1-1-2013

A LASSO Chart for Monitoring the Covariance Matrix

Edward M. Maboudou-Tchao
University of Central Florida

Norou Diawara

Recommended Citation

Maboudou-Tchao, Edward M. and Diawara, Norou, "A LASSO Chart for Monitoring the Covariance Matrix" (2013). *Faculty Bibliography 2010s*. 4357.
https://stars.library.ucf.edu/facultybib2010/4357

A LASSO Chart for Monitoring the Covariance Matrix

Edgard M. Maboudou-Tchao & Norou Diawara

QTQM

A LASSO Chart for Monitoring the Covariance Matrix

Edgard M. Maboudou-Tchao[1] and Norou Diawara[2]

[1]Department of Statistics, University of Central Florida, Orlando, FL, USA
[2]Mathematics & Statistics Department, Old Dominion University, Norfolk, VA, USA

Abstract: Multivariate control charts are essential tools in multivariate statistical process control. In real applications, when a multivariate process shifts, it occurs in either location or scale. Several methods have been proposed recently to monitor the covariance matrix. Most of these methods use rational subgroups and are used to detect large shifts. In this paper, we propose a new accumulative method, based on penalized likelihood estimators, that uses individual observations and is useful to detect small and persistent shifts in a process when sparsity is present.

Keywords: Average run length (ARL), covariance matrix, multi standardization, penalized likelihood function.

1. Introduction

Multivariate process control is becoming an important tool in process control instead of using a collection of univariate charting techniques applied to individual components. "Shewhart-type" charts use rational subgroups as opposed to "accumulative" methods that pool information across successive observations. The best known work in multivariate control charts to monitor the mean vector is that of Hotelling [14] which is a direct multivariate extension of the univariate Shewhart $\bar{X}$ chart.

Several methods have been proposed to monitor the covariance matrix Σ. Two current standard Shewhart-type methodologies are available. One, due to Montgomery and Wadsworth [23], uses the generalized variance; the other, due to Alt [1] uses the generalized likelihood ratio. Sakata [31], Calvin [6], Levinson *et al.* [18], and Vargas and Lagos [35] also proposed Shewhart-type charts based on the likelihood ratio test (LRT). Yen and Shiau [39] proposed effective control charts for monitoring the covariance matrix based on a one-sided and two-sided likelihood ratio test based control charts.

These Shewhart-type control charts do not take into account the history of the process. Some authors suggested to use 'accumulative' methods that pool information across successive observations-methods such as cumulative sum or exponentially weighted moving average charts. Shewhart-type charts are effective for detecting large shifts, while the accumulative methods are needed for smaller but persistent shifts. For that purpose, some authors, like Chan and Zhang [7] and Runger and Testik [30], proposed multivariate cumulative sum (MCUSUM) control charts.

Recent proposals suggested to accumulate historical observations by taking some exponentially weighted moving average (EWMA)-type of some statistics based on individual observations. For that purpose, Yeh *et al.* [37] proposed the maximum multivariate exponential weighted moving variability (MaxMEWMV), Huwang *et al.* [15] suggested the

multivariate exponential weighted mean squared deviation (MEWMS) and the multivariate exponential weighted variance. Hawkins and Maboudou-Tchao [13] defined an analogous chart, the Multivariate Exponentially Weighted Moving Covariance Matrix (MEWMC). Memar and Niaki [22] suggested to monitor the covariance matrix using EWMA control charts based on squared deviation of observations from target. Instead of using the trace, the authors employed the ℓ_1-norm and ℓ_2-norm-based distances between diagonal elements of the estimators and their expected values to design new control charts for monitoring the covariance matrix of a multivariate process. Memar and Niaki [22] evaluated their proposed statistic using simulations. Zang and Chang [42] proposed the Multivariate Exponentially Weighted Moving deviation (MEWMD) which is an extension of the Multivariate Exponentially Weighted Moving average (MEWMA). The MEWMD chart is designed for detecting changes only on the variance components. Zhang *et al.* [43] proposed a chart to monitor simultaneously the mean vector and the covariance matrix of a process. Reynolds and Cho [27] proposed a chart for monitoring the mean and covariance matrix of a multivariate process when the joint distribution of the process variables follows a multivariate normal distribution. The emphasis of Reynolds and Cho's [27] study was on the use of combinations of multivariate exponentially weighted moving average (MEWMA) control charts based on sample means and the sum of the squared deviations from the target. Reynolds and Cho [27] concluded that the best overall performance was achieved using a combination of MEWMA charts based on the sample means and the sum of squared regression adjusted deviations from the target. Reynolds and Cho [28] extended their previous work by using variable sampling intervals.

For a bivariate process, Machado and Costa [20] suggested a double sampling and EWMA charts based on the sample variances. The monitoring statistic was based on the standardized sample variance of p quality characteristics called the maximum of the variances (VMAX) statistics. The points plotted on the chart corresponded to the maximum value of these p variances. The VMAX statistic was faster than the generalized variance at detecting shifts. Other charts have also been proposed based on this method using a synthetic statistic (Machado *et al.* [21]), a double-sampling scheme, and an EWMA-type control chart (Machado and Costa [20]). Aparisi *et al.* [2] present work shows the design of generalized variance chart $|\mathbf{S}|$ with adaptive sample size to control processes defined by two quality characteristics, comparing the power of this new chart versus the $|\mathbf{S}|$ chart with the sample size. Their approach is also intended for monitoring changes only on the variance components of a bivariate process. Still with the bivariate process, Quinino *et al.* [25] proposed a new chart for controlling the covariance matrix of bivariate processes using the mixed variance (VMIX) statistic.

Some authors proposed other schemes different from the above. Tang and Barnett [32-33] proposed to use Shewhart procedures based on the decomposition of the covariance matrix while Hao *et al.* [12] suggested the multivariate projection chart. Zamba and Hawkins [41] used a multivariate change-point model for detecting change in the mean and/or covariance matrix. For more details on monitoring the covariance matrix, the readers are referred to the papers by Bersimis *et al.* [5] and Yeh *et al.* [38].

Recently, Wang and Jiang [36] and Zou and Qiu [44] constructed control charts for monitoring multivariate mean vector using the Least Absolute Shrinkage and Selection Operator (LASSO) type penalty in the context of statistical process control. Zou and Qiu [44] used the adaptive LASSO to construct a control chart to monitor the mean vector. Zou *et al.* [45] proposed a LASSO chart to monitor a multivariate linear profile. Li *et al.* [17] suggested to monitor the covariance matrix with a penalized likelihood estimation based on the

LASSO.

This paper is motivated by the fact that in multivariate or high dimensional applications, when a change is detected, it is often rare that all quality characteristics changed. This means that the probability that all variables change simultaneously is very small. Also, not all quality characteristics are correlated with one another; so, it can happen that one subset of variables is uncorrelated with another subset of variables. This brings to bear the idea of 'sparsity' property. Our goal is to design a good process monitoring tool that uses this property. Note that the accumulative methods using the EWMA-type estimates of the covariance matrix do not take into consideration this sparsity property of the process. By not including that sparsity property in the design of the chart, the chart will contain information not relevant to the true process variability and this may harm the performance of the chart. Therefore, for these applications or high dimensional processes, the sparsity property should be considered in the design of the control chart and is the main objective of this present paper. So, we propose a new control chart to monitor the process variability using information pertinent to the process such as sparsity.

1.1. Monitoring Shift in the Covariance Matrix

There are two current standard Shewhart-type methodologies for monitoring the covariance matrix $\boldsymbol{\Sigma}$. One uses the generalized variance, Montgomery [24], and the other uses the generalized likelihood ratio, Alt [1]. Suppose that process readings are p component vectors $\mathbf{x}_i$, $i = 1, 2, \ldots$, and that while the process is in control, these process readings are independent multivariate normal random vectors with mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}_0$. We assume that the interest focuses on control of the covariance matrix to see whether this has departed from an in-control (IC) value $\boldsymbol{\Sigma}_0$. Also, we assume that the in-control mean vector $\boldsymbol{\mu}$ and the covariance matrix $\boldsymbol{\Sigma}_0$ are known exactly. The standard methods of monitoring the covariance matrix are Shewhart-Type charts based on rational subgroups of size m greater than p. If $\mathbf{S}_i$ is the sample covariance matrix of the ith rational subgroup, Montgomery and Wadsworth [23] proposed to chart the determinant of $\mathbf{S}_i$. Reynolds and Cho [27] suggested the charting of the trace of $\mathbf{S}_i$. A different approach was suggested by Alt [1]. His approach is based on the generalized likelihood ratio (GLR) statistic for testing $H_0 : \boldsymbol{\Sigma} = \boldsymbol{\Sigma}_0$ vs. $H_1 : \boldsymbol{\Sigma} \neq \boldsymbol{\Sigma}_0$. So, Alt's charting statistic is

$$W_i = -(m-1)\left[p + ln\frac{|\mathbf{S}_i|}{|\boldsymbol{\Sigma}_0|} - tr(\boldsymbol{\Sigma}_0^{-1}\mathbf{S}_i) \right], \tag{1}$$

where $\mathbf{S}_i$ is the $p \times p$ sample covariance matrix for sample i, tr is the trace operator, the subscript i represents the rational subgroup number, m is the rational subgroup size, and p is the dimension.

