Malaysian Journal of Mathematical Sciences 10(2) : 233-244 (2016)

MALAYSIAN JOURNAL OF MATHEMATICAL SCIENCES

Journal home page: http://einspem.upm.edu.my/journal

On Distributional Behavior of Jennrich's Statistic

Maman Abdurachman Djauhari

Institute for Mathematical Research (INSPEM),
Universiti Putra Malaysia, 43400 UPM Serdang, Selangor, Malaysia

E-mail: maman_abd@upm.edu.my

ABSTRACT

For testing the hypothesis that correlation structure is stable from sample to sample, Jennrich's statistic is the most appropriate tool. However, when the dynamics of that structure is of our interest, it becomes useless. This is due to a serious limitation it possesses; as a finite sum of independent statistics, the distribution of each term is unknown even asymptotically. On the other hand, understanding such dynamics is very important. It will enable us to monitor where and when the correlation structure has significantly been shifted. To overcome this problem, we introduce a correction factor for each term of that statistic and then derive the asymptotic distributional behavior of the corrected statistic.

Keywords: commutation matrix, Mahalanobis distance, multivariate normal distribution, *vec* operator

1. Introduction

Consider a sequence of m independent samples, each of which is drawn from a multivariate normal distribution with positive definite correlation matrix P_i ; i = 1, 2, …, m. One of the major problems in multivariate analysis is to identify the samples, if any, where correlation structure is shifted (Annaert et al., 2006). This is, in general, a problem of modeling the correlation structure dynamics (Rosenow et al., 2003, Onnela et al., 2005, Goetzmann et al., 2005 and Annaert et al., 2006).

The purpose of this paper is to develop a statistic for testing the occurrence of correlation structure dynamics and detecting where and when the instability takes place. We start by examining Jennrich's statistical test (Jennrich, 1970) which has been developed for testing correlation structure stability. What make it special, such that it is considered as the most appropriate statistical tool for that hypothesis testing, are its remarkable properties. Its asymptotic distribution is familiar and it is computationally efficient (Larntz and Perlman, 1985 and Deblauwe and Le, 2007).

Rage, 2003, Annaert et al., 2006 and Fischer, 2007, have remarked that the test becomes a standard tool in financial market analysis. It plays a vital role in testing the hypothesis of correlation structures stability (Deblauwe and Le, 2007 and Fischer, 2007). Unfortunately, if the hypothesis is rejected, it does not give any information about the dynamics of the structure. To reach our purpose, due to its remarkable properties mentioned above, Jennrich's statistic will be developed. An equivalent form will be proposed and, since it is a sum of independent statistics, a correction factor for each term will be introduced.

We start our discussion in Section 2 with the background that motivates this paper. Later on, in Section 3, a theorem on an equivalent form in terms of Mahalanobis distance is introduced. Then, a correction factor for each term is proposed in Section 4. It is the corrected version that allows us to analyze the correlation structure dynamics. To illustrate the adventages of the corrected statistic, an example is given in Section 5. Concluding remarks in the last section will be highlighted to close this presentation.

2. Background and Motivation

Testing the hypothesis of correlation structure stability is an active research area. In 1970, Jennrich introduced a statistic for testing that hypothesis. Nowadays it becomes popular as the most appropriate test (Deblauwe and Le, 2007). Actually, such hypothesis testing has a long history. We can see, for example, an early development in Hotelling, 1940, and Lawley, 1963 and Aitkin et al., 1968, for more recent works. In what follows we recall briefly that test and show its limitations, and then develop some ideas for improvement.

