This working paper series is intended to facilitate discussion and encourage the exchange of ideas. Inclusion here does not preclude publication elsewhere. It is the original work of the author(s) and subject to copyright regulations.
VECTOR AUTOREGRESSION AND MONITORING

MULTIVARIATE

AUTOCORRELATED PROCESSES

Xia Pan, Ph.D.
University of Illinois at Springfield

Jeffrey E. Jarrett, Ph.D.
University of Rhode Island
(jejarrett@mail.uri.edu)
ABSTRACT

The literature on statistical quality control considers both multivariate control charts for independent processes and univariate control charts for autocorrelated processes separately. This paper combined the two situations and proposes VAR control charts for monitoring multivariate (cross-sectional) and serially correlated processes. We discuss also the estimation and model selection issue. Estimating vector AR(p) model instead of vector ARMA model for the systematic cause is justified. We suggest the design of the VAR control chart. We examine the effects of parameter shifts. One will observe examples of using the VAR chart and the results indicate the feasibility of VAR control charts.
INTRODUCTION


Process control charts that monitor multivariate autocorrelated processes can be constructed by combining the Alwan and Roberts's (1988) residual chart and the traditional Multivariate Hotelling $T^2$ chart. This generalization extending Alwan and Roberts’s special cause approach to multivariate cases is new. The residuals, or the one-step ahead forecasting errors of the multivariate autocorrelated processes, need to be acquired at first. Since the multivariate one-step-ahead forecasting errors are approximately independent identical multivariate normal distributions, they can be monitored with traditional multivariate chart for independent process. Several methods can be used to estimate the multivariate time series processes to acquire the one-step-ahead forecasting errors.

Many practical production processes can be both multivariate and autocorrelated and it may be effective to use a control chart(s) particularly designed for monitoring these processes.
Examples of these kinds of processes can be found in traditional industries and high tech industries. For instance, in optical communication products manufacturing, the production of fiber optic is based on SiO$_2$ rod made from condensation of silicon and oxygen gases. The preparation of SiO$_2$ rod needs to monitor variables such as temperature, pressure, densities of different components, and intense of the molecular beams etc. Similar processes can also be found in chemical industry and semiconductor industry where materials are prepared and made. In service industry, many processes can be also autocorrelated because of, say, the habit and inertia of human behavior, and there are interactions between people's behaviors. For instance, the number of visits to a restaurant at a tourist attraction may be serially dependent and related to the occupation rate of a nearby hotel and the prices of airlines connecting to the place, while the latter quantities are also autocorrelated and cross-correlated each other.

As general phenomena in economy and business, variables are both cross-sectionally and serially correlated and studying these variables as a system can capture more information from the system. Examples include that a company's sales may be related to internal variables such as inventory, account receivable, labor and materials costs, and external variables such as outputs, competitors' prices, specific demands, economic situations, etc. They are all autocorrelated too. The company's operational performance may be better assessed with several variables simultaneously together. Therefore, monitoring multivariate autocorrelated processes instead of separately monitoring each of the processes is more effective for understanding the system. In this paper, we will introduce vector autoregression (VAR) chart for monitoring this kind of multivariate autocorrelated processes.

In addition, up-to-date literature in statistical process control emphasizes the properties of charts with little attention paid to process parameter estimation. For ordinary Shewhart chart the mainly concerned process parameter is the variance or standard deviation. Estimation is less important because the effects of different estimators of process variance are usually indifferent under the rough criterion of average run length (ARL). For autocorrelated processes, estimation is
the key procedure in chart construction. Adopting practically workable estimators is also an important issue.

The remaining of this paper is arranged as follows: In Section II, the principles of the traditional multivariate charts for independent data and the traditional special cause chart for univariate correlated data that provide the two origins of our new approach, will be reviewed. We discuss the issues of parsimonious estimation of ARIMA models. Section III and Section IV describe the principles and design procedures of VAR chart respectively. Section V examines the effects of parameter shifts. Finally, in Section VI, we present examples of VAR charts.