As usual for Shewhart charts, these are effective for large transient changes in the covariance matrix, but are less suitable for smaller persistent shifts. Recent proposals by Huwang $et\ al.$ [15] and Hawkins and Maboudou-Tchao [13] defined an analogous chart for monitoring the covariance matrix. We will focus on the proposal of Hawkins and Maboudou-Tchao, the Multivariate Exponentially Weighted Moving Covariance Matrices (MEWMC), as it is relevant for the remainder of the paper. Assume the p-component process reading vectors $\mathbf{x}_n \sim N(\boldsymbol{\mu},\boldsymbol{\Sigma})$, $n = 1, 2, \ldots$. In the MEWMC, the process readings $\mathbf{x}_n$ are first multi-standardized to $\mathbf{u}_n = \boldsymbol{\Sigma}^{-1/2}(\mathbf{x}_n - \boldsymbol{\mu})$. While in control, the $\mathbf{u}_n$ are $N(\mathbf{0}, \mathbf{I}_p)$. The standardized MEWMC is defined by the recursion

$$\mathbf{W}_n = (1-\lambda)\mathbf{W}_{n-1} + \lambda\mathbf{u}_n\mathbf{u}_n' \ , \tag{2}$$

where $\mathbf{W}_0 = \mathbf{I}_p$, and λ is the smoothing constant. In most EWMA methods, λ can take the value 1, but it should be clear that for this definition of $\mathbf{W}_n$, it is prohibited for λ to take the value 1, as when $\lambda = 1$, $\mathbf{W}_n = \mathbf{u}_n \mathbf{u}'_n$ and is not positive definite.

The MEWMC chart statistic is obtained by

$$C_n = tr(\mathbf{W_n}) - log \mid \mathbf{W_n} \mid -p, \tag{3}$$

where tr is the trace operator and $\mid . \mid$ represents the determinant.

The control chart used then consists of plotting C_n against n, and signaling a loss of control if $C_n > h$ where h is chosen to achieve a specified in-control ARL.

2. Penalized Likelihood Estimation of Inverse Covariance Matrix

In the general case, assume that we observe a set of n vectors $\mathbf{x}_i$, $i = 1, 2, \ldots, n$, where each vector, $\mathbf{x}_i$ is p dimensional. Without loss of generality, assume that $\mathbf{x}_i$ has zero mean. If the vectors $\mathbf{x}_i$ are identically distributed, then the sample covariance matrix is given by

$$\mathbf{S} = \frac{1}{n} \sum_{i=1}^{n} \mathbf{x}_i \mathbf{x}'_i$$

and $\mathbf{S}$ is an unbiased estimator of the true covariance matrix $\Sigma = E(\mathbf{S})$. While $\mathbf{S}$ is an unbiased estimate of the true covariance matrix, it is also singular when $n < p$. In practice, n may be much smaller than p and so most of the eigenvalues of Σ are incorrectly estimated as zero. Several methods have been proposed to regularize the estimate of Σ so that it is not singular. Shrinkage estimators are a class of estimators which regularize the covariance matrix by shrinking it toward some target structure. The concept of robust estimation of an inverse covariance matrix was first introduced by Dempster [9] who suggested that the number of parameters to be estimated be reduced by setting some elements of the precision matrix or inverse covariance matrix Ω to zero.

Recently, a number of methods have been proposed for regularizing the estimate by making either the covariance or its inverse sparse. In the absence of model assumptions when $p > n$, an active line of statistical research is based on imposing various restrictions on the model – for instance, sparsity.

Given a random sample $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n$ following a multivariate normal distribution $N(\mu, \Sigma)$, the log likelihood for μ and $\Omega = \Sigma^{-1}$ can be expressed as

$$\frac{n}{2} log \mid \Omega \mid -\frac{1}{2} \sum_{i=1}^{n} (\mathbf{x}_i - \mu)' \Omega (\mathbf{x}_i - \mu) \tag{4}$$

up to a constant not depending on the parameters. The matrix Ω is sometimes referred to as the concentration matrix, precision matrix, or dispersion matrix. The maximum likelihood estimator of (μ, Σ) is $(\bar{\mathbf{x}}, \hat{\Sigma})$, where

$$\hat{\Sigma} = \frac{1}{n} \sum_{i=1}^{n} (\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{x}_i - \bar{\mathbf{x}})'$$

is the maximum likelihood estimate of Σ with $\bar{\mathbf{x}}$ the sample mean. Assuming that the observations are properly centered, it follows that the sample mean is zero. To obtain the maximum likelihood estimator of the covariance matrix is equivalent to minimize

$$-\frac{2}{n} l(\Omega) = -log \mid \Omega \mid + tr(\Omega \hat{\Sigma}). \tag{5}$$

When $p > n$, the maximum likelihood estimate is no longer useful; in particular, it is not positive definite and has rank no greater than n. Finding the most sparse inverse covariance matrix to fit a data set is a NP-hard problem (Banerjee *et al*. [3]). To overcome this problem, several ℓ_1-regularization methods have been proposed for sparse estimation of the inverse covariance matrix.

To achieve sparse structure, penalized likelihood methods using the ℓ_1 penalty have been considered by Yuan and Lin [40], Banerjee *et al*. [4], D'Aspremont *et al*. [8], Friedman *et al*. [11], and Rothman *et al*. [29], who have all proposed different algorithms for computing this estimator. The resulting estimator $\hat{\boldsymbol{\Omega}}$ should minimize the following objective function

$$Q(\boldsymbol{\Omega}) = -log\,|\boldsymbol{\Omega}| + tr(\boldsymbol{\Omega}\hat{\boldsymbol{\Sigma}}) + \rho\,\|\boldsymbol{\Omega}\|_1\,,$$

where $\|\boldsymbol{\Omega}\|_1 = \sum_{j=1}^{p}\sum_{i=1}^{p}|\omega_{ij}|$ is the ℓ_1 norm of $\boldsymbol{\Omega}$ and ρ is a data dependent tuning parameter which can be tuned to achieve different levels of sparsity of the $\boldsymbol{\Omega}$ estimate. The term $\|\boldsymbol{\Omega}\|_1$ encourages sparseness of the precision matrix or conditional independence among variables. This approach produces a sparse estimate of the inverse covariance matrix and an estimate of the covariance matrix that is positive definite. Yuan and Lin [40] solved the corresponding optimization problem using the MAXDET algorithm, while Vandenberghe *et al*. [34] focused on statistical properties of the estimates. D'Aspremont *et al*. [8] proposed two efficient first-order numerical algorithms with a low memory requirement using semidefinite programming algorithms, which obey the positive-definiteness constraint of the precision matrix. Rothman *et al*. [29] and Lam and Fan [16] showed that the Frobenius norm between the inverse correlation matrix and its ℓ_1 penalized likelihood estimator is $O_p(\sqrt{S\log p/n})$, where S is the number of the nonzero elements of the inverse correlation matrix. Consequently, the sparse inverse correlation matrix is highly estimable and the dimensionality only costs an order of $\log p$, a remarkable improvement on the general result of Fan and Peng [10]. The ℓ_1 penalty is convex and leads to a desirable convex optimization problem when the log-likelihood function is convex. Using a coordinate descent procedure, Friedman *et al*. [11] proposed the graphical Least Absolute Shrinkage and Selection Operator (GLASSO) algorithm to estimate the sparse inverse covariance matrix using the LASSO penalty.

