Applied Mathematical Sciences, Vol. 8, 2014, no. 70, 3491 - 3499
HIKARI Ltd, www.m-hikari.com
http://dx.doi.org/10.12988/ams.2014.44298

On Monitoring Shift in the Mean Processes with Vector Autoregressive Residual Control Charts of Individual Observation

J. K. Wororomi

Department of Statistics
Institut Teknologi Sepuluh Nopember, Surabaya, Indonesia

M. Mashuri, Irhamah

Department of Statistics
Institut Teknologi Sepuluh Nopember, Surabaya, Indonesia

A. Z. Arifin

[3]Department of Informatics Engineering
Institut Teknologi Sepuluh Nopember, Surabaya, Indonesia

Abstract

In this paper, we purpose a monitoring scheme of individual observation that can be modeled as a first-order vector autoregressive process and develop it to mean square successive difference matrix based on residuals. Due to the effect of shift in mean process on VAR residual charts, we investigate the T^2-type charts performance of $T^2_{\mathrm{VAR(1)}}$ residual and T^2_{D} control charts as an individual observation. Then we compare both T^2-type charts using upper control limit approximation with α=0.005 (fix) based simulation. From the results of comparison, it shows that the T^2_{D} chart for a small to moderate level of shift in the process mean vector gives a better performance than the $T^2_{\mathrm{VAR(1)}}$ charts.

Conversely, the T_{D}^2 charts indicate a fault detection in large level of shift and the $T_{\mathrm{VAR(1)}}^2$ chart is more reliable on outliers detection. Thus, the result may not only provide a good option for quick response on monitoring scheme for residual, but also provide the way to avoid fault detection in a multivariate process.

Keywords: Vector autoregressive models (VAR) residual, $T_{VAR(1)}^2$ and T_D^2 chart

1 Introduction

In univariate autoregressive model, disturbances are a set of effects of various factors. These factors have serial correlations that come from variables process in which contribute variation in system. Some of them are unmeasureable, while others variables probably either measurable or conversely even if modeled in univariate simple model. In this situation, some of disturbance components should be potential as an auto correlating processes in the vector of autoregressive component. On the other hand, in VAR model, ones a component would be predict, it is usually refer to order determination for ensure a consistency criteria of model. VAR models are important class for analyzing multivariate time series data. In this case, it also requires an optimum procedures or criteria to optimize the VAR model.

Due to general aspect on order determination in time series theory, it should be determine by a simulation empirically, while the approximation of mean square matrix always increase with increasing of *p* value. On the other hand, the advantage of VAR(*p*) models approach is referring to dynamic process that make it meaningfully because more reasonable to the real processes. A real example can be seen in VAR(*p*) models that, all of systematic variables will be accommodated in this model, and just white noise and error measurement as a part of disturbance. Other advantage of residual VAR model is it can be estimated with OLS procedure instead of the more complex MLE procedure [4], [5] and [6]. In this study, we considers the problem of monitoring the mean vector of a process in which observation can be modeled as a first-order vector autoregressive VAR(1) process and development study of mean square successive differences process base on residuals. Due to the effect of shift in mean process on VAR residual chart by [6], we investigate the performance for T^2-type charts to determine the control limit of monitoring system.

Moreover, the circumstance of assumption about autocorrelation structure in a process still remain other easy circumstance for resolve either through on data generating process as well as on multivariate normal generating based on uncorrelated residuals. Properties of stationary or time invariant on data generating using VAR(*p*) models are important requirement that can be used in determining estimator behavior and forecast calculation, as well as in interval forecasting. The properties mention above, also using to ensure mean, variance, and auto-covariance invariant over time. Main examination of test can be dealt to model stability checking, according to fact that stability hence stationary [4].

Following to non random effect on multivariate time series, shift of structure parameter as impact of accumulation effect on shock occurrence in variables. Commonly, those accumulation effects can be determined via coefficient matrices of MA representation by using numeric approximation due to structural change as an impact of shift processes. Such as in [4], [5] and [6], there were an approximation by using control chart constructing due to controlling of structure changes on mean shift, disperse shift, and coefficient matrices shift. There are several alternative to constructing T^2 statistic due to structure chance. For example, the successive-difference estimator proposed by Holmes and Mergen (1993) which recommended by [2] and [3] when random variation is present in the in-control process data.