Like in the univariate case, we consider only the case of individual samples because the autocorrelation depends on the time-interval of sampling. If the sample size were not one, there would be autocorrelation both within and between samples. We treat this situation as unequal time-interval sampling of individual observations. In this paper, we only discuss models based on equal time-interval sampling of individual samples.

**VECTOR AUTOREGRESSION CONTROL CHARTS**

For a multivariate autocorrelated process system of \( n \) variables, denote \( \tilde{y}_t = (y_{1t}, y_{2t}, \cdots, y_{nt})' \) as the \((n \times 1)\) vector of the \( n \) variables. These variables are autocorrelated and cross-correlated and the highest order of autocorrelation is \( p \). The processes follow a vector autoregression (VAR) model

\[
\tilde{y}_t = \bar{c} + (\Phi_1 L + \Phi_2 L^2 + \cdots + \Phi_p L^p)\tilde{y}_{t-1} + \varepsilon_t = \bar{c} + \Phi(L)\tilde{y}_{t-1} + \varepsilon_t
\]  

(1)

where \( L \) is the backshift operator, \( \bar{c} = (c_1, c_2, \cdots, c_n)' \) is the constant vector, and \( \varepsilon_t = (\varepsilon_{1t}, \varepsilon_{2t}, \cdots, \varepsilon_{nt})' \) is the error term vector. The mean of the variable vector is \( \tilde{\mu} = \bar{c} + \Phi(L)\tilde{\mu} \). Each \( \Phi_j \) is an \((n \times n)\) coefficient matrix for the \( j \)th lag. The error term vector \( \varepsilon_t \) is serially uncorrelated but cross-sectionally correlated. That is, \( E(\tilde{\varepsilon}_t \tilde{\varepsilon}'_t) = \Omega \) is invariant.
about time but may not be a diagonal \((n\times n)\) matrix. Strictly we can assume \(\tilde{e}_t \sim \text{i.i.d. } N(0,\Omega)\). Some of the variables may not be autocorrelated with the same order of \(p\) themselves as well as each other, so that some elements of the matrix \(\Phi_j\) may be zero.

In this model, the AR coefficient matrix \(\Phi = \begin{pmatrix} \Phi_1 & \Phi_2 & \cdots & \Phi_p \end{pmatrix}\) and the constant term \(\tilde{c}\) represent the systematic causes of the processes, and the error term vector \(\tilde{e}_t\) represents nonsystematic disturbances. We view the disturbances as the whitened remains of the processes after one filters out the systematic pattern. The occurrences of special-causes are the unusual changes in the disturbances. The basic idea for monitoring the process is the same as Alwan and Robert’s residual approach for the univariate cases that filters the systematic patterns from the processes and then monitors the residuals for special-causes with multivariate Shewhart chart.

One estimates the systematic patterns from analyzing the history of the process. For in-control status, the systematic patterns are time-invariant. As in univariate case, we apply a two-phase construction procedure of control chart. In Phase I where the process is in-control, we estimate the systematic patterns and we acquire the one-step-ahead forecasting errors. Multivariate regression needs larger data size for asymptotic estimates. In Phase II, Hotelling \(T^2\) statistic of the multivariate residuals is monitored chart according to chi-square degree of freedom of order equal to the number of variables in the system.

Specifically, denote \(\Pi^\prime = \begin{pmatrix} \tilde{c} & \Phi \end{pmatrix}\) being an \(n \times (1+np)\) matrix, and \(\tilde{x}_t' = (1, \tilde{y}_{t-1}', \tilde{y}_{t-2}', \cdots, \tilde{y}_{t-p}')'\) being a \(1 \times (1+np)\) vector. Then (14) becomes

\[
\tilde{y}_t = \Pi^\prime \tilde{x}_t + \tilde{e}_t
\]

