance shift ($k = 1.0{,}1.05{,}1.10{,}1.15, 1.20$) and different correlations ($\rho = 0.2{,}0.5{,}0.7{,}0.9$).

Let $\sigma^2 = (\sigma_1, \sigma_2, \ldots, \sigma_p)$ represent the population standard deviation vector of the *p* variables where $\sigma_i = \sqrt{\sigma_{ii}}$ for $i = 1{,}2, \ldots, p$, and σ_{ii} is corresponding variance of the *i*-th variable. We consider in each data set, it consist of different the number of variables, as well as sample size, *n*. At the same time, for all pairs of variables and sample size will have the same correlations ρ . These ρ values are representatives of low, moderate, high and very high correlation. As an example, size of shift, $k = 1.05$ then, the hypothesis testing is, $H_0: \Sigma_m = \Sigma_0$ versus $H_1: \Sigma_m = \Sigma_1$ where $\Sigma_0 = I_p$ and Σ_1 are as follows,

$$\Sigma_1 = \begin{bmatrix} 1.05 & 0.5*1.05 & \cdots & 0.5*1.05 \\ 0.5*1.05 & 1.05 & \cdots & 0.5*1.05 \\ \vdots & \vdots & \ddots & \vdots \\ 0.5*1.05 & 0.5*1.05 & \cdots & 1.05 \end{bmatrix}; \rho = 0.5$$

Generally, the power of a test or known as $1 - \beta$ is the probability of correctly rejecting the null hypothesis when it is false. It actually refer to the sensitivity of the statistical test where the ability to detect a true. From the power of test, the sensitivity level of the S* test and Box's M test to the shift in covariance structure can be determined.

RESULTS AND DISCUSSIONS

Table 1 and 2 display the simulations results based on several conditions as listed previously. When $p = 20, n = 10, \rho = 0.70$ and $k = 1.00$, S* test has reach the maximum power of test but M test reveal the least power of test and only obtain the highest power of test when $p = 20, n = 30, \rho = 0.90$ and $k =$

1.00. These indicates S* test outperform M test at lower correlation (ρ) and smaller sample size (n). Interesting results obtain when $p = 30$and $n = 10$, no power of test can be computed for M test. It is very obvious S* test performs well than M test when $n < p$ but even better when $n > p$. S* test also shows sensitivity when there is a small shift from $k = 1.00$ to 1.05 as compared to $k = 1.00$ to 1.20. Although S* test has low power of test when $\rho = 0.20$ but still outperform M test. Overall S* test has high power of test when $\rho = 0.50$ and above for all p and sample sizes (n). This finding aligns with real world application regarding the inter-relationships of variables (p) where the variables are correlated. One example is regarding the rubber gloves where the measurements of the five fingers should be correlated as to reflect the proportionate measurements of human fingers.

CONCLUSIONS

In this study, we successfully developed a new test for quality control name as S* test for high dimension data set. The power of test reveals S* test outperformed Box's M test when sample size (n) less than number of variables or dimensions (p) which fulfilled the condition of high dimensions data set. This S* test can be used in real world settings when there is a need to select small sample size for quality testing especially in SME.