Suppose a sequence of *m* independent samples are available; each of which is drawn from $\mathcal{N}_p\left(\mu_i, \Sigma_i\right)$ with correlation matrix P_i where Σ_i is positive definite. Let n_i and R_i be the size and the correlation matrix of sample *i*. Jennrich's statistic is to test $\mathrm{H}_0 : \mathrm{P}_1 = \mathrm{P}_2 = \ldots = \mathrm{P}_m \; (= \mathrm{P}_0$, say). It is defined, see Jennrich, 1970, as

$$J = \sum_{i=1}^{m} J_i \tag{1}$$

where $J_i = \frac{1}{2} Tr\left(Z_i^2\right) - v_i^t G^{-1} v_i$ and,

(i) $Z_i = \sqrt{n_i - 1}\, R_{pooled}^{-1} \left(R_i - R_{pooled} \right)$,

(ii) $R_{pooled} = \frac{1}{N-m} \sum_{i=1}^{m} (n_i - 1) R_i$, $N = \sum_{i=1}^{m} n_i$; the pooled correlation matrix,

(iii) $v_i = \left(z_{11}^{i}, z_{22}^{i}, ..., z_{pp}^{i} \right)^t$ where z_{kk}^{i} is the k-th diagonal element of Z_i ,

(iv) $G = \left(g_{ij} \right)$ is a matrix defined by $g_{ij} = \delta_{ij} + r_{pooled;ij}\, r_{pooled;ij}^{-1}$, δ_{ij} is Kronecker's delta, and $r_{pooled;ij}$ and $r_{pooled;ij}^{-1}$ are the general element of R_{pooled} and R_{pooled}^{-1} .

It is well-known that $J \xrightarrow{d} \chi^2_{(m-1)k}$ with $k =$ $k = \frac{1}{2} p(p-1)$. Thus, to use the test (1), we need sufficiently large n_i ; i runs from 1 to m. In practice, H_0 is rejected if $J > \chi^2_{\alpha;(m-1)k}$; the $(1-\alpha)$-th quantile of $\chi^2_{(m-1)k}$.

As long as our concern is to test H_0, there is nothing wrong with J. However, when H_0 is rejected and we need further information about the samples at which the correlation structure has been shifted, J is useless. It cannot be used to explain the dynamics of correlation structure. Why? Because the distribution of the term J_i in (1) is unknown. On the other hand, knowing how to identify the particular samples where the correlation structure was shifted is important in order to conduct further analyses. This is what motivates this paper.

In his paper, Jennrich, 1970, has remarked that the term J_i needs not asymptotically to be a chi-square variable. However, he does not specifically mention that distribution. In this paper, we investigate the distributional behavior of J_i by means of Mahalanobis distance and introduce a correction factor. We need an equivalent form of J_i a correction factor to investigate the distribution of its corrected version.

3. An Equivalence Theorem

3.1. Basic theorem

We start by recalling the distribution of R_i. Consider the *vec* operator which transforms a matrix * into vector *vec*(*) by stacking each column of * underneath the other. Let K be a commutation matrix,

$$\mathrm{K} = \sum_{i=1}^{p} \sum_{j=1}^{p} \mathrm{H}_{ij} \otimes \mathrm{H}_{ij}^{t} .$$

Here H_{ij} is a matrix of size $(p \times p)$ where all of its elements are 0 but 1 at for (i, j)-th element (Kollo and von Rosen,2005, and Schott, 2007). We borrow this theorem from Kollo and von Rosen, 2005.

Theorem 1. If K_D is a diagonal matrix where its diagonal elements are those of K, and A = $\left(\mathrm{P}_i \otimes \mathrm{I}_p + \mathrm{I}_p \otimes \mathrm{P}_i\right)\mathrm{K}_D$, then,

$$\sqrt{n_i - 1}\left\{vec\left(R_i - \mathrm{P}_i\right)\right\} \xrightarrow{d} \mathcal{N}_{p^2}\left(0, \Gamma\right),$$

where Γ = A_1 – A_2 + A_3 with

(i) $\mathrm{A}_1 = \left(\mathrm{P}_i \otimes \mathrm{P}_i\right)\left(\mathrm{I}_{p^2} + \mathrm{K}\right)$,

(ii) $\mathrm{A}_2 = \mathrm{A}\left(\mathrm{P}_i \otimes \mathrm{P}_i\right) + \left(\mathrm{P}_i \otimes \mathrm{P}_i\right)\mathrm{A}^t$, and

(iii) $\mathrm{A}_3 = \frac{1}{2}\mathrm{A}\left(\mathrm{P}_i \otimes \mathrm{P}_i\right)\mathrm{A}^t$.