From the assumption about \(\tilde{e}_t\), with ordinary least squares in each row equation, we estimate the coefficient matrix \(\Pi^\prime\). For Gaussian AR\((p)\) processes, \(\tilde{e}_t \sim \text{i.i.d. } N(0,\Omega)\), these OLS estimates
are the same as the conditional maximum likelihood estimates based on historical observations (see, i.e., Hamilton, 1994). The estimates are given by

$$\hat{\Theta} = \left[ \sum_{t=1}^{T} \hat{y}_t \hat{x}_t \mid \sum_{t=1}^{T} \hat{x}_t \hat{x}_t' \right]^{-1}$$

(3)

where $T$ is the number of observations in Phase I. The estimated systematic model is

$$\hat{y}_t = \hat{\Theta} \hat{x}_t$$

(4)

and the error term is estimated as

$$\hat{\epsilon}_t = \hat{y}_t - \hat{y}_t$$

(5)

One establishes the special-cause chart based on $\hat{\epsilon}_t$. The covariance of $\hat{\epsilon}$ is estimated through MLE as

$$\hat{\Omega} = (1/T) \sum_{t=1}^{T} \hat{\epsilon}_t \hat{\epsilon}_t' \rightarrow_p \Omega$$

(6)

If the processes are in-control and the model in (7) is appropriate, under the assumption of i.i.d. normality for the error term, the estimated error $\hat{\epsilon}_t$ should be also asymptotically multivariate normal and asymptotically uncorrelated about time, specifically, it is asymptotically $N(0, \Omega)$. This is guaranteed by that the coefficient estimates, though probably biased for small samples, are consistent (see, e.g., Hamilton, 1994),

$$\sqrt{T} [ vec(\hat{\Theta}') - vec(\Theta')] \rightarrow_L N(0, \Omega \otimes Q^{-1})$$

(7)

where $(1/T) \sum_{t=1}^{T} \hat{x}_t \hat{x}_t' \rightarrow Q$. Thus, we treat each of the elements of $\hat{\Theta}'$ as an approximate normal distribution with mean at the true value. Generalized Wald test can be used to test significance of each of the element estimates with restriction $R$,

$$T (R vec(\hat{\Theta}) - r)' (R(\hat{\Theta} \otimes Q^{-1})R')^{-1} (R vec(\hat{\Theta}) - r) \sim \chi^2(m)$$

(8)
although for chart control chart purposes, it is usually not necessary to test the elements except some specific parameter elements are of special interest.

The multivariate Hotelling $T^2$ chart for independent process can be applied to $\hat{e}_i$. Thus, in Phase I the Hotelling $T^2$ statistic is, noting the individual samples, approximately $\chi^2(n)$ when $T-n$ is large.

$$T^2 = (\hat{e}_i - \bar{e})' \hat{\Omega}^{-1} (\hat{e}_i - \bar{e}) \sim \chi^2(n)$$  \hspace{1cm} (9)

where $\bar{e}$ is mean the of Phase I $\hat{e}_i$, $\bar{e} = \frac{1}{T} \sum_{i=1}^{T} \hat{e}_i$, which should be close to zero for large number of observations in Phase I. If the asymptotic mean is used, the Phase I statistic is

$$T^2 = \hat{e}_i' \hat{\Omega}^{-1} \hat{e}_i \sim \chi^2(n)$$  \hspace{1cm} (10)

Since the estimates of the AR(p) model are asymptotic available, we should use chi-square rather than Beta distribution.

For a Phase II observations $\hat{y}_i$, $(i>T$, we denote the first Phase II observation as the $T+1^{th}$ observation), the estimated error is substituted by the one-step-ahead forecasting error, while the first Phase II observation is compared with the one-step-ahead forecast from Phase I. That is, from the known $T$ observation in Phase I, the first Phase II status can be predicted as

$$\hat{y}_{T+1} = \Pi \hat{x}_{T+1}$$  \hspace{1cm} (11)

and the forecasting error is the difference between the real observation and the forecast,

$$\hat{e}_{T+1} = \hat{y}_{T+1} - \hat{y}_{T+1}.$$  \hspace{1cm} (12)