This procedure has been referred to as graphical lasso, GLASSO, (Friedman *et al*. [11]) or Sparse Permutation Invariant Covariance Estimator, SPICE, (Rothman *et al*. [29]). The resulting estimation procedure has excellent theoretical properties, as shown by Rothman *et al*. [29] and Ravikumar *et al*. [26].

In this paper, LASSO type penalty function is applied to multivariate control charts. Recently, Wang and Jiang [36] and Zou and Qiu [44] constructed control charts for monitoring multivariate mean vector using LASSO type penalty in the context of statistical process control (SPC). This work extends the research on recent penalized-likelihood-based SPC method (Wang and Jiang [36], Zou and Qiu [44]) to the monitoring of covariance matrix.

3. A LASSO Chart for Monitoring a Shift in the Covariance Matrix

Suppose that the process readings are p component vectors $\mathbf{x}_i, i = 1,2,\ldots,$ and that while the process is in control, these process readings are independent multivariate normal random vectors with mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$. Moreover, we assume that the in-control mean vector and the covariance matrix $\boldsymbol{\Sigma}$ are known exactly.

Our proposal is an accumulative method in line with the accepted wisdom that

accumulative methods work better with individual observations than with rational groups. Also, it is convenient to work with multistandardized data vectors instead of the original process readings $\mathbf{x}$. To achieve this, we use a matrix $\mathbf{A}$ with the property $\mathbf{A\Sigma A'} = \mathbf{I}_p$, where $\mathbf{I}_p$ is a $p \times p$ identity matrix, and transform $\mathbf{x}$ to $\mathbf{u} = \mathbf{A}(\mathbf{x} - \mathbf{\mu})$. While the process is in control, the $\mathbf{u}$'s are $N(\mathbf{0}, \mathbf{I}_p)$. Our proposal is a two-step mechanism. In the first step, for each observation n, we compute $\mathbf{U}_n = \mathbf{u}_n \mathbf{u}'_n$ which is a $p \times p$ symmetric matrix but rank deficient and not positive definite. Using $\mathbf{U}_n$, we find an estimate of the inverse covariance matrix by

$$\hat{\mathbf{\Omega}}_n = \underset{\mathbf{\Omega} > 0}{arg\,max} \left\{ tr(\mathbf{\Omega}_n \mathbf{U}_n) - \log|\mathbf{\Omega}_n| + \rho\|\mathbf{\Omega}_n\| \right\}. \tag{6}$$

Now $\rho > 0$ performs regularization and the estimated covariance is invertible and positive definite. We obtain an estimate of the covariance matrix $\mathbf{V}_n$ by inverting $\hat{\mathbf{\Omega}}_n$. We then define the LASSO multivariate exponentially weighted moving covariance matrix (LEWMC) by the recursion $\mathbf{S}_0 = \mathbf{I}_p$, and for $n = 1, 2, \ldots$

$$\mathbf{S}_n = (1 - \lambda)\mathbf{S}_{n-1} + \lambda\mathbf{V}_n, \tag{7}$$

where λ is the smoothing constant. Also, $\mathbf{S}_n$ is positive definite for all values of n, the proof of this statement is available in the Appendix. Unlike the MEWMC, λ can take the value 1 for the LEWMC. When $\lambda = 1$, $\mathbf{S}_n = \mathbf{V}_n$ and is positive definite.

In the last step, we compare the matrix $\mathbf{S}_n$ with the identity matrix using the chart statistic

$$c_n = tr(\mathbf{S}_n) - \log|\mathbf{S}_n| - p. \tag{8}$$

The control chart involves plotting c_n against n and signaling a loss of control if $c_n > h$, where the control limit h is chosen to achieve a specified in-control average run length (IC ARL). Since $\mathbf{S}_n$ is positive definite for all values of n, the chart statistic c_n is defined for all values of n.

When $\rho = 0$, no regularization is done and the LEWMC is the same as the MEWMC.

3.1. Algorithm Summary

Our method for monitoring the covariance matrix of a process using a penalized likelihood estimation consists of the following steps.

1. Multistandardize the data by transforming the original variables $\mathbf{x}$ into $\mathbf{u}$.

2. For each observation n, compute $\mathbf{U}_n = \mathbf{u}_n \mathbf{u}'_n$.

3. For each observation n, transform $\mathbf{U}_n$, a non-positive definite matrix into a positive definite matrix, i.e. find $\hat{\mathbf{\Omega}}_n$ and then $\mathbf{V}_n$. Use the graphical LASSO algorithm of Friedman *et al.* [11] for this step to directly obtain $\mathbf{V}_n$.

4. Using $\mathbf{S}_0 = \mathbf{I}_p$, compute $\mathbf{S}_n = (1 - \lambda)\mathbf{S}_{n-1} + \lambda\mathbf{V}_n$ for $n = 1, 2, \ldots$

5. Compute the chart statistic $c_n = tr(\mathbf{S}_n) - \log|\mathbf{S}_n| - p$ for $n = 1, 2, \ldots$ and detect an out of control behavior when $c_n > h$.

This will be referred to as "LEWMC chart" in the remainder of the paper.

3.2. Selection of the Control Limit h

The chart involves three constants, λ, ρ, and h. The first constant, also called the smoothing constant λ, is used for tuning the chart to different sizes of change. A small value of λ is used to detect a small shift and a large value of λ is used if our goal is to detect a large

shift. The second element of interest is the sparsity tuning parameter ρ. A small value of ρ will make the estimated covariance matrix less sparse whereas a large value will make the estimate sparse. We will focus more on this parameter in the next section. The control limit h defines the run lengths. Different combinations of λ, ρ, and p lead to different values of h. Getting the appropriate value of h to achieve a specific IC ARL is not an easy task because the statistic used by the LEWMC chart does not have a known standard distribution. So, the values of h are obtained by simulation.

3.3. Out-of-Control Settings

The element studied here is the out-of-control settings. Three shifts were investigated. In the first scenario ("variance shift"), we changed the covariance matrix from its in-control value of $\mathbf{I}_p$ to a matrix having $1 + \delta$ in the (1, 1) position with the other elements unchanged.

In the second scenario ("correlation shift"), the covariance matrix was left as $\mathbf{I}_p$, except for putting a correlation δ in the (1, 2) and (2, 1) positions with the other elements unchanged.

The third scenario ("simultaneous variance and correlation shifts") changed the identity covariance matrix by putting δ in the (1, 2) and (2, 1), and also $1 + \delta$ in the (1, 1) and (2, 2) positions while the other elements remained unchanged.

4. Performance Study

The performance of quality control charts is commonly measured by their average run length (ARL) following an out of control shift. As a benchmark, we will compare the out-of-control ARL of our proposed method with that of the Multivariate Exponentially Weighted Moving Covariance matrix (MEWMC) of Hawkins and Maboudou-Tchao [13]. Also, as pointed out by a referee, we will compare, in the bivariate case, the LEWMC with the VMAX chart proposed by Machado and Costa [20].

The performance comparison requires adjusting the approaches to have comparable in-control behavior, and then selecting informative out-of-control settings to evaluate. We compare the two methods for p = 5, 10, 20, and λ = 0.1. The in-control ARL is set to 200.

4.1. Performance under Variance Shifts

A full study of the chart performance is infeasible, as it would involve arbitrary changes in the $p(p+1)/2$ elements of the covariance matrix. Hence, we sketch just a scenario which changes the covariance matrix from the identity matrix to a matrix with $1 + \delta$ in the (1, 1) position while the other elements remain unchanged, and that leaves the mean vector at $\mathbf{0}$. We investigate the performance of δ varying from 0 to 1, simulating a minimum of 20,000 independent series at each δ value. The resulting out-of-control ARL values are shown in Table 1. The table displays the comparison between the MEWMC and LEWMC with ρ = 0.1, 0.2, 0.5, 1, 1.5, 1.8, and 2. In terms of performance, the LEWMC bettered the MEWMC for variance increases for all dimensions considered. This result should not be surprising as in this case, only the variances of the process variables change (variation of the diagonal elements). Then, the estimated covariance matrix has most of the off-diagonal elements equal to zero, and the "changed" diagonal elements will be correctly estimated. So, the shift in variance will be detected very quickly by the LEWMC.

4.2. Performance under Correlation Shifts

We again investigate p = 5, 10, and 20, and ρ = 0.1, 0.2, 0.5, 1, 1.5, 1.8, and 2. The correlation inserted in the (1, 2) element of the covariance matrix is varied from 0 to 0.9.