This theorem is very important but, unfortunately, it cannot directly be used to investigate the distribution of J_i because Γ is singular. To overcome this problem of singularity, we consider only the upper (lower) diagonal part of R_i as the information contained therein is equal to that in R_i. Suppose we choose the upper part. For this purpose, we use *squareform* operator (MATLAB, 2009) which transforms R_i into a vector representing its upper diagonal elements.

3.2. An equivalent theorem

The *squareform* operator transforms R_i into a vector, $sqf\left(R_{i,u}\right)$ say, representing all its upper diagonal elements. Similarly, $sqf\left(\mathrm{P}_{i,u}\right)$ is the *squareform* of P_i. These vectors are in $\mathcal{R}^k$. The transformation that changes R_i into $sqf\left(R_{i,u}\right)$ can be described formally as follows. Let us define a matrix M = $\left(\mathrm{M}_1|\mathrm{M}_2|.....|\mathrm{M}_p\right)$ of size $\left(k\times p^2\right)$ partitioned into p blocks $\mathrm{M}_r = \left(m_{ij}^r\right)$ of size $(k\times p)$, where M_1 is zero matrix and

$$m_{ij}^r = \begin{cases} 1; & (i,j)=\left(\mathrm{C}_2^r - r + s + 1, s\right) \text{ and } s \text{ runs from } 1 \text{ until } r-1 \\ 0; & \text{elsewhere} \end{cases}$$

for r from 2 until p, and C_2^r denotes the combination of 2 out of r objects.

Then, M transforms $\mathcal{R}^{p^2}$ into $\mathcal{R}^k$ where,

$$sqf\left(R_{i,u}\right) = \mathrm{M}\,vec\left(R_i\right). \tag{2}$$

Two consequences arise from (2). First, the Frobenius norm $\left(R_i - \mathrm{P}_i\right)$ is equivalent to the distance between $sqf\left(R_{i,u}\right)$ and $sqf\left(\mathrm{P}_{i,u}\right)$ in Euclidean space. Therefore, the distribution of the former is equivalent to that of the latter. Second, since the correlation matrix $\Lambda = \mathrm{M}\Gamma\mathrm{M}^t$ of $sqf\left(R_{i,u}\right)$ is non-singular, it is customary to investigate the distribution of the latter in the sense of Mahalanobis distance. This is formulated in Theorem 2 which is an equivalent form of Jennrich's statistic (1).

Theorem 2. Let Γ_0 be the value of Γ under H_0 and $\Lambda_0 = \mathrm{M}\Gamma_0\mathrm{M}^t$. If $\hat{\Lambda}_0$ is a consistent estimator of Λ_0, then

$$\sum_{i=1}^{m}\left(n_i - 1\right)\left\{sqf\left(R_{i,u} - R_{pooled,u}\right)\right\}^t \hat{\Lambda}_0^{-1}\left\{sqf\left(R_{i,u} - R_{pooled,u}\right)\right\} \xrightarrow{d} \chi^2_{(m-1)k}.$$

Proof. We need the following lemma which is a consequence of Theorem 2.2.2 in Kollo and von Rosen, 2005.

Lemma. For all i from 1 until m, we have

$$(n_i - 1)\left\{sqf\left(R_{i,u} - \mathrm{P}_{i,u}\right)\right\}^t \Lambda^{-1}\left\{sqf\left(R_{i,u} - \mathrm{P}_{i,u}\right)\right\} \xrightarrow{d} \chi_k^2 .$$

Corollary.