There are several possible schemes for Phase II chart. If we treat $\hat{\Omega}$ as Wishart distribution, the chart statistic is

$$T_{T+1}^2 = (\hat{e}_{T+1} - \bar{e})' \hat{\Omega}^{-1} (\hat{e}_{T+1} - \bar{e}) \sim n(T+1)(T-1) \frac{F_{\alpha,n,T-n}}{T(T-n)}$$  \hspace{1cm} (13)
However, in autoregression environment, we only have asymptotic properties, $\hat{\Omega} \rightarrow \Omega$. Therefore, the more appropriate way is to use the approximation:

$$T_{T+1}^2 = (\hat{e}_{T+1} - \bar{e})' \hat{\Omega}^{-1} (\hat{e}_{T+1} - \bar{e}) \sim \chi^2(n)$$

(14)

For the following observations, the forecasting errors are acquired recursively one-step-ahead each. Generally, for the forecasting error of the $i^{th}$ observation, $\hat{e}_i = \hat{\Pi} \bar{x}_i - \bar{y}_i$, the chart statistic is

$$T_i^2 = (\hat{e}_i - \bar{e})' \hat{\Omega}^{-1} (\hat{e}_i - \bar{e}) \sim \chi^2(n).$$

(15)

In fact, $F$ distribution is asymptotically close to be chi-square distribution. Under the same confidence level, the asymptotic distributions of the estimates provide more narrow confident intervals for true parameters than the exact distribution for limited sample size $T$. Therefore, control charts using chi-square distribution are more sensitive than those using $F$ distribution.

We may substitute the estimation of ARMA models by the estimation of AR models. This substitution is more appropriate in the multivariate autocorrelated processes. To see this, consider the underlying principles of processes with ARIMA model. It is rational a real process is autoregressive, for example, many processes are continuous flow processes and the status of continuous flow must be determined by the its previous status. On the other hand, the moving average terms is composed of disturbances. For a univariate model, the disturbances can be regarded as a collection of the effects of other immeasurable or unmeasured influential factors. These factors can be other process variables and environmental variables. Some of them are immeasurable, others are measurable but not measured in the simple univariate model. If these disturbance factors are autocorrelated themselves, then it is reasonable to model the disturbances as moving average.

Another source of the disturbances is measurement errors. This disturbance is not from the process system and its environment, but from the measurement instruments. The systematic
pattern should not include measurement errors, because the measurement disturbance would not be induced if the quality inspection were not applied. Thus, if the true disturbance factors are not autocorrelated, the systematic pattern is not moving average. In this case the presence of moving average in the data is due to the measurement and it should be deleted as much as possible by improving the measurements.

In multivariate processes, more measurable variables are included into the model. These measurable variables might have effects in the disturbance term if they were not included in the model explicitly--such as in univariate case. Thus, the effects of the disturbance factors in the multivariate model should be less than that in the univariate case. In other words, some of the potentially autocorrelated parts of the disturbances, which would have been presented as moving average (MA) in the univariate model, are moved, or transferred, to the autoregressive (AR) part in the multivariate model.

Further, more, even if the remaining moving average terms in the disturbance (say, due to measurement) can not be deleted through including more systematic variables explicitly in the model, they should tend to have their own autocorrelation patterns. For instance, measurement of temperature is usually independent to the measurement of pressure, because the principles and equipment of the measurement (not the temperature and the pressure themselves) are different. These independent patterns of measurement errors jointly form the disturbance so that the joint effect on the disturbance should be less autocorrelated. That is the moving average terms will be weakened in multivariate model, although deleting autocorrelation in measurement errors is always preferable. In summery, the systematic moving average plays less important role in multivariate case than in univariate case, or say, the importance of the autoregressive terms in the multivariate case is enhanced in comparison with that in the univariate case.