However, for some multivariate non-random distributions, the T^2 chart based on known in-control parameters has an excessive false-alarm rate as well as a reduced probability for detecting shift in the mean vector can be performed by the T_D^2 statistic on [7] will be purposed here as a monitoring scheme to improve the $T_{VAR(1)}^2$ for residual.

The remaining part of this paper is organized as follows: section 2 contains discussion of materials and methods which contain the concept of vector autoregressive model and FPE criteria in time series theory, also the control charts of mean square successive will be reviewed in this section. The result of the research work based simulation and discussion will be discussed in section 3. Finally, the conclusion will be made in section 4.

2 Material and methods

2.1 Vector autoregressive model

Alwan and Roberts's (1988) monitoring the ARMA residuals on standard Shewhart control charts and then extended to the multivariate case. In order to facilitate the development of estimators in quality monitoring system, the MLE or LS estimators can be used as standard estimator in estimating the model parameters [4], [5], and [6]. This estimator is a common method used in estimating a multivariate normal random variable such as the residual. Let *n* variable a VAR process as $\mathbf{Y}_t = (\mathbf{Z}_{1t}, \mathbf{Z}_{2t}, \ldots, \mathbf{Z}_{nt})'$ with size $(n \times 1)$ and VAR model with *p*-th or VAR(*p*) expressed as

$$\mathbf{Z}_t = \boldsymbol{c} + \left(\mathbf{\Phi}_1 L + \mathbf{\Phi}_2 L^2 + \cdots + \mathbf{\Phi}_p L^p\right)\mathbf{Z}_t + \mathbf{u}_t \quad (1)$$

with L as backshift operator, $\boldsymbol{c} = (\boldsymbol{c}_1, \boldsymbol{c}_2, \ldots, \boldsymbol{c}_n)'$ constant vector, and $\mathbf{u}_t = (\mathbf{u}_{1t}, \mathbf{u}_{2t}, \ldots, \mathbf{u}_{nt})'$ as error vector. Here, the vector error assumed independent of time, without the possibility there are also several sub-matrix cross-correlated. Nevertheless, the initial assumption of the error vector to be used is normally distributed or $\mathbf{u}_t \sim N\left(\mathbf{0}, \mathbf{\Sigma}_{\mathbf{u}_t}\right)$. In this case, a process which is under the control indicates that the assignable cause is fixed (invariant) with respect to time [5] and [6].

The procedures in numerical simulations based on random observation was modeled on [5] and [6] begins with estimating the in-control process parameters on phase I. Let, $\mathbf{B}' = (\boldsymbol{c}'\boldsymbol{\Phi})$ where $\boldsymbol{\Phi} = (\boldsymbol{\Phi}_1, \boldsymbol{\Phi}_2, \dots, \boldsymbol{\Phi}_p)$ and vector $\mathbf{X}' = (1, \mathbf{Y}_{t-1}, \mathbf{Y}_{t-2}, \dots, \mathbf{Y}_{t-p})'$

$$\mathbf{Y}_t = \mathbf{B}'\mathbf{X}_t + \mathbf{u}_t \quad (2)$$

In the case of VAR (p) as a Gaussian process, with $\mathbf{u}_t \sim N(\mathbf{0}, \boldsymbol{\Sigma}_\mathbf{a})$, the coefficient matrix $\boldsymbol{\Pi}'$ can be estimated through OLS procedure as

$$\widehat{\mathbf{B}}' = [\textstyle\sum_{t=1}^{T} \mathbf{Y}_t\mathbf{X}_t'][\sum_{t=1}^{T} \mathbf{X}_t\mathbf{X}_t']^{-1} \quad (3)$$

with T as the number of observations in phase I. Estimates of the VAR model and residual respectively given as $\widehat{\mathbf{Y}}_t = \widehat{\mathbf{B}}'\mathbf{X}_t$ and $\widehat{\mathbf{u}}_t = \mathbf{Y}_t - \widehat{\mathbf{Y}}_t$

Thus, to explore the distribution of residuals of $\widehat{\mathbf{u}}_t$, variance-covariance matrix is estimated as

$$\widehat{\boldsymbol{\Sigma}}_{\mathbf{u_t}} = \frac{1}{T}\textstyle\sum_{t=1}^{T} \widehat{\mathbf{u}}_t\widehat{\mathbf{u}}_t' \quad (4)$$

The residual $\widehat{\mathbf{u}}_t$ as i.i.d asymptotically normal with mean zero will be satisfied. In this case, the Hotelling T^2 control charts for $\widehat{\mathbf{u}}_t$ can be expressed as

$$\boldsymbol{T}_i^2 = \widehat{\mathbf{u}}_t'\widehat{\boldsymbol{\Sigma}}_{\mathbf{u_t}}^{-1}\widehat{\mathbf{u}}_t \sim \chi^2 \quad (5)$$

On an identification of data pattern, fitting plot will be performed to the VAR(p) residual based on (5). The identification of patterns referring to the fitting residual plots at each different lag-k of the input time series data. Estimation of VAR(p) residuals parameter via FPE or MLE given in the following section.