Table 1: Power of test when $p = 20$

ρ	k	$n = 10$		$n = 20$		$n = 30$		$n = 40$		$n = 50$	
		S* test	**M test**	**S* test**	**M test**	**S* test**	**M test**	**S* test**	**M test**	**S* test**	**M test**
0.2	1.00	0.111	0.0011	0.3554	0.0029	0.4641	0.0042	0.6658	0.0131	0.7891	0.0374
	1.05	0.0569	0.0022	0.1756	0.0038	0.2444	0.0111	0.3769	0.0167	0.5009	0.0512
	1.10	0.0207	0.0035	0.0681	0.0042	0.1109	0.0106	0.1527	0.0298	0.2235	0.0715
	1.15	0.0065	0.0052	0.0221	0.0055	0.0399	0.0151	0.045	0.0459	0.0579	0.1124
	1.20	0.0035	0.0069	0.0078	0.0055	0.0117	0.0219	0.008	0.0709	0.0129	0.1574
0.5	1.00	0.9964	0.0002	1	0.0034	1	0.0217	1	0.5243	1	0.9953
	1.05	0.9894	0.0003	1	0.0026	1	0.0208	1	0.4416	1	0.9843
	1.10	0.9741	0.0002	0.9999	0.0026	1	0.0198	1	0.3733	1	0.9642
	1.15	0.9473	0.0002	0.9995	0.0029	0.9997	0.02	1	0.3189	1	0.9351
	1.20	0.8875	0.0002	0.9982	0.0044	0.9992	0.0202	1	0.2961	1	0.901
0.7	1.00	**1**	**0.0001**	1	0.0044	1	0.8251	1	1	1	1
	1.05	1	0	1	0.0039	1	0.7143	1	1	1	1
	1.10	1	0	1	0.004	1	0.6175	1	1	1	1
	1.15	1	0	1	0.0045	1	0.5135	1	1	1	1
	1.20	1	0	1	0.0049	1	0.4304	1	1	1	1
0.9	1.00	1	0	1	0.3394	**1**	**1**	1	1	1	1
	1.05	1	0	1	0.2816	1	1	1	1	1	1
	1.10	1	0	1	0.2384	1	1	1	1	1	1
	1.15	1	0	1	0.2004	1	1	1	1	1	1
	1.20	1	0	1	0.1679	1	1	1	1	1	1

Table 2: Power of test when $p = 30$

ρ	k	$n = 10$		$n = 20$		$n = 30$		$n = 40$		$n = 50$	
		S* test	**M test**	**S* test**	**M test**	**S* test**	**M test**	**S* test**	**M test**	**S* test**	**M test**
0.2	1.00	0.3083	-	0.6014	0.0011	0.8103	0.0023	0.9074	0.0049	0.9653	0.0123
	1.05	0.1455	-	0.3177	0.0019	0.5163	0.0025	0.6445	0.0069	0.7773	0.0247
	1.10	0.0652	-	0.116	0.0029	0.2263	0.0031	0.293	0.0124	0.416	0.046
	1.15	0.0211	-	0.0295	0.0048	0.0526	0.0032	0.0781	0.0175	0.1248	0.075
	1.20	0.0064	-	0.0051	0.0079	0.0104	0.0044	0.0127	0.0298	0.0194	0.2287
0.5	1.00	1	-	1	0.0003	1	0.0016	1	0.0087	1	0.1662
	1.05	1	-	1	0.0002	1	0.0021	1	0.0091	1	0.1328
	1.10	0.9996	-	1	0.0003	1	0.0013	1	0.0077	1	0.112
	1.15	0.999	-	1	0.0002	1	0.0015	1	0.0088	1	0.1214
	1.20	0.9943	-	0.9999	0.0004	1	0.0019	1	0.0084	1	0.0966
0.7	1.00	1	-	1	0.0001	1	0.0035	1	0.7298	1	1
	1.05	1	-	1	0.0001	1	0.0027	1	0.5882	1	1
	1.10	1	-	1	0.0001	1	0.0028	1	0.4474	1	1
	1.15	1	-	1	0	1	0.0028	1	0.3512	1	1
	1.20	1	-	1	0	1	0.0026	1	0.2615	1	1
0.9	1.00	1	-	1	0.0001	1	0.854	1	1	1	1
	1.05	1	-	1	0.0001	1	0.7596	1	1	1	1
	1.10	1	-	1	0.0001	1	0.6666	1	1	1	1
	1.15	1	-	1	0.0001	1	0.5685	1	1	1	1
	1.20	1	-	1	0.0001	1	0.4914	1	1	1	1

REFERENCES

[1] Anderson, T. W. (2003). *Introduction to Multivariate Statistical Analysis.* New York: John Wiley & Sons, Inc.