From the table (Table 2), the LEWMC is outperformed by the MEWMC. However, the performance of the LEWMC is getting closer to the MEWMC as ρ gets small. In fact, the result seen here should be expected as in this case, only the off-diagonal elements change after a shift. Here, the LEWMC chart will hardly detect changes (unless they are quite large, especially with respect to the penalty weight), given that by itself it will force the off-diagonal estimated elements towards zero.

Table 1. Performance comparison of the charts for variance shift.

p	δ	MEWMC	LEWMC(ρ)						
			0.1	0.2	0.5	1	1.5	1.8	2
5	0.00	199.7	199.9	199.4	201.4	199.8	200.4	199.0	199.7
5	0.10	173.7	163.2	153.5	157.2	151.4	157.9	148.3	151.0
5	0.20	145.2	129.7	120.7	119.3	115.5	122.2	115.5	116.8
5	0.30	117.8	101.0	93.10	92.95	90.82	94.46	90.89	93.33
5	0.40	96.06	79.33	73.62	74.47	73.62	75.83	72.88	73.02
5	0.50	77.96	63.68	58.93	59.53	58.73	62.06	58.92	60.85
5	0.60	64.22	51.57	48.09	48.94	48.52	51.38	49.53	50.27
5	0.70	53.21	43.23	40.33	40.57	41.49	43.57	41.87	42.94
5	0.80	44.28	37.05	34.35	35.16	35.17	37.07	36.39	37.28
5	0.90	38.54	31.91	30.07	30.43	31.19	32.15	31.57	32.28
5	1.00	33.45	27.59	26.11	26.88	27.31	28.59	28.12	28.19
10	0.00	201.3	199.5	198.2	199.7	200.4	199.5	199.4	199.3
10	0.10	190.1	175.9	171.5	165.3	166.0	164.0	162.8	161.0
10	0.20	174.4	149.1	141.0	139.3	137.6	136.5	135.7	133.8
10	0.30	151.7	125.7	119.2	114.2	114.1	112.7	112.6	112.3
10	0.40	133.9	104.5	96.72	95.27	95.85	94.49	95.89	94.10
10	0.50	114.2	88.11	81.55	79.74	80.28	79.93	79.32	81.12
10	0.60	99.02	74.12	69.59	67.41	66.44	68.42	69.19	68.84
10	0.70	83.67	63.60	58.90	57.86	58.56	58.85	59.24	59.03
10	0.80	72.70	53.48	50.29	49.42	51.02	50.45	50.91	51.78
10	0.90	62.94	46.97	43.27	42.99	44.44	44.76	45.57	45.54
10	1.00	55.32	40.78	38.16	37.62	39.67	39.90	40.76	41.13
20	0.00	200.3	199.7	200.6	201.1	199.2	199.5	201.9	200.5
20	0.10	196.3	180.9	179.2	174.7	172.9	173.6	178.5	179.5
20	0.20	186.6	163.4	161.8	155.2	151.8	121.6	157.8	156.4
20	0.30	172.4	145.1	141.1	136.2	132.5	132.2	137.7	134.6
20	0.40	161.1	127.6	123.0	117.9	115.4	115.3	119.5	118.6
20	0.50	146.1	113.2	106.7	102.7	100.5	102.8	105.7	105.5
20	0.60	134.3	99.59	95.05	90.47	89.11	89.51	91.56	92.66
20	0.70	121.0	86.72	82.08	78.96	78.78	79.10	82.14	82.44
20	0.80	109.0	78.58	72.04	70.57	68.82	72.31	73.59	73.43
20	0.90	97.31	68.45	64.29	62.66	61.45	63.37	65.18	65.64
20	1.00	88.24	60.90	57.27	56.01	55.90	57.77	59.22	58.72

4.3. Performance under Variance and Correlation Shifts

Next, we considered a shift in the variance and correlation. The constant δ was varied from 0 to 0.9. We used p = 5, 10, and 20, and set the penalty term ρ = 0.1, 0.2, 0.5, 1, 1.5, 1.8, and 2. The results are shown in Table 3. It can also be seen from the table that the LEWMC control chart had a much faster reaction to shifts than the MEWMC.

Table 2. Performance comparison of the charts for correlation shift.

p	δ	MEWMC	LEWMC(ρ)						
			0.1	0.2	0.5	1	1.5	1.8	2
5	0.00	196.2	199.9	199.4	201.4	198.1	201.3	200.2	199.8
5	0.10	180.8	189.3	186.6	198.4	195.5	197.3	194.6	196.4
5	0.20	143.6	159.3	165.3	182.9	188.8	194.4	189.9	190.7
5	0.30	104.4	122.1	137.4	161.4	177.6	181.9	181.5	184.0
5	0.40	72.78	93.74	107.9	144.4	164.2	171.8	171.5	172.6
5	0.50	50.40	68.32	85.02	123.4	147.4	160.2	158.4	161.0
5	0.60	35.94	50.85	66.48	102.4	132.3	144.9	145.4	146.8
5	0.70	26.09	37.90	52.28	88.17	120.6	134.2	133.0	136.6
5	0.80	19.63	29.28	41.41	76.35	108.6	121.7	123.1	123.3
5	0.90	15.60	22.60	32.95	64.86	97.22	110.7	112.4	114.8
10	0.00	199.0	199.9	198.5	201.7	201.4	200.1	199.6	199.6
10	0.10	189.1	192.9	195.1	198.7	198.1	196.9	193.9	193.6
10	0.20	170.2	175.8	181.4	190.7	193.9	197.8	192.2	193.4
10	0.30	139.1	150.6	161.0	178.0	189.5	189.8	187.8	189.7
10	0.40	108.7	124.5	138.3	165.5	180.1	181.5	183.3	180.5
10	0.50	81.30	101.5	117.4	149.4	173.4	174.7	175.5	178.2
10	0.60	60.03	78.83	97.12	134.1	158.9	166.3	166.0	167.0
10	0.70	43.61	62.45	80.28	119.8	148.5	154.5	158.6	159.1
10	0.80	32.02	48.34	65.81	105.0	137.4	144.3	148.7	152.4
10	0.90	24.10	37.33	53.28	93.21	127.4	137.4	140.4	140.7
20	0.00	200.8	199.5	200.8	201.5	199.5	201.3	200.9	201.1
20	0.10	198.4	195.0	197.1	201.2	196.7	200.6	200.3	200.1
20	0.20	186.5	186.9	190.1	196.0	195.9	200.0	199.1	200.4
20	0.30	166.8	174.1	181.7	189.0	192.8	197.9	196.9	198.6
20	0.40	145.1	158.5	167.9	184.6	186.6	190.6	193.5	195.0
20	0.50	119.8	136.8	153.9	174.8	181.1	188.6	190.8	192.6
20	0.60	96.87	120.1	139.5	163.3	174.2	183.7	184.7	186.5
20	0.70	74.77	104.2	122.7	153.1	166.5	177.0	179.2	179.9
20	0.80	54.88	85.58	106.1	140.3	156.8	169.3	172.9	173.8
20	0.90	40.74	70.97	92.67	130.0	152.4	163.3	164.6	169.6

4.4. Selection of the Tuning Parameter ρ

The chart involves two constants, ρ and h. The penalty weight ρ is an input parameter and works as a threshold below which correlation are thought of as zero. When $n < p$, $\rho > 0$ corresponds to a sparse solution that fits the data less well. There is no clear connection between sparsity in $\mathbf{\Omega}$ and $\mathbf{\Sigma} = \mathbf{\Omega}^{-1}$ in general. However, our simulations suggest that large values of ρ yield sparse estimate of $\mathbf{S}$. So, a challenging question is how to choose ρ. For that purpose, we run a simulation with $p = 5$, 10, and 20 and an in-control ARL of 200. Also, we set the value of the tuning parameter ρ to 0.1, 0.2, 0.5, 1, 1.5, 1.8, and 2. We look at changes in the variance, correlation, and both variance and correlation. From Table 1, for a "variance shift", and Table 3 for "simultaneous variance and correlation shifts", it is hard to recommend a certain value of ρ which will perform well for all shifts. However, for the "correlation shift", Table 2, the LEWMC reacts faster to a shift when the tuning

parameter ρ is small. This result observed is not surprising as in this case, only the off-diagonal elements change after a shift. The LEWMC chart will hardly detect changes (unless the changes are quite large, especially with respect to the penalty weight), given that by itself it will force the off-diagonal estimated elements towards zero. So, when the tuning parameter is small, the estimate is less sparse and the chart will react faster to a change in the off-diagonal elements. Then the choice of ρ depends on the change patterns. Therefore, it is difficult to recommend one ρ value that will be adequate for all out-of-control scenarios.