2.2 Information criteria based FPE

To obtain a unique solution the white noise covariance matrix $\boldsymbol{\Sigma}_\mathbf{u}$ or variance-covariance matrix in (4) can be replaced by an estimate [4]. Here, $\widetilde{\boldsymbol{\Sigma}}_\mathbf{u}(m)$ is the ML estimator of $\boldsymbol{\Sigma}_\mathbf{u}$ obtained by fitting a VAR(m) model. It's refer to the LS estimator with degrees of freedom adjustment in form by

$$\widehat{\boldsymbol{\Sigma}}_\mathbf{u}(m) = \frac{T}{T-Km-1}\widetilde{\boldsymbol{\Sigma}}_\mathbf{u}(m) \quad (6)$$

And when take the determinant of (6). The criterion is called the Final Prediction Error (FPE) criteria, and can be expressed as

$$\mathrm{FPE}(m) = det\left[\frac{T + Km + 1}{T}\,\frac{T}{T - Km - 1}\widetilde{\boldsymbol{\Sigma}}_u(m)\right]$$
$$= \left[\frac{T+Km+1}{T-Km-1}\right]^K det\,\widetilde{\boldsymbol{\Sigma}}_u(m) \quad (7)$$

So, optimization criterion of the residuals covariance matrix through the FPE can be chosen so that

$$\mathrm{FPE}[\hat{p}(\mathrm{FPE})] = min\{\mathrm{FPE}(m) | m = 0,1, \dots, M\} \quad (8)$$

According to [4], the VAR model of order $m = 0,1, \dots, M$ are not only an estimated of the $\mathrm{FPE}(m)$ value, but also as the parameters restriction of the VAR model due to coefficients matrix, $\boldsymbol{\Phi}$. In this case, an order that minimizes the FPE chosen as the estimated value for p can be obtained by minimized the forecast mean square error (MSE) of equation (4).

On the other hand, computing of the $\boldsymbol{T_i^2}$ value in equation (5) for each observation and it comparison of the various values of UCL will be used for identifying outliers in a multivariate processes. In this process identification, the UCL values approximation from [1], [2], and [3], should be represented as the upper α^{th} quantile of the Chi-square, Beta and F distribution. Observation with $\boldsymbol{T_i^2}$ value greater than UCL are identifying as an outliers. Here, the three type of UCL has been proposed by [1], [2], and [3] to approximate Hotelling T^2 charts and we apply to different level of shift in mean vector proportionally.

2.3 Mean square successive difference control charts

According to study about performance and control charts based on non-random observation require some capable scheme of monitoring system. Moreover, the use of the sample mean vector and the mean square successive difference matrix in T^2 control chart is sensitive in detecting process mean shift or trend but less sensitive in detecting outliers [7]. Therefore, in this paper the estimator in monitoring the variability refers to the statistical mean successive difference matrix. This statistic has been used previously by [7] in estimating the covariance matrix of the T^2 control chart to detect a trend or shift of the mean process. Estimation of location and dispersion denoted as

$$\bar{\mathbf{x}} = \frac{1}{m}\sum_{j=1}^{m}\mathbf{x}_j \text{ , and}$$

$$\mathbf{S}_D = \frac{1}{2(m-1)}\sum_{j=2}^{m}(\mathbf{x}_j - \mathbf{x}_{j-1})(\mathbf{x}_j - \mathbf{x}_{j-1})' \quad (9)$$

with positive definite matrix $\mathbf{S}_D$ as unbiased estimator from $\boldsymbol{\Sigma}$. Next, the statistic T^2 was proposed by Sullivan using $\mathbf{S}_D$ can be rewrite as

$$T_D^2 = (\mathbf{x} - \bar{\mathbf{x}})\mathbf{S}_D^{-1}(\mathbf{x} - \bar{\mathbf{x}})' \quad (10)$$

In a further development, [7] had proposed the weighted sample mean successive mean vector and the weighted mean square successive difference matrix.