Under H_0, $(n_i - 1)\left\{sqf\left(R_{i,u} - \mathrm{P}_{0,u}\right)\right\}^t \Lambda_0^{-1}\left\{sqf\left(R_{i,u} - \mathrm{P}_{0,u}\right)\right\} \xrightarrow{d} \chi_k^2$. Thus, since R_1, R_2, …, R_m are independent,

$$\sum_{i=1}^{m}(n_i - 1)\left\{sqf\left(R_{i,u} - \mathrm{P}_{0,u}\right)\right\}^t \Lambda_0^{-1}\left\{sqf\left(R_{i,u} - \mathrm{P}_{0,u}\right)\right\} \xrightarrow{d} \chi_{mk}^2 .$$

From this corollary, since R_{pooled} and $\hat{\Lambda}_0$ are consistent estimators of P_0 and Λ_0, we have Theorem 2.

By construction, see Jennrich, 1970, the i-th term in Theorem 2 is equivalent to J_i in (1). Therefore, the statistic in that theorem is equivalent to J. By using this result, in the next section we introduce a correction factor for each term J_i.

4. A Correction Factor for J_i

Let D_i denote the i-th term of the statistic in Theorem 2. Since it is equivalent to J_i, in what follows we propose a correction factor for D_i. For that purpose, consider testing repeatedly the hypothesis H_{00}: $\mathrm{P}_i = \mathrm{P}_0$ for I from 1 until m. This is equivalent to testing H_0 because the m samples are independent (Montgomery, 2009). Under H_{00}, we have this theorem.

Theorem 3. Let $n_{-i} = \sum_{\substack{j=1 \\ j \neq i}}^{m}\left(n_j - 1\right)$. If $n_i \to \infty$, then for i from 1 until m, we have

$$\frac{N-m}{n_{-i}} D_i \xrightarrow{d} \chi_k^2 .$$

Proof.

It sufficient to investigate the distribution of $sqf\left(R_{i,u} - R_{pooled,u}\right)$. For that purpose we consider $R_i - R_{pooled}$ and we write,

$$R_i - R_{pooled} = \frac{n_{-i}}{N-m} R_i - \frac{1}{N-m} \sum_{\substack{j=1 \\ j \neq i}}^{m} \left(n_j - 1\right) R_j \tag{3}$$

The first term on the right hand side of (3) leads us to search for the distribution of $\frac{n_{-i}}{N-m} sqf\left(R_{i,u}\right)$. In this case, by using Theorem 2.2.2 in Kollo and von Rosen, 2005, we have,

$$\sqrt{n_i - 1}\left\{sqf\left(R_{i,u} - \mathrm{P}_{0,u}\right)\right\} \xrightarrow{d} \mathcal{N}_k\left(0, \Lambda\right). \tag{4}$$

This means that the distribution of $\frac{n_{-i}}{N-m} sqf\left(R_{i,u}\right)$ can be approximated by,

$$\mathcal{N}_k\left(\left(\frac{n_{-i}}{N-m}\right) sqf\left(\mathrm{P}_{0,u}\right), \left(\frac{n_{-i}}{N-m}\right)^2 \frac{\Lambda}{\left(n_i - 1\right)}\right). \tag{5}$$

Similarly, the second term of (3) leads us to conclude that,

$$\mathcal{N}_k\left(\frac{n_{-i}}{N-m} sqf\left(\mathrm{P}_{0,u}\right), \frac{n_{-i}}{\left(N-m\right)^2} \Lambda\right). \tag{6}$$

can be used to approximate the distribution of $\frac{1}{N-m} \sum_{\substack{j=1 \\ j \neq i}}^{m} \left(n_j - 1\right) sqf\left(R_{j,u}\right)$.

Therefore, from (5) and (6), the distribution of $sqf\left(R_{i,u} - R_{pooled,u}\right)$ can be approximated by,

$$\mathcal{N}_k\left(0, \left(\frac{n_{-i}}{\left(N-m\right)}\right) \frac{\Lambda}{\left(n_i - 1\right)}\right).$$

This implies that

$$\sqrt{\frac{\left(N-m\right)}{n_{-i}}} \sqrt{n_i - 1}\left\{sqf\left(R_{i,u} - R_{pooled,u}\right)\right\} \xrightarrow{d} \mathcal{N}_k\left(0, \Lambda\right) \tag{7}$$

for all *i* from 1 until *m*. Thus, we get the theorem.