Table 3. Performance comparison of the charts for variance and correlation shift.

p	δ	MEWMC	LEWMC(ρ)						
			0.1	0.2	0.5	1	1.5	1.8	2
5	0.00	199.2	199.8	199.2	199.9	199.1	199.3	199.8	199.9
5	0.10	149.2	128.1	121.6	118.9	116.3	118.3	116.6	117.0
5	0.20	96.61	81.36	75.47	73.76	74.25	74.39	74.16	74.62
5	0.30	64.15	53.69	50.17	49.94	50.51	51.53	50.90	51.20
5	0.40	44.29	37.96	35.43	35.77	36.48	37.37	37.02	38.10
5	0.50	33.57	28.32	27.01	27.80	28.50	29.27	29.02	29.42
5	0.60	26.20	22.49	21.43	21.83	22.67	23.21	23.25	23.81
5	0.70	21.19	18.37	17.56	18.09	18.81	19.28	19.50	19.41
5	0.80	17.87	15.41	15.15	15.32	15.95	16.73	16.29	16.51
5	0.90	15.66	13.47	13.21	13.38	13.92	14.34	14.11	14.41
10	0.00	200.7	199.5	199.8	200.3	199.6	199.4	199.1	199.6
10	0.10	170.7	148.1	142.7	140.1	140.5	136.1	136.2	134.2
10	0.20	132.2	106.7	100.1	98.41	97.56	95.62	96.25	94.88
10	0.30	96.99	76.08	71.04	70.08	69.84	68.83	69.00	69.59
10	0.40	71.43	54.74	52.07	51.01	52.22	52.88	52.50	52.83
10	0.50	54.84	42.21	39.01	39.75	40.74	41.52	41.10	41.46
10	0.60	42.73	33.23	31.62	31.50	33.05	33.48	33.64	33.48
10	0.70	35.02	27.18	25.11	25.91	27.29	27.67	27.96	28.25
10	0.80	29.03	22.50	21.38	21.56	22.54	23.35	23.35	23.91
10	0.90	24.21	19.48	18.67	18.82	19.80	20.14	20.20	20.64
20	0.00	200.6	199.5	199.3	201.3	199.8	200.8	200.0	201.1
20	0.10	186.7	163.1	158.9	158.1	150.7	153.5	152.1	156.4
20	0.20	160.4	128.2	122.4	120.9	116.1	120.2	117.1	119.5
20	0.30	134.6	101.3	95.96	91.47	91.80	91.69	93.29	94.22
20	0.40	109.8	79.45	74.48	72.58	72.61	73.96	74.75	74.19
20	0.50	88.69	62.80	59.98	56.96	57.46	59.37	60.42	59.25
20	0.60	72.18	50.25	47.16	47.01	47.32	48.67	49.34	49.57
20	0.70	59.75	41.38	38.59	38.62	40.11	40.69	41.47	41.37
20	0.80	49.97	34.41	32.97	32.48	33.12	34.22	35.35	35.12
20	0.90	41.85	29.67	28.04	28.20	28.69	29.85	30.30	30.57

4.5. Bivariate Case

As pointed out by one referee, we will investigate the performance in the bivariate case (2 variables X and Y) by comparing the proposed chart to the EWMA scheme based on the VMAX statistic. The VMAX statistic is based on the sample variances of the two quality characteristics. This means that in order to compute the VMAX statistic, we need to have a

sample size m. Now, to detect a change in the covariance matrix, use the statistic

$$Z_i = \lambda Y_i + (1-\lambda)Z_{i-1}, \quad i = 1, 2, \ldots, \tag{9}$$

where $Y_i = \max\{S_{x_i}^2, S_{y_i}^2\}$, $S_{x_i}^2$ and $S_{y_i}^2$ are the sample variances of X and Y and λ is a smoothing parameter $0 < \lambda \leq 1$. Z_0 is often set to the expected in-control value of Z. An out-of-control signal is obtained when $Z_i > $ CL, where CL is the control limit for the EWMA chart.

Note that the LEWMC, defined in this paper, is based on individual observations. So, for the two charts to be fairly compared, we propose a slight modification to the LEWMC chart by using samples of size m instead of individual observations. This modified chart will be called "Sample based LEWMC" and denoted by "S-LEWMC" chart.

To define the S-LEWMC chart, it is convenient to work with multistandardized data vectors instead of the original process readings $\mathbf{x}$. To achieve this, we use a matrix $\mathbf{A}$ with the property $\mathbf{A}\Sigma_0\mathbf{A}' = \mathbf{I}_p$, where $\mathbf{I}_p$ is a $p \times p$ identity matrix, and transform $\mathbf{x}$ to $\mathbf{u} = \mathbf{A}(\mathbf{x} - \mathbf{\mu}_0)$. While the process is in control, the $\mathbf{u}$'s are $N(\mathbf{0}, \mathbf{I}_p)$. We denote the ith sample of size m by $\mathbf{U}_i = [\mathbf{u}_{1i}, \mathbf{u}_{2i}, \ldots, \mathbf{u}_{mi}]'$ (an $m \times p$ matrix) where $\mathbf{u}_{ki}$ represents the kth vector $\mathbf{u}$ of the ith sample.

The S-LEWMC chart can be outlined as follows:

1. Multistandardize the data by transforming the original variables $\mathbf{x}$ into $\mathbf{u}$

2. For each sample i, compute $\mathbf{V}_i = \mathbf{U}_i'\mathbf{U}_i$

3. For each sample i, find $\hat{\mathbf{\Omega}}_i$ and then $\mathbf{W}_i$. Use the graphical LASSO algorithm of Friedman *et al.* [11] for this step to directly obtain $\mathbf{W}_i$.

4. Using $\mathbf{S}_0 = \mathbf{I}_p$, compute $\mathbf{S}_i = (1-\lambda)\mathbf{S}_{i-1} + \lambda\mathbf{W}_i$ for $i = 1, 2, \ldots$

5. Compute the chart statistic $c_i = tr(\mathbf{S}_i) - \log|\mathbf{S}_i| - p$ and detect an out of control behavior when $c_i > h$.

Next, we compare the EWMA scheme based on the VMAX statistic with the LEWMC and the S-LEWMC. We set $\lambda = 0.1$ and the in-control ARL to 200. For the EWMA and the S-LEWMC, we use samples of size $m = 4$. The LEWMC, as mentioned earlier, is based on the individual observations. We investigate the following out-of-control scheme. Case I will be a change in the variance of X, i.e., we changed the covariance matrix from its in-control value of $\mathbf{I}_2$ to a matrix having $1 + \delta$ in the (1, 1) position with the other elements unchanged. Next, case II will be a change in both the variance of X and Y, i.e., we changed the covariance matrix from its in-control value of $\mathbf{I}_2$ to a matrix having $1 + \delta$ in the (1, 1) and (2, 2) positions with the other elements unchanged. For both cases I and II, we vary δ from 0 to 1.

Table 4 and Table 5 show the performance comparison between the three charts. For both schemes, the EWMA based on the VMAX outperforms the LEWMC. This is reasonable as the EWMA chart is based on samples of size m while the LEWMC used individual observations. Next, the LEWMC based on samples of size m (S-LEWMC) reacts faster than the EWMA chart based on the VMAX statistic.

5. Diagnostics After Detecting an Out-of-Control Signal

Looking more closely at the issue of diagnosis, matters are complicated by the fact that a signal in the LASSO chart does not necessarily mean that the covariance matrix has shifted.

The signal can be triggered by several factors.

1.   A change in the mean vector

2.   A change in the relationship between the variables

3.   A change in the variability

An efficient way to monitor a multivariate process is to run simultaneously two charts. One to monitor the mean vector (Hotelling T^2 chart, MEWMA, MCUSUM for example) and the other for monitor the process variability (Alt's chart, MEWMC, CE chart for example). So, a change in the covariance matrix of the data impacts the run length behavior of the location chart, so a signal from the location chart could be an indirect indicator of an increase in the variability of the data. On the other side, as the scale chart is implicitly centered at the in-control mean vector, a shift in the mean vector accelerates signaling in the scale chart. This means that a location chart signal could be caused by a variance increase, and a scale chart signal could be caused by a shift in mean. Thus careful interpretation is necessary after a signal.