3 Results

In this study, we use woodmod dataset from [8] in order to determine sensitivity of limit control based $T_{VAR(1)}^2$ and T_D^2 control charts. The result of the simulation is depicting by figure 2- 6, and will be briefed by description on figure 1 as follows:

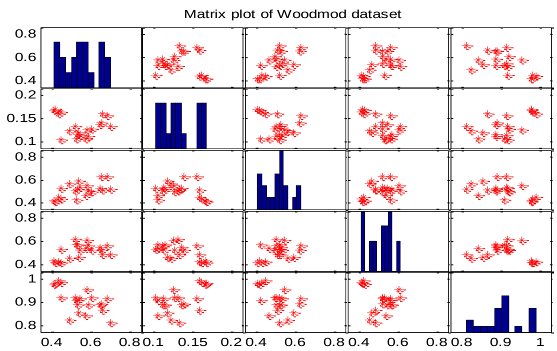

Figure 1. Matrix plot of woodmod dataset

We explore the performance of two type control charts. Each observation of woodmod dataset, $\boldsymbol{x}_{ip} = (\boldsymbol{x}_{i1}, \boldsymbol{x}_{12}, \boldsymbol{x}_{i3}, \boldsymbol{x}_{14}, \boldsymbol{x}_{i5})'$ shows five variables which described on the figure1 are respect to number of fibers per square milliliter in springwood, number of fibers per square milliliter in summerwood, fraction of springwood, fraction of light absorption by springwood, and fraction of light absorption by summerwood.

The example as mention above is a part of real-life production process and we wish to show the effectiveness of using VAR residual chart to detect special causes in multiple and or/ multivariate time series. The raw data are found having serial correlation in five variables.

According to generation of VAR residuals chart procedure for obtain an appropriate covariance matrix from woodmod dataset, we estimated the VAR parameters using least squared method, and choose a VAR(1) for unconstrained models as appropriate VAR process for residual with smallest FPE criteria as parsimonious model.

3.1 VAR(1) modeling

In VAR(1) and VAR(2) modeling, we chose VAR(1) based on FPE criteria by VAR(1) unconstraint as a parsimonious criteria which minimizing MSE of residuals, and its presented table 1.

Table 1. Information criterion based on minimum FPE

Information criteria				
Models	VAR(1)		VAR(2)	
criteria	unconst.	consts.	unconts.	conts.
AICC	-1.650	-0.647	5.162	7.692
HQC	-2.380	-1.852	-1.441	-1.536
AIC	-2.590	-2.105	-1.782	-1.911
SBC	-1.347	-0.614	0.691	0.810
FPEC	**0.080**	0.136	0.342	0.400

By refer to table 1, the equation (2) can be rewrite in VAR(1) unconstraint (i.e. unconst.) as

$$\mathbf{y}_t = \begin{bmatrix} -0.05 \\ 0.05 \\ 0.02 \\ -0.01 \\ 0.09 \end{bmatrix} + \begin{bmatrix} 0.23 & 0.75 & -0.12 & 0.20 & -0.64 \\ -0.11 & -0.45 & 0.42 & -0.40 & -0.16 \\ 0.13 & 0.39 & -0.15 & 0.04 & -0.31 \\ -0.41 & 0.64 & -0.16 & 0.83 & -0.56 \\ -0.25 & -0.71 & 0.23 & -0.04 & 0.22 \end{bmatrix} \mathbf{x}_{t-1} + \mathbf{u}_t \qquad (11)$$

With

$$\Sigma_{\mathbf{u}_0} = \begin{bmatrix} 0.48 & -0.11 & 0.38 & 0.35 & -0.25 \\ -0.11 & 0.53 & -0.24 & -0.49 & 0.20 \\ 0.37 & -0.24 & 0.84 & 0.37 & -0.33 \\ 0.35 & -0.49 & 0.37 & 0.76 & -0.18 \\ -0.25 & 0.20 & -0.33 & -0.18 & 0.48 \end{bmatrix}$$

Based on this fact, a recursive form of the system on (11) will be the main options that can be performed in this model.