Corollary. The distribution (4) still remains if $\mathrm{P}_{0,u}$ is replaced by $R_{pooled,u}$ except for a constant multiplier $\sqrt{\frac{(N-m)}{n_{-i}}}$ as showed in (7).

Theorem 3 is what we need to solve our problem. The correlation structure has significantly been shifted at sample *i* if $D_i > \frac{n_{-i}}{N-m}\chi^2_{(1-\alpha);k}$. For practical purpose, since the statistics D_i and J_i have the same value, the latter is computationally more preferable. If D_i needs an inversion matrix of size $(k \times k)$, J_i only needs the inversion of a $(p \times p)$ matrix.

5. Example

To illustrate the advantages of the corrected statistic in Theorems 3, in what follows, an example on its application in teaching and learning process is given. Students' scores in the three subjects Mathematics (MA), Science (SC) and Biology (BI) issued from the final year exam of year 2014 is analyzed to understand the disparity of correlation structure among 13 classes in a public primary school. Data in Table 1 represent the correlation between MA and SC ($\boldsymbol{r}_{12}$), MA and BI ($\boldsymbol{r}_{13}$), and SC and BI ($\boldsymbol{r}_{23}$).

Table 1: Correlations among MA, SC and BI

Class	n	r_{12}	r_{13}	r_{23}
1	37	0.905	0.915	0.930
2	38	0.734	0.846	0.809
3	31	0.696	0.797	0.781
4	33	0.733	0.772	0.862
5	40	0.885	0.847	0.853
6	37	0.832	0.862	0.806
7	26	0.758	0.690	0.810
8	27	0.743	0.703	0.863
9	28	0.805	0.903	0.765
10	37	0.743	0.624	0.621
11	34	0.661	0.681	0.738
12	33	0.271	0.691	0.264
13	35	0.643	0.596	0.751
R_{pooled}		0.727	0.766	0.758

To test whether the disparity of correlation structure among classes occurs, Jennrich's statistic in (1) is enough and appropriate to be used. However, to identify the classes in which the correlation structure differs significantly, we need the statistic that we introduce in Theorem 3. Its implementation needs the calculation of the statistic J_i, correction factor (CF) and corrected Jennrich's statistic ($CorJ_i$). The results are in Table 2.

Table 2: Statistic J_i, correction factor and corrected statistic

Class	J_i	n_{-i}	CF	$CorJ_i$
1	7,979	387	1,093	8,721
2	2,327	386	1,096	2,550
3	1,032	393	1,076	1,111
4	3,475	391	1,082	3,759
5	4,809	384	1,102	5,298
6	2,521	387	1,093	2,756
7	3,125	398	1,063	3,321
8	4,940	397	1,065	5,264
9	3,994	396	1,068	4,266
10	12,241	387	1,093	13,380
11	1,703	390	1,085	1,847
12	72,567	391	1,082	78,505
13	7,756	389	1,087	8,433

For 5% significance level, the critical point is $\chi^2_{0.95;3}$ = 7.815. Therefore, based on Theorem 3, we do not only test the occurrence of correlation structure dynamic but also at the same time identify the classes where the structure differs significantly. The results are visually presented in Figue 1. The horizontal axis refers to the number of the class under study and vertical axis represents the value of $CorJ_i$. The dashed line is the value of the critical point for $\chi^2_{0.95;3}$ = 7.815. We learn from the figure that the shift occurs significantly in Class 1, Class10, Class 12 and Class 13. Severe changed is in Class 12.

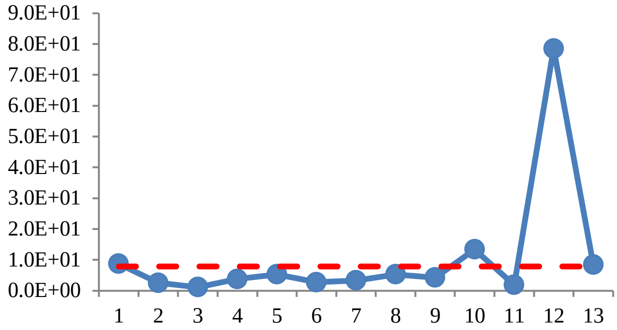

Figure 1: Dynamic of correlation structure for 5% significance level

In order for the school management to eliminate the above disparity, this figure suggests to study further the root causes why the correlation structure is shifted in those four classes. From statistics point of view, to solve this problem we need a special statistic. That statistic will be developed in future research.