After a shift is detected, we recommend using the method proposed in Hawkins and Maboudou-Tchao [13] as a diagnostic tool to help identify the cause of the shift.

Table 4. Performance comparison of the LEWMC with the VMAX for case I shift.

p	δ	VMAX	LEWMC	S-LEWMC
2	0.00	200.6	201.2	200.1
2	0.10	105.7	132.2	92.47
2	0.20	62.06	91.91	51.60
2	0.30	40.03	66.10	32.71
2	0.40	28.21	51.01	22.57
2	0.50	21.09	40.30	17.16
2	0.60	16.66	32.51	13.50
2	0.70	13.39	27.69	10.99
2	0.80	11.37	23.46	9.39
2	0.90	9.64	20.29	8.03
2	1.00	8.43	18.27	7.10

Table 5. Performance comparison of the LEWMC with the VMAX for case II shift.

p	δ	VMAX	LEWMC	S-LEWMC
2	0.00	200.3	201.2	200.4
2	0.10	66.24	132.2	51.94
2	0.20	31.50	91.91	23.00
2	0.30	18.98	66.10	13.66
2	0.40	13.19	51.01	9.47
2	0.50	9.84	40.30	7.10
2	0.60	7.78	32.51	5.76
2	0.70	6.52	27.69	4.69
2	0.80	5.49	23.46	4.01
2	0.90	4.75	20.29	3.46

6. Example

We apply the new method to the real data set used in Hawkins and Maboudou-Tchao [13]. The data set used is from a long-standing research project in ambulatory monitoring (see www.msi.umn.edu/halberg for deeper background). In this work subjects were equipped with instruments that measure and record physiological variables. The wearer blood pressure and heart rate were measured and recorded every 15 minutes for 6 years. Before analysis using statistical process control (SPC) methods, each week raw data are condensed into weekly summary numbers, which include mean systolic blood pressure (SBP), mean diastolic blood pressure (DBP), mean of heart rate (HR), and overall mean arterial pressure (MAP). We set the smoothing constant λ to .1, the sparsity parameter ρ to 0.5, and the IC ARL to 200; thus the control limit is $h = 1.28$. The IC distribution of the readings for this particular person is assumed to be known exactly from his (long) historical sequence. It is multivariate normal with mean vector

$$\boldsymbol{\mu} = (126.61, 77.48, 80.95, 97.97)'$$

and covariance matrix

$$\boldsymbol{\Sigma}_0 = \begin{pmatrix} 15.04 & 8.66 & 10.51 & 12.04 \\ 8.66 & 5.83 & 5.56 & 7.5 \\ 10.51 & 5.56 & 15.17 & 8.79 \\ 12.04 & 7.5 & 8.79 & 10.57 \end{pmatrix}.$$

The matrix $\mathbf{A}$ used to multistandardize the data to $N(\mathbf{0}, \mathbf{I})$ is

$$\mathbf{A} = \begin{pmatrix} 0.26 & 0 & 0 & 0 \\ -0.63 & 1.09 & 0 & 0 \\ -0.38 & 0.21 & 0.36 & 0 \\ -0.45 & -1.03 & -0.13 & 1.42 \end{pmatrix}.$$

As pointed out by a referee, we give some details about the computation of the statistic c_n for the first observation. The computations are done in the following steps

1. After the multistandardization, the first observation is

$$\mathbf{u}_1 = (0.496, -0.259, -1.249, 0.398)'.$$

2. The next task is to compute $\mathbf{U}_1 = \mathbf{u}_1 \mathbf{u}_1'$. We get

$$\mathbf{U}_1 = \begin{pmatrix} 0.246 & -0.129 & -0.620 & 0.198 \\ -0.129 & 0.067 & 0.324 & -0.103 \\ -0.620 & 0.324 & 1.559 & -0.497 \\ 0.198 & -0.103 & -0.497 & 0.158 \end{pmatrix}.$$

3. By using the graphical LASSO package from R, we compute directly $\mathbf{V}_1$,

$$\mathbf{V}_1 = \begin{pmatrix} 0.747 & 0.000 & -0.121 & 0.000 \\ 0.000 & 0.567 & 0.000 & 0.000 \\ -0.121 & 0.000 & 2.059 & 0.000 \\ 0.000 & 0.000 & 0.000 & 0.658 \end{pmatrix}.$$

4. Now, with $\mathbf{S}_0 = \mathbf{I}_4$, we compute $\mathbf{S}_1 = (1-\lambda)\mathbf{S}_0 + \lambda\mathbf{V}_1$ to obtain

$$\mathbf{S}_1 = \begin{pmatrix} 0.907 & -0.012 & -0.068 & 0.018 \\ -0.012 & 0.873 & 0.029 & -0.009 \\ -0.068 & 0.029 & 1.156 & -0.045 \\ 0.018 & -0.009 & -0.045 & 0.890 \end{pmatrix}.$$

5. Finally, we can use $\mathbf{S}_1$ to compute $c_1 = tr(\mathbf{S}_1) - \log|\mathbf{S}_1| - p = 0.03876546$

In the same way, we compute $c_2, c_3, \ldots$ for $n = 2, 3, \ldots$. An R code to compute the c_n statistic is available in the appendix. So, calculating the LEWMC chart statistic c_n for the next 20 observations gives the results shown in Figure 1. The chart goes outside its control limit at the 19th observation. Note that the MEWMC also detects a signal at the 19th observation. The diagnosis analysis is discussed in Hawkins and Maboudou-Tchao [13] and is not repeated in this paper.
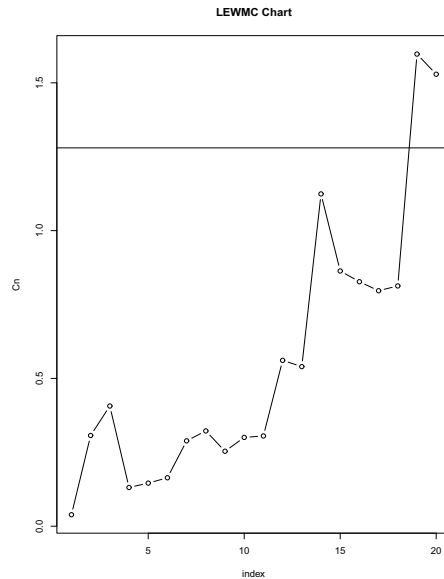

Figure 1. LEWMC chart for the Ambulatory Monitoring data.

7. Conclusion

In this paper, we have proposed a new method using the graphical LASSO to monitor small and persistent changes in a covariance matrix of a multivariate Gaussian process. The graphical LASSO is used here because of certain desirable properties of the graphical LASSO estimators such as being positive definite and full rank even when the empirical covariance is singular. We have derived some properties for the LEWMC such as being positive definite among others.

The performance of this chart was discussed and compared to some existing charts, the MEWMC chart of Hawkins and Maboudou-Tchao [13] and, in the bivariate case, the EWMA based on the VMAX statistic of Machado and Costa [20]. The simulations show that

the LEWMC outperforms the MEWMC for all but one of the scenarios considered in this current paper. The current version of this chart is designed to detect changes in the covariance matrix only. Also, it is based on the assumptions that the observations are independent of each other, follow a multivariate normal distribution, and their in-control distribution is known. In the case when the IC distribution is unknown, a self starting methodology, such as Maboudou-Tchao and Hawkins [19], may be applied to the proposed chart.

References

1. Alt, F.A. (1984). Multivariate Quality Control. *The Encyclopedia of Statistical Sciences*. In: N. L. Johnson, S. Kotz and C.R. Read (Eds.), Wiley, New York.

2. Aparisi, F., Jabaloyes, J. and Carrión, A. (2001). Generalized variance chart design with adaptive sample sizes, the bivariate case. *Communications in Statistics: Simulation and Computation*, 30, 931-948.

3. Banerjee, O., El Ghaoui, L., d'Aspremont, A. and Natsoulis, G. (2006). Convex optimization techniques for fitting sparse Gaussian graphical models. *International Conference on Machine Learning (ICML)*.

4. Banerjee, O., El Ghaoui, L. and d'Aspremont, A. (2008). Model selection through sparse maximum likelihood estimation for multivariate Gaussian or binary data. *Journal of Machine Learning Research*, 9, 485-516.