3.2 The T_{D}^2 control charts based on VAR(1) residual model

The true process is unknown, but we can investigate the consistency of the VAR(1) models based on T_D^2 term. Particular, investigation of the VAR residual model from equation (5) and (10) that can be interpreted on $T_{VAR(1)}^2$ and T_D^2 before and after occurring of shift on mean vector as shown in fig. 2 using UCL_B respectively.

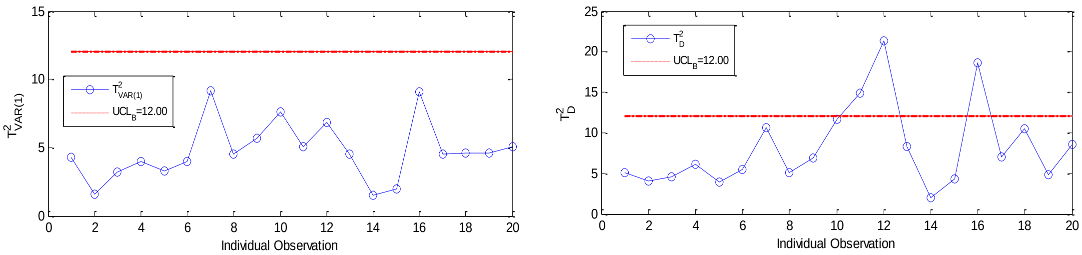

Figure 2. $T_{VAR(1)}^2$ and $\mathrm{T}_{\mathrm{D}}^2$ control chart using $UCL_B \approx 12.00$

We set shift in mean vector (δ) equal to 0.005, 0.01, 0.025, 0.1, 0.2, 0.5, 0.6, and 0.7 for evaluate the performance of both the T^2-type charts using $UCL_B \approx 12.00$ for small shift in mean vector, in these case $\delta = 0.005, 0.01, \text{and } 0.025$. From fig.2 to fig.3 (a)-(b), T_{D}^2 control charts shows more sensitive than $T_{VAR(1)}^2$ before and or/ after shift occurring in mean process.

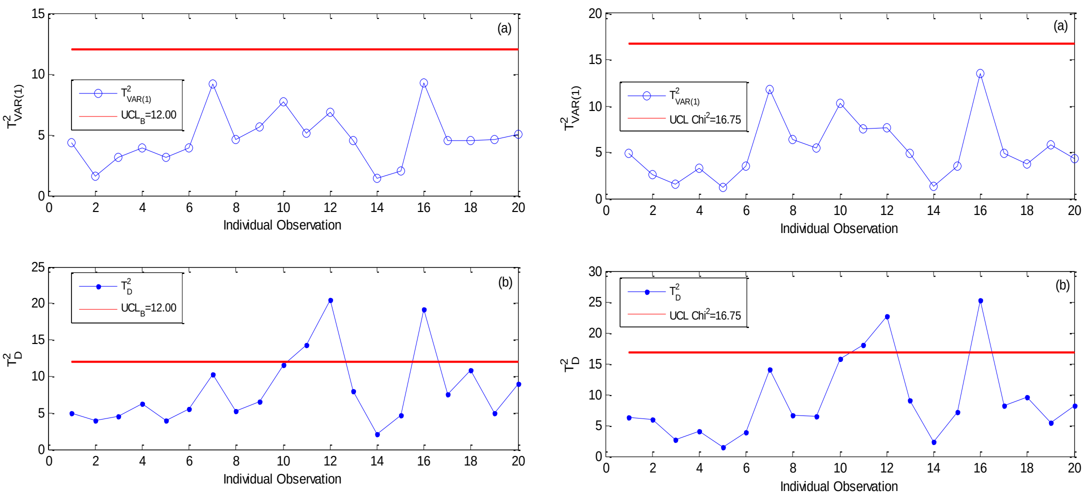

Figure 3. T^2-type control charts using $UCL_B \approx 12.00$ and $UCL_{Chi^2} \approx 16.75$

Consider to [8], to avoid a fault detection, here we are applying the T^2-type control chart using UCL_B, UCL_{Chi^2} and UCL_F with false alarm rate $\alpha = 0.005$ (fixed) for the moderate $(0.1 \leq \delta \leq 0.2)$ and large $(0.2 \leq \delta \leq 0.7)$ shift in mean vector.