6. Concluding Remarks

Two theorems are introduced in this paper. The first, Theorems 2, presents a statistic equivalent to Jennrich's. It is a weighted sum of squares of Mahalanobis distances where each summand D_i is equivalent to J_i. The second is Theorems 3. It shows that D_i, corrected by factor $\frac{N-m}{n_{-i}}$, converges in distribution to χ_k^2. It is this corrected statistic that allows us to study the dynamics of correlation structure. Since the computational complexity of D_i is of order $O\left(p^4\right)$ while J_i is $O\left(p^2\right)$, the latter is preferable in practice.

Acknowledgments

The author gratefully acknowledges the anonymous referees for their comments. Special thanks go to the Institute for Mathematical Research (INSPEM), Universiti Putra Malaysia, for providing research facilities during his appointment as Research Fellow.

References

Aitkin, M. A., Nelson, W. C., and Reinfurt, K. H. (1968). Test for correlation matrices. *Biometrika*, **55**: 327–334.

Annaert, J., Claes, A.G.P., and de Ceuster, M.J.K. (2006). Inter-temporal stability of the European credit spread co-movement structure. *European Journal of Finance*, **12** : 23–32.

Deblauwe, F., and Le, H. (2007). *Stability of correlation between credit and market risk over different holding periods*. University of Antwerp Management School, Antwerp, Belgium. www.linkedin.com/in/ frdeblauwe (accessed on 23 July 2013).

Fischer, M. (2007*). Are correlations constant over time? Application of the CC-Trig–Test to Return Series from Different Asset Classes*; SFB 649 Discussion Paper 2007-012; University of Erlangen-Nűrmberg: Nűrmberg, Germany.

Goetzmann, W.N., Li, L., and Rouwenhorst, K.G.J. (2005). Long-term global market correlations. *Journal of Business*, **78**: 1–38.

Hotelling, H. (1940). The selection of variates for use in prediction, with some comments on the general problem of nuisance parameters. *Annales of Mathematical Statistics*, **11**: 271–283.

Jennrich, R.I. (1970). An asymptotic chi-square test for the equality of two correlation matrices. *Journal of American Statistical Association*, **65**: 904–912.

Kollo, T., and von Rosen, D. (2005). *Advanced multivariate statistics with matrices*. Netherlands: Springer, Dordrecht.

Larntz K., and Perlman, M.D. (1985). *A simple test for the equality of correlation matrices.* Technical Report No. 63, Department of Statistics, University of Washington, Seattle, USA.

Lawley, D.N. (1963). On testing a set of correlation coefficients for equality. *Annales of Mathematical Statistics*, **34**: 149–151.

MATLAB (2009). MATLAB version 7.8.0, 2009a.

Montgomery, D.C. (2009). *Introduction to statistical quality control,* 6th Ed.; New York, USA: John Wiley & Sons, Inc.

Onnela, J.P., Chakraborti, A., Kaski, K., Kert'esz, J., and Kanto, A. (2003). Dynamic of market correlations: Taxonomy and portfolio analysis. *Physical Review E*, **68**: 056110.

Ragea, V. (2003). Testing correlation stability during hectic financial markets. *Financial Market Portfolio Management*, **17**: 289–380.

Rosenow, B., Gopikrishnan, P., Plerou, V., and Stanley, H.E. (2003). Dynamics of cross-correlations in the stock market. *Physica A*, **324**: 241–246.

Schott, J.R. (2007). Testing the equality of correlation matrices when sample correlation matrices are dependent. *Journal of Statistical Planning and Inference*, **137**: 1992–1997.