[2] Djauhari, M. A. (2005). Improved Monitoring of Multivariate Process Variability. *Journal of Quality Technology*, 37, pp. 32-39.

[3] Djauhari, M.A. (2009). Asymptotic Distribution of Sample Covariance Determinant. *Journal MATEMATIKA*, 25 (1).

[4] Djauhari, M.A. (2010b). A Multivariate Process Variability Monitoring Based on Individual Observations. *Journal of Modern Applied Science*, 4(10).

[5] Djauhari, M.A. (2011). Geometric Interpretation of Vector variance. *Journal MATEMATIKA*, 25(1).

[6] Kollo, T. & von Rosen, D. (2005). *Advanced Multivariate Statistics with Matrices*. The Netherlands: Springer.

[7] Mardia, K. V., Kent, J. T. and Bibby, J. M. (1979). *Multivariate Analysis.* London: Academic Press.

[8] Mittelhammer, R. C., & Mittelhammer, R. C. (1996). *Mathematical statistics for economics and business* (Vol. 78). New York, NY, USA: Springer.

[9] Montgomery, D.C. (2005). Introduction to statistical quality control, 5th Edition. John Wiley & Sons, Inc., New York.

[10] Sharif, S. (2013). A new statistic to the theory of correlation stability testing in financial market. (Unpublished PhD Thesis), Universiti Teknologi Malaysia, Johor, Malaysia

[11] Wan Yusoff, W.N.S. (2013). A new covariance structure stability test based on vector variance. (Unpublished PhD Thesis), Universiti Teknologi Malaysia, Johor, Malaysia

[12] Yue, Shang. Paul Pilon & George Cavadias. (2002). Power of the Mann-Kendall and Spearman's rho tests for detecting monotonic trends in hydrological series. *Journal of Hydrology, 259, 254-271.*

APPENDIX

Construction of Generalised Transformation Covariance Matrices

Case $p = 2$

$$S = \begin{bmatrix} s_{11} & s_{12} \\ s_{21} & s_{22} \end{bmatrix},\ V(S) = \begin{bmatrix} s_{11} \\ s_{21} \\ s_{12} \\ s_{22} \end{bmatrix} \in R^{4=2^2}.$$

Choosing the lower triangular elements of S, we construct the covariance vector $V(S_L)$ as follows

$$V(S_L) = \begin{bmatrix} s_{11} \\ s_{21} \\ s_{22} \end{bmatrix} \in R^{3=\frac{2(2+1)}{2}}.$$

Now,

$$T_2 V(S) = V(S_L)$$

Since the sizes of $V(S)$ and $V(S_L)$ are 4×1 and 3×1 respectively, T_2 must be of the size 3×4,i.e

$$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} s_{11} \\ s_{21} \\ s_{12} \\ s_{22} \end{bmatrix} = \begin{bmatrix} s_{11} \\ s_{21} \\ s_{22} \end{bmatrix}$$

We can further partition T_2 into two blocks as follows

$$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

The location of entries with element 1 for each block can be presented as below:

First block: $(1, p-1)$, $(2, p)$

Second block: $(3, 2p)$

<u>Case $p = 3$</u>

$$S = \begin{bmatrix} s_{11} & s_{12} & s_{13} \\ s_{21} & s_{22} & s_{23} \\ s_{31} & s_{32} & s_{33} \end{bmatrix},\ V(S) = \begin{bmatrix} s_{11} \\ s_{21} \\ s_{31} \\ s_{12} \\ s_{22} \\ s_{32} \\ s_{13} \\ s_{23} \\ s_{33} \end{bmatrix} \in R^{9=3^2}.$$

We then construct covariance $V(S_L)$ by choosing the lower triangular elements of S, i.e

$$V(S_L) = \begin{bmatrix} s_{11} \\ s_{21} \\ s_{31} \\ s_{22} \\ s_{32} \\ s_{33} \end{bmatrix} \in R^{6=\frac{3(3+1)}{2}}.$$