The above discussion can also be seen from the mathematical examinations in the multivariate ARMA expressions. First, if the 'true' underlying process is a vector AR(p), then the
univariate model for individual series \( y_{jt} \) will be an ARMA(np,(n-1)p), where n is the number of variables in the vector system. That is, when we try to use a univariate ARIMA model to describe the individual series \( y_{jt} \) which is actually generated from an n variable multivariate autoregressive process of order p, we will get an ARMA model with the maximum autoregressive order of np, and maximum moving average order of (n-1)p. This result was given in Reinsel (1993). For example, a vector AR(1) model \((I - \Phi L)\vec{y}_t = \epsilon_t\) also means

\[
\text{det}(I - \Phi L)\vec{y}_t = (I - \Phi_1^*L - \cdots - \Phi_{n-1}^*L^{n-1})\epsilon_t \quad (16)
\]

where \(Adj(I - \Phi_1^*L - \cdots - \Phi_{n-1}^*L^{n-1}) = Adj(I - \Phi L) = \text{det}(I - \Phi L)(I - \Phi L)^{-1}\). Hence the \(i^{th}\) individual series follows \(\text{det}(I - \Phi L)y_{jt} = \eta_j(L)a_{jt}\), which is an ARMA(n, (n-1)) model--the determination of \(I - \Phi L\) is polynomials of degree n, and the adjunct matrix of \(I - \Phi L\) is of lag order n-1. Generally, a multivariate AR(p) gives univariate individual series with ARMA(np,(n-1)p).

On the other hand, if a individual series, say, \(y_{it}\), follows an ARMA(p,q) process,

\[
(1 - \phi_1 L - \cdots - \phi_p L^p)y_{it} = (1 + \eta_1 L + \cdots + \eta_q L^q)a_{it}, \quad (17)
\]

it is at least possible to form \(y_{2t} = (\eta_1 L + \cdots + \eta_q L^q)a_{it}\), which is independent with \(a_{it}\) at time t--since \(a_{it}\)'s are independent along time--so that

\[
(1 - \phi_1 L - \cdots - \phi_p L^p)y_{1t} - y_{2t} = a_{it}. \quad (18)
\]

Moreover since \(y_{2t}\) is a combination of q-1 innovations, of course it can be expressed as one innovation plus some others. Hence repeating the steps as from (17) to (18), we can get q+1 equations with zero order moving average. Note here \(a_{it}\)'s at t, t-1, ..., t-q, are just i.i.d. random realizations, we can just denote, say, \(\epsilon_{jt} = a_{it-j}\), as another innovation at time t that is different from \(a_{it}\). So \(i=0,1,\ldots,q\), and \(j=1,2,\ldots,q+1\). Although we may not be able to identify the physical
or realistic meanings of the variables $y_{2t}, \ldots, y_{q+1t}$, at least they form a vector AR(p) process of q+1 variables together with the real $y_{1t}$.

In real problems, it only makes sense when some real variables are identified in the system. We can find $\tilde{\epsilon}_t$, another set of q+1 serially i.i.d. innovations so that $\tilde{\epsilon}_t = D\tilde{\epsilon}_t$. With $\tilde{\epsilon}_t$, it is possible that a new set of $y_{2t}, \ldots, y_{q+1t}$ can be found that these variables have real meanings. For instance, $y_{q+1t} = D^{-1}\tilde{\epsilon}_t$ may have real meaning. If fortunately we found all of the q+1 real variables that are occasionally in the form of these $y_{1t}, \ldots, y_{q+1t}$, we actually expressed the individual univariate AMRA models with a vector autoregressive model.

We may not be so fortunate in identifying model variables correctly. It is quite possible that some explanatory variables are even unknown and not observed at all. Therefore, variables included in the model may not describe the true system appropriately. However, there is still the possibility of identifying all appropriate variables so that a VAR model represents the system and this enhances and justifies VAR modeling.

Estimating a multivariate AR(p) model instead of an multivariate ARMA model is more reasonable than that in the separate univariate models. Although it is impossible to know the true model exactly, and separate univariate AR models may not be very close to the "true" processes, the multivariate AR(p) model is more possibly close to the proxy of the underlying. In an ideal case, all of the systematic variables may be included in the monitoring vector, and only the white noise and measurement errors are in the disturbance. By these reasons, in summetry we use the multivariate autorrgression model to filer the systematic causes from the process. Estimating an AR model is also much easier than estimating an MA model. As the mater of fact, estimation of multivariate moving average models is even unavailable in many software packages. Our VAR chart can avoid this problem.
However, we have to admit that not all processes are appropriate for VAR modeling. In some cases, e.g., large value coefficients in moving average terms, fitting the VAR(p) have to lag to high order \( p \). This may indicate the "true" underlying process is mixed vector ARMA. Using high order VAR to approximate low order ARMA needs to estimate too many parameters. This is not appropriate especially when the data size is limited.