5. Bersimis, S., Psarakis, S. and Panaretos, J. (2007). Multivariate statistical process control charts: an overview. *Quality and Reliability Engineering International*, 23, 517-543.

6. Calvin, J. A. (1994). One-sided test of covariance matrix with a known null value. *Communication in Statistics: Theory and Methods*, 23, 3121-3140.

7. Chan, L. K. and Zhang, J. (2001). Cumulative sum control charts for the covariance matrix. *Statistica Sinica*, 11, 767-790.

8. d'Aspremont, A., Banerjee, O. and El Ghaoui, L. (2008). First-order methods for sparse covariance selection. *Siam Journal on Matrix Analysis and Applications*, 30, 56-66.

9. Dempster, A. (1972). Covariance selection. *Biometrics*, 28, 157-175.

10. Fan, J. and Peng, H. (2004). Nonconcave penalized likelihood with a diverging number of parameters. *Annals of Statistics*, 32, 928-961.

11. Friedman, J., Hastie, T. and Tibshirani, R. (2008). Sparse inverse covariance estimation with the graphical LASSO. *Biostatistics*, 9, 432-441.

12. Hao, S., Zhou, S. and Ding, Y. (2008). Multivariate process variability monitoring through projection. *Journal of Quality Technology*, 40, 214-226.

13. Hawkins, D. M. and Maboudou-Tchao, E. M. (2008). Multivariate Exponentially Weighted Moving Covariance Matrix. *Technometrics*, 50, 155-166.

14. Hotelling, H. (1947). *Multivariate Quality Control, Illustrated by the Air Testing of Sample Bombsights*. In: C. Eisenhart, M. Hastay, and W.A.Wallis, (Eds.). Techniques of Statistical Analysis, McGraw-Hill, New York.

15. Huwang, L., Yeh, A. B. and Wu, C. W. (2007). Monitoring multivariate process variability for individual observations. *Journal of Quality Technology*, 39, 258-278.

16. Lam, C. and Fan, J. (2009). Sparsistency and rates of convergence in large covariance matrices estimation. *Annals of Statistics*, 37(6B), 4254-4278.

17. Li, B., Wang, K. and Yeh, A. B. (2013). Monitoring covariance matrix via penalized likelihood estimation. *IIE Transactions*, 45(2), 132-146.

18. Levinson, W., Holmes, D. S. and Mergen, A. E. (2002). Variation charts for multivariate processes. *Quality Engineering*, 14, 539-545.

19. Maboudou-Tchao, E. M. and Hawkins, D. M. (2011). Self-starting multivariate control charts for location and scale. *Journal of Quality Technology*, 43, 2, 113-126.

20. Machado, M. A. G. and Costa, A. F. B. (2008). The double sampling and the EWMA charts based on the sample variances. *International Journal of Production Economics*, 114, 134-148.

21. Machado, M. A. G., Costa, A. F. B. and Rahim, M. A. (2009). The synthetic control chart based on two sample variances for monitoring the covariance matrix. *Quality and Reliability Engineering International*, 25, 5, 595-606.

22. Memar, A. O. and Niaki, S. T. A. (2009). New control charts for monitoring covariance matrix with individual observations. *Quality and Reliability Engineering International*, 25(7), 821-838.

23. Montgomery, D. C. and Wadsworth, H. M. (1972). *Some Techniques for Multivariate Quality Control Applications.* Transactions of the ASQC, Washington, D.C.

24. Montgomery, D. C. (2001). *Introduction to Statistical Quality Control*, 4th edition. Wiley.

25. Quinino, R., Costa, A. and Ho, L. L. (2012). A single statistic for monitoring the Covariance matrix of bivariate processes. *Quality Engineering*, 24, 3, 423-430.

26. Ravikumar, P., Liu, H., Raskutti, G. and Yu, B. (2009). Model selection in Gaussian graphical models: High- dimensional consistency of l_1 -regularized MLE. *Advances in Neural Information Processing Systems*, 22, Cambridge, MA, MIT Press.

27. Reynolds, R. M. and Cho, G. Y. (2006). Multivariate control charts for monitoring the mean vector and covariance matrix. *Journal of Quality Technology*, 38, 230-253.

28. Reynolds, R. M. and Cho, G. Y. (2011). Multivariate control charts for monitoring the mean vector and covariance matrix with variable sampling intervals. *Sequential Analysis: Design Methods and Applications*, 30, 1, 1-40.

29. Rothman, A., Bickel, P., Levina, E. and Zhu, J. (2008). Sparse permutation invariant covariance estimation. *Electronic Journal of Statistics*, 2, 494-515.

30. Runger, G. C. and Testik, M. C. (2004). Multivariate extensions to cumulative sum control charts. *Quality and Reliability Engineering International*, 20, 587-606.

31. Sakata, T. (1987). Likelihood ratio test for one-sided hypothesis of covariance matrices of two normal populations. *Communications Statistics: Theory and Methods*, 16, 3157-3168.

32. Tang, P. F. and Barnett, N. S. (1996). Dispersion control for multivariate processes. *Australian Journal of Statistics*, 38, 235-251.

33. Tang, P. F. and Barnett, N. S. (1996). Dispersion control for multivariate processes-some comparisons. *Australian Journal of Statistics*, 38, 253-273.

34. Vandenberghe, L., Boyd, S. and Wu, S. P. (1998). Determinant maximization with linear matrix inequality constraints. *Siam Journal on Matrix Analysis and Applications*, 19, 499-533.

35. Vargas, N. J. A. and Lagos, C. J. (2007). Comparison of multivariate control charts for process dispersion. *Quality Engineering*, 19, 191-196.

36. Wang, K. and Jiang, W. (2009). High dimensional process monitoring and fault isolation via variable selection. *Journal of Quality Technology*, 41, 247-258.

37. Yeh, A. B., Huwang, L. C. and Wu, C. W. (2005). A multivariate EWMA control chart for monitoring process variability with individual observations. *IIE Transactions on Quality and Reliability Engineering*, 37, 1023-1035.

38. Yeh, A. B., Lin, D. K.J. and McGrath, R. N. (2006). Multivariate control charts for monitoring covariance matrix: a review. *Quality Technology and Quantitative Management*, 3, 415-436.

39. Yen, C. L. and Shiau, J. J. (2010). A multivariate control chart for detecting increases in process dispersion. *Statistica Sinica*, 20, 1683-1707.

40. Yuan, M. and Lin, Y. (2007). Model selection and estimation in the Gaussian graphical model. *Biometrika*, 94, 19-35.

41. Zamba, K. D. and Hawkins, D. M. (2009). A multivariate change-point model for change in mean vector and/or covariance structure. *Journal of Quality Technology*, 41, 3, 285-303.

42. Zhang, G. and Chang, S. I. (2008). Multivariate EWMA control charts using individual observations for process mean and variance monitoring and diagnosis. *International Journal of Production Research*, 46 (24), 6855-6881.

43. Zhang, J., Li, Z. and Wang, Z. (2010). A multivariate control chart for simultaneously monitoring process mean and variability. *Computational Statistics* and *Data Analysis*, 54, 10, 2244-2252.

44. Zou, C. and Qiu, P. (2009). Multivariate statistical process control using LASSO. *Journal of American Statistical Association*, 104, 1586-1596.

45. Zou, C., Ning, X. and Tsung, F. (2012). LASSO-based multivariate linear profile monitoring. *Annals of Operation Research*, 192, 3-19.