Since $T_3 V(S) = V(S_L)$ and the the sizes of $V(S)$ and $V(S_L)$ are now 9×1 and 6×1 respectively, T_3 must be of the size 6×9,i.e

$$\begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} s_{11} \\ s_{21} \\ s_{31} \\ s_{12} \\ s_{22} \\ s_{32} \\ s_{13} \\ s_{23} \\ s_{33} \end{bmatrix} = \begin{bmatrix} s_{11} \\ s_{21} \\ s_{31} \\ s_{22} \\ s_{32} \\ s_{33} \end{bmatrix}$$

We partition T_3 into three blocks as depicted below

$$\begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

whose the location of entries with element 1 is given by

First block: $(1, p-2)$, $(2, p-1)$, $(3, p)$

Second block: $(4,2p-1)$, $(5,2p)$

Third block: $(6,3p)$

<u>Case $p = 4$</u>

$$S = \begin{bmatrix} s_{11} & s_{12} & s_{13} & s_{14} \\ s_{21} & s_{22} & s_{23} & s_{24} \\ s_{31} & s_{32} & s_{33} & s_{34} \\ s_{41} & s_{42} & s_{43} & s_{44} \end{bmatrix},\ V(S) = \begin{bmatrix} s_{11} \\ s_{21} \\ s_{31} \\ s_{41} \\ s_{12} \\ s_{22} \\ s_{32} \\ s_{42} \\ s_{13} \\ s_{23} \\ s_{33} \\ s_{43} \\ s_{14} \\ s_{24} \\ s_{34} \\ s_{44} \end{bmatrix} \in R^{16=4^2}.$$

$V(S_L)$ is constructed using the same strategy as before, i.e by choosing the lower triangular elements of S

$$V(S_L) = \begin{bmatrix} s_{11} \\ s_{21} \\ s_{31} \\ s_{41} \\ s_{22} \\ s_{32} \\ s_{42} \\ s_{33} \\ s_{43} \\ s_{33} \end{bmatrix} \in R^{10=\frac{4(4+1)}{2}}.$$

Since $T_4 V(S) = V(S_L)$, T_4 has the size of 10×16 as the sizes of $V(S)$ and $V(S_L)$ are 16×1 and 10×1 respectively

$$\begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} s_{11} \\ s_{21} \\ s_{31} \\ s_{41} \\ s_{12} \\ s_{22} \\ s_{32} \\ s_{42} \\ s_{13} \\ s_{23} \\ s_{33} \\ s_{43} \\ s_{14} \\ s_{24} \\ s_{34} \\ s_{44} \end{bmatrix} = \begin{bmatrix} s_{11} \\ s_{21} \\ s_{31} \\ s_{41} \\ s_{22} \\ s_{32} \\ s_{42} \\ s_{33} \\ s_{43} \\ s_{33} \end{bmatrix}$$

Partition T_4 into four blocks as follows

$$\begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

The indices of entries with element 1 is given by

First block: $(1, p-3)$, $(2, p-2)$, $(3, p-1)$, $(4, p)$

International Journal of Applied Engineering Research ISSN 0973-4562 Volume 12, Number 16 (2017) pp. 6241-6249
© Research India Publications. http://www.ripublication.com

Second block: $(5,2p-2)$, $(6,2p-1)$, $(7,2p)$

Third block: $(8,3p-1)$, $(9,3p)$

Fourth block: $(10,4p)$

Using the same procedures as described previous we manage to obtain the indices of entries with element 1 for $p=5, p=6$ and $p=7$ as follows