**THE DESIGN OF VAR CHARTS**

One VAR statistic is given in (15). By selecting the expression of the mean item of \( T_i^2 \), we can have more chart schemes. One option is to directly choose \( \overline{\hat{e}} \) as its asymptotic value, zero.

Then the item \( \overline{\hat{e}} \) in (15) disappears and the chart statistic is on the target zero:

\[
T_i^2 = \hat{\Sigma}_{i-1}^{-1} \hat{\Sigma}_{i} \sim \chi^2(n). 
\]

(19)

The corresponding Phase I statistic is (10).

Another option is to use recursive mean. This is not exactly a two-phase scheme. The mean for the \( i+1^{th} \) observation’s statistic can be defined as \( \overline{\hat{e}}_i = (\hat{e}_i + (i-1)\hat{e}_{i-1}) / i \) so that the statistic is

\[
T_i^2 = (\hat{e}_i - \overline{\hat{e}}_i)^T \hat{\Sigma}_{i}^{-1} (\hat{e}_i - \overline{\hat{e}}_i) \sim \chi^2(n) 
\]

(20)

Although the recursive mean varies every time when a new observation is added, the control limit of the chart does not change. This is much better in vision than the moving central line in the univariate case (see Montgomery and Mastrangelo 1991). In fact, all the above schemes have the same chart that the UCL is determined by \( \chi^2(n) \), say, \( \chi_{0.01}^2(n) \) or \( \chi_{0.0027}^2(n) \). Only the statistic values are different for different schemes. Hence, the VAR chart is actually simple in its vision. In addition, its model estimation is actually not complicated as well. Among the above, the statistic in (19) is the most convenient and appropriate, because the asymptotic mean of \( \overline{\hat{e}} \) is zero so (19) is theoretically justified for large sample size.
**By Using Likelihood Ratio Test,** we may determining the autoregressive order \( p \) may be the issue before chart setting. To find an appropriate VAR order, the following likelihood ratio test can be induced.

\[
H_0: \text{the VAR order is } p_0 \\
H_a: \text{the VAR order is } p_1 > p_0
\]

The test statistic is a likelihood ratio:

\[
2(L_1 - L_0) = T \ln \left( \frac{\hat{\Omega}_0}{\hat{\Omega}_1} \right) \sim \chi^2 (n^2 (p_1 - p_0)) \tag{21}
\]

where \( T \) is the number of observations used in model estimation. \( \hat{\Omega}_0 \) is the estimated covariance matrix of the error term given in (12) under the null hypothesis being true. \( \hat{\Omega}_1 \) is also from (12) but under the alternative hypothesis. The chi-square distribution has a degree of freedom \( n^2 (p_1 - p_0) \) because each variable has \( (p_1 - p_0) \) fewer lags under the null hypothesis and each of the \( n \) equations includes \( n \) variables so that \( n^2 \) restrictions for each lag were added. We reject the null hypothesis for large value of \( 2(L_1 - L_0) \). A workable way to determine the order \( p \) is to test \( p_1 = p_0 + 1 \) and let \( p_0 \) increase gradually until the null hypothesis cannot be rejected.