Appendix

Properties of the LEWMC

In this section, we assume that $0 < \lambda < 1$ and $\rho > 0$. While the process is in-control, by decomposing $\mathbf{S}_{n-1}$ in terms of $\mathbf{S}_{n-2}$, $\mathbf{S}_{n-2}$ in terms of $\mathbf{S}_{n-3}$, and so on, it follows

Lemma 1.

$$\mathbf{S}_n = (1-\lambda)^n \mathbf{I}_p + \lambda \sum_{k=0}^{n-1} (1-\lambda)^k \mathbf{V}_{n-k}. \tag{A1}$$

Proof. We substitute the quantity $\mathbf{S}_{n-1}$ on the right-hand side of Equation (7) to obtain

$$\mathbf{S}_n = \lambda \mathbf{V}_n + (1-\lambda)[\lambda \mathbf{V}_{n-1} + (1-\lambda)\mathbf{S}_{n-2}]$$
$$= \lambda \mathbf{V}_n + \lambda(1-\lambda)\mathbf{V}_{n-1} + (1-\lambda)^2 \mathbf{S}_{n-2}.$$

Continuing to substitute recursively for $\mathbf{S}_{n-k}$, $k = 2, 3, \ldots$, we obtain

$$\mathbf{S}_n = (1-\lambda)^n \mathbf{I}_p + \lambda \sum_{k=0}^{n-1} (1-\lambda)^k \mathbf{V}_{n-k}.$$

The above formula shows that $\mathbf{S}_n$ is a linear combination of the identity matrix weighted by a coefficient $(1-\lambda)^n$ and the random matrices $\mathbf{V}_1, \ldots, \mathbf{V}_n$ weighted by the coefficients $\lambda(1-\lambda)^{n-1}, \ldots, \lambda$. For this reason, the sequence $\mathbf{S}_1, \ldots \mathbf{S}_n, \ldots$ is called a LASSO Multivariate Exponentially Weighted Moving Covariance Matrix (LEWMC) sequence. When $\lambda \to 0$, the sequence $\mathbf{S}_1, \ldots \mathbf{S}_n$ tends to be a smoother version of the initial sequence $\mathbf{V}_1, \ldots, \mathbf{V}_n$ and when $\lambda = 0$, then $\mathbf{S}_n = \mathbf{S}_{n-1} = \ldots = \mathbf{S}_0 = \mathbf{I}_p$.

Next, we compute the mean and variance of the matrix $\mathbf{S}_n$. The quantity $\mathbf{V}_n$ defined earlier is a random matrix and we assume that it has mean $E(\mathbf{V}_n) = \boldsymbol{\Psi}$ and covariance matrix $Cov(\mathbf{V}_n) = \boldsymbol{\Phi}$. The mean of $\mathbf{S}_n$ is given in the theorem below

Theorem 2.

$$E(\mathbf{S}_n) = (1-\lambda)^n(\mathbf{I}_p - \mathbf{\Psi}) + \mathbf{\Psi}. \tag{A2}$$

Proof.

$$E(\mathbf{S}_n) = (1-\lambda)^n\mathbf{I}_p + \lambda \sum_{k=0}^{n-1}(1-\lambda)^k E[\mathbf{V}_{n-k}]$$

$$= (1-\lambda)^n\mathbf{I}_p + \lambda\mathbf{\Psi}\sum_{k=0}^{n-1}(1-\lambda)^k$$

$$= (1-\lambda)^n\mathbf{I}_p + (1-(1-\lambda)^n)\mathbf{\Psi}$$

$$= (1-\lambda)^n(\mathbf{I}_p - \mathbf{\Psi}) + \mathbf{\Psi}.$$

As $n \to \infty$,

$$E(\mathbf{S}_n) \to \mathbf{\Psi}.$$

When $\rho = 0$, there is no regularization and $\mathbf{\Psi} = \mathbf{I}_p$. It follows that $E(\mathbf{S}_n) = \mathbf{I}_p$ which is the MEWMC case.

Furthermore, the variance of $\mathbf{S}_n$ is

Theorem 3.

$$Cov(\mathbf{S}_n) = \frac{\lambda}{2-\lambda}[1-(1-\lambda)^{2n}]\mathbf{\Phi} \tag{A3}$$

Proof.

$$Cov(\mathbf{S}_n) = \lambda^2\sum_{k=0}^{n-1}(1-\lambda)^{2k}Cov[\mathbf{V}_{n-k}]$$

$$= \lambda^2\mathbf{\Phi}\sum_{k=0}^{n-1}(1-\lambda)^{2k}$$

$$= \lambda^2\frac{(1-(1-\lambda)^{2n})}{\lambda(2-\lambda)}\mathbf{\Phi}.$$

Next, as $n \to \infty$,

$$Cov(\mathbf{S}_n) \to \frac{\lambda}{2-\lambda}\mathbf{\Phi}.$$

This means that the asymptotic mean and asymptotic covariance matrix of the process $\mathbf{S}_n$ are respectively $\mathbf{\Psi}$ and $[\lambda/(2-\lambda)]\mathbf{\Phi}$.

Theorem 4. $\mathbf{S}_n$ is a positive definite matrix.

Proof. The proof uses an argument by induction. Recall that the sequence $\{\mathbf{V}_n, n=1,2,\ldots\}$ is the penalized estimate obtained by the LASSO, so it is positive definite. Then,

$$\mathbf{S}_1 = (1-\lambda)\mathbf{S}_0 + \lambda\mathbf{V}_1 = (1-\lambda)\mathbf{I}_p + \lambda\mathbf{V}_1,$$

$\forall \mathbf{x} \in \mathbf{R}^p$ and $\mathbf{x} \neq \mathbf{0}$,

$$\mathbf{x}'\mathbf{S}_1\mathbf{x} = \mathbf{x}'((1-\lambda)\mathbf{I}_p + \lambda\mathbf{V}_1)\mathbf{x}$$

$$= (1-\lambda)\mathbf{x}'\mathbf{x} + \lambda\mathbf{x}'\mathbf{V}_1\mathbf{x} > 0.$$

For $n = 2, 3, \dots$, we use the inductive hypothesis $\mathbf{x}'\mathbf{S}_{n-1}\mathbf{x} > 0$ and show that $\forall \mathbf{x} \in \mathbf{R}^p$ and $\mathbf{x} \neq \mathbf{0}$,

$$\mathbf{x}'\mathbf{S}_n\mathbf{x} = \mathbf{x}'((1-\lambda)\mathbf{S}_{n-1} + \lambda\mathbf{V}_n)\mathbf{x}$$
$$= (1-\lambda)\mathbf{x}'\mathbf{S}_{n-1}\mathbf{x} + \lambda\mathbf{x}'\mathbf{V}_n\mathbf{x} > 0.$$

So, $\mathbf{S}_n$ is positive definite and the statistic $c_n = tr(\mathbf{S}_n) - \log|\mathbf{S}_n| - p$ is defined for all values of n.

Computer Code

An R code for the computation of the LEWMC chart is presented here. This function uses for argument a multistandardized data $\mathbf{X}$, a tuning parameter ρ, the control limit h, and the smoothing constant λ. The function LEWMC.Chart can be modified slightly if one wants to not penalize the diagonal elements of the covariance matrix. To achieve that, we just need to use V<-glasso(U,rho, penalize.diagonal = FALSE) $ w instead. Note that by doing so, we also need to change the control limit h. Remember that you need first to install the package "glasso" and load it before using the code below.

```
LEWMC.Chart <- function(x,rho,cl,lam){
##
## x: data
## rho: LASSO penalty term
## cl: Control limit
## lam: Smoothing constant term
##
#——————————————————————
if (rho<=0)
     stop("The tuning parameter must be a positive number.")
if (lam<=0 || lam>=1)
     stop("The smoothing constant must be a number between 0 and 1.")
if (cl <= 0)
     stop("The control limit must be a positive number.")
#——————————————————————
n<-dim(x)[1]; p< - dim(x) [2]
sn<-diag(p);cn<-rep( 0 ,n); sample<-rep(0 ,n);k<-0
cn<-rep(0,n)
for(i in 1:n){
k<-k+1
u<-x[i,];sample[i]<-k
U<-u%*%t(u)
V<-glasso(U,rho)$w
sn<-(1-lam)*sn +lam*V
trsn<-sum(diag(sn))
cn[i]<-trsn-log(det(sn))-p
}
plot(sample,cn,type="b")
abline(h=cl)
list(cn=cn,sample=sample)
}
```

Authors' Biographies:

Edgard M. Maboudou-Tchao is an Assistant Professor of Statistics at the Department of Statistics at the University of Central Florida - College of Sciences, Orlando, FL, USA. He holds a Ph.D. in Statistics from the School of Statistics at the University of Minnesota-Twin Cities, Minneapolis, MN. His publications have appeared in peer-reviewed journals including Technometrics and Journal of Quality Technology. His current research is focused on Multivariate Statistics, Statistical Process Control, Change-Point problems, Data Mining, and High Dimensional Problems.

Norou Diawara is an Associate Professor of Statistics at the Mathematics & Statistics Department at Old Dominion University in Norfolk, VA, USA. He holds a Ph.D. in Mathematical Statistics from Auburn University in Auburn, AL. His research is focused on Functional Data Analysis, Multivariate Data Analysis and Modeling, Probability Theory and its Applications, Biostatistics, and Time Series.