<u>Case $p=5$</u>

First block: $(1,p-4)$, $(2,p-3)$, $(3,p-2)$, $(4,p-1)$, $(5,p)$

Second block: $(6,2p-3)$, $(7,2p-2)$, $(8,2p-1)$, $(9,2p)$

Third block: $(10,3p-2)$, $(11,3p-1)$, $(12,3p)$

Fourth block: $(13,4p-1)$, $(14,4p)$

Fifth block: $(15,5p)$

<u>Case $p=6$</u>

First block: $(1,p-5)$, $(2,p-4)$, $(3,p-3)$, $(4,p-2)$, $(5,p-1)$, $(6,p)$

Second block: $(7,2p-4)$, $(8,2p-3)$, $(9,2p-2)$, $(10,2p-1)$, $(11,2p)$

Third block: $(12,3p-3)$, $(13,3p-2)$, $(14,3p-1)$, $(15,3p)$

Fourth block: $(16,4p-2)$, $(17,4p-1)$, $(18,4p)$

Fifth block: $(19,5p-1)$, $(20,5p)$

Sixth block: $(21,6p)$

<u>Case $p=7$</u>

First block: $(1,p-6)$, $(2,p-5)$, $(3,p-4)$, $(4,p-3)$, $(5,p-2)$, $(6,p-1)$, $(7,p)$

Second block: $(8,2p-5)$, $(9,2p-4)$, $(10,2p-3)$, $(11,2p-2)$, $(12,2p-1)$, $(13,2p)$

Third block: $(14,3p-4)$, $(15,3p-3)$, $(16,3p-2)$, $(17,3p-1)$, $(18,3p)$

Fourth block: $(19,4p-3)$, $(20,4p-2)$, $(21,4p-1)$, $(22,4p)$

Fifth block: $(23,5p-2)$, $(23,5p-1)$, $(24,5p-2)$, $(25,5p)$

Sixth block: $(26,6p-1)$, $(27,6p)$

Seventh block: $(28,7p)$

The entries of T_p consist of 0 and 1 only. In general, for any $p \geq 2$, the indices of entries of T_p with element 1 is given by

<u>First block ((p) elements):</u>

$\big(1,p-(p-1)\big), \big(2,p-(p-2)\big), \big(3,p-(p-3)\big), \big(4,p-(p-4)\big), \big(5,p-(p-5)\big), \ldots, (p,p-(p-p)) = (p,p)$

<u>Second block ($(p-1)$ elements):</u>

$\big(p+1,2p-(p-2)\big), \big(p+2,2p-(p-3)\big), \big(p+3,2p-(p-4)\big), \big(p+4,2p-(p-5)\big), \ldots, \big(p+1+(p-1),2p-(p-p)\big) = (2p,2p)$

<u>Third block ($(p-2)$ elements):</u>

$\big(2p+1,3p-(p-3)\big), \big(2p+2,3p-(p-4)\big), \big(2p+3,3p-(p-5)\big), \ldots, \big(2p+(p-2),3p-(p-p)\big) = (3p-2,3p)$

Fourth block ($(p-3)$ elements):

$\big((3p-2)+1, 4p-(p-4)\big)$, $\big((3p-2)+2, 4p-(p-5)\big)$,
$\dots, \big((3p-2)+(p-3), 4p-(p-p)\big) = (4p-5, 4p)$

.

.

.

$(p-1)th$ block (2 elements):

$(23, 5p-2)$ $(23, 5p-1)$ $(24, 5p-2)$ $(25, 5p)$

$(p)th$ block (1 element):

$\left(\frac{p(p+1)}{2},\ pp+(p-p)\right) = \left(\frac{p(p+1)}{2}, p^2\right)$

Total number of entries having 1 as the element is

$p+(p-1)+(p-2)+(p-3)+\cdots+\big(p-(p-2)\big)+\big(p-(p-1)\big)$

which is an arithmetic series with $a = m$ and $d = -1$. The sum of this series that represents the total number of entries having 1 is

$$SUM_p = \frac{p}{2}[2p+(p-1)(-1)] = \frac{p}{2}(2p-p+1)$$
$$= \frac{p}{2}(p+1)$$

which proves that for size of covariance vector $V(S_L)$ is indeed $\frac{p}{2}(p+1)$.