Fig.3 and fig.4 depict $T^2_{VAR(1)}$ and T^2_{D} control charts on part (a) and part (b) respectively using UCL_B, UCL_{Chi^2} and UCL_F; and it has been showed that shift in mean vector increasing in moderate level still no make sense for these chart using UCL_{Chi^2} (similar to the case of the small level one). Even if, when we use UCL_F on large level of shift in mean vector, both of these T^2-type control charts show a different response, but quick in outlier detection. However, the $T^2_{VAR(1)}$ chart is more reliable for this level (large shift). Conversely, the T^2_{D} charts indicate the fault detection in this level though UCL_F with largest interval from its lower limit control.

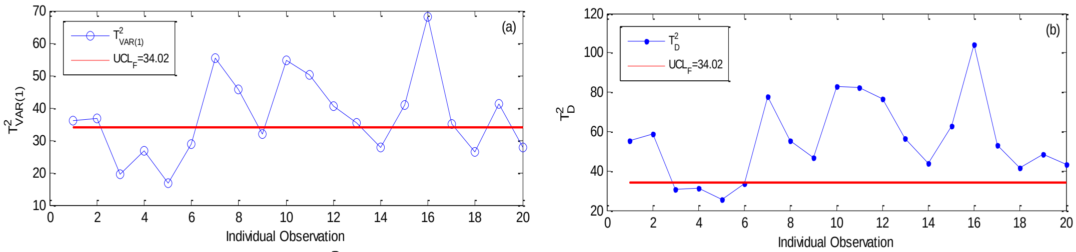

Figure 4. T^2-type control charts using $UCL_F \approx 34.02$

4 Conclusions

In multivariate process system with presence of serial correlation, vector autoregressive models (VAR) should be used to approximate the system, estimate and monitor the VAR residuals as a serially independent multivariate series. In this study, we purpose T^2-type control charts of the VAR residual model as $T^2_{VAR(1)}$ and T^2_D control chart from mean square successive difference estimator for monitoring the shift effect in mean vector.

In examining the outlier detection in the parameters process can refer to level of shift in mean vector which correspondence to approximate of upper control limit based on individual observation with fixed false alarm rate.

Moreover, we found that T^2_D charts more sensitive and quick react to detect the outliers than $T^2_{VAR(1)}$ charts in small to moderate shifts level, include the treatment for the shift occurring. Conversely, in the case of shift in mean vector increases to large level, the performance of T^2_D charts shows poor in outliers detection on such level, because it's indicate a fault detection (figure 6b) using UCL_F. In these situation, $T^2_{VAR(1)}$ charts shows it's capability on detect the outliers and more reliable than the other one.

Thus, based on this simulation by applying the woodmod dataset as an individual observation, we conclude that our scheme on monitoring shift in mean process with VAR(1) residual and the mean square successive difference estimator for residual should be improve T^2 VAR residuals in [6] to monitor multivariate process in the presence of serial correlation. Nevertheless, the use of both control charts showed their potential usefulness that not only provide a good option for quick response on monitoring scheme for residual, but also provide the way to avoid fault detection manner in a multivariate process.

Acknowledgements.
This work is supported by Hibah Doktor 2013-Direktorat Pendidikan Tinggi (DIKTI), Republik Indonesia.

References

[1] R.L. Mason et.al. *A practical approach for interpreting multivariate T^2 control chart signals*, Journal of quality technology, vol.29/no.4/p.369-501, 1997.
[2] R.L. Mason et.al. *Systematic patterns in T^2 charts*, Journal of quality technology; 35,1;ABI/INFORM global, 2003.
[3] D.C. Montgomery, *Introduction to statistical quality control*, 5e edition, Wiley international edition, (2005).
[4] H. Lütkepohl, *New introduction to multiple time series analysis*, Springer-Verlag Berlin Heidelberg, 2005.
[5] X. Pan, *Notes on shift effects for T^2-type charts on multivariate ARMA residuals*, Computers and industrial engineering 49 (2005) 381-392.
[6] X. Pan and J. Jarrett, *Using vector autoregressive residual to monitor multivariate processes in presence of serial correlation*, Int. J. Production economics 106 (2007) 204-216.
[7] J.N Pan, and S.C. Chen, *New robust Estimator for detecting non-random pattern in multivariate control charts: a simulation approach*, Journal of statistical computation and simulation, vol.81, no.3, (2011), 289-300.
[8] G. Stefatos, and Ben Hamza, *Fault detection using robust multivariate control chart*, Expert system with applications 36 (2009) 5888-5.

Received: April 24, 2014