The Akaike information criteria can be also used in multivariate cases. Specifically for VAR model, the order of \( p \) is determined by minimizing one of the following criteria that are, respectively, FPE (Akaike, 1971):

\[
FPE_p = \left[ (1 + pn / T)(T / (T - pn)) \right] \hat{\Omega}_p \tag{22a}
\]

or AIC (Akaike, 1976):

\[
AIC_p = \ln \left( \hat{\Omega}_p \right) + 2pn^2 / (T - p) \tag{22b}
\]

or BIC (Schwarz, 1978):

\[
BIC_p = \ln \left( \hat{\Omega}_p \right) + pn^2 \ln(T) / T \tag{22c}
\]

or HQ (Hannan and Quin, 1979):
\[ H_{\Omega_p} = \ln(\left| \hat{\Omega}_p \right|) + 2pn^2 \ln(\ln T)/T \]  

(22d)

where \( \hat{\Omega}_p \) is the residual covariance matrix for fitting VAR(p) model. Detailed discussions of this can be found in Reinsel (1993).

As mentioned previously, although VAR scheme can cover a great part of the real processes, it is possible in some cases a low order of VAR(p) cannot be acquired. If the order \( p \) has to be quite large, it usually indicates the possibility of parsimonious moving average terms. Then charts especially for VMA(q) or mixed VARMA may be more appropriate than VAR(p).

There are many variables in reality available that may be observed. A group of variables may be cross-correlated and serially correlated, others may be irrelevant to that group. The multivariate autocorrelated processes can be modeled effectively for the group of variables that have internal interactions. Nevertheless, it is not effective to describe irrelevant variables in a multivariate model. Before modeling, the relationship among all the available variables is usually unknown. Therefore, identify appropriate variable groups for the multivariate process is an issue.

We know filtering systematic patterns from the processes can monitor the changes due to special causes efficiently. However, the variables in the systematic patterns should be relevant. If some of the variables do not have influence to the said systematic patterns, they are irrelevant and do not help filtering the systematic patterns. Since data size in Phase I is limited, estimating coefficients for those irrelevant variables will reduce efficiency of estimating the systematic patterns. Therefore, the irrelevant variables should not be included in the VAR chart.

Suppose there are \( n \) variables in a system. A group of \( n_1 \) variables, Group 1, is of main interests, while the remaining \( n_2 \) variables, Group 2, though of other interests, may not have influence on Group 1. \( n=n_1+n_2 \). For the current main interests, monitoring Group 1 for special causes is wanted, so we do not want include the non-influential Group 2 in the chart in order to
optimize the efficiency of the chart. Suppose the two groups, \( \tilde{y}_{1t} \), an \((n_1 \times 1)\) vector for Group 1 and \( \tilde{y}_{2t} \), an \((n_2 \times 1)\) vector for Group 2, follows vector autoregressive model of order p,

\[
\tilde{y}_{1t} = \tilde{c}_1 + A_1 \tilde{x}_{1t} + A_2 \tilde{x}_{2t} + \tilde{e}_{1t}
\]

\[
\tilde{y}_{2t} = \tilde{c}_2 + B_1 \tilde{x}_{1t} + B_2 \tilde{x}_{2t} + \tilde{e}_{2t}
\]

where \( \tilde{x}_{1t}' = (\tilde{y}_{1,t-1}', \tilde{y}_{1,t-2}', \ldots, \tilde{y}_{1,t-p}') \) and \( \tilde{x}_{2t}' = (\tilde{y}_{2,t-1}', \tilde{y}_{2,t-2}', \ldots, \tilde{y}_{2,t-p}') \), \( A_1, A_2, B_1, \) and \( B_2 \) are matrixes of coefficients, the constant items and the error term items are all vectors for the respective variable groups.

If Group 1 does not have influential effects on Group 2, then \( \tilde{y}_{1t} \) should not be dependent on the historical values of \( \tilde{y}_{2t} \). In this case, the matrix \( A_2 \) should be zero, it is called there is no Granger causality from \( \tilde{y}_{2t} \) to \( \tilde{y}_{1t} \). Specifically, to test the Granger causality from Group 2 to Group 1, the hypothesis is

\[
H_0: \quad A_2 = 0
\]

\[
H_a: \quad A_2 \neq 0
\]

The likelihood ratio statistic is

\[
2(L_1 - L_0) = T\left(\log|\hat{\Omega}_{11}(0)| - \log|\hat{\Omega}_{11}|\right) \sim \chi^2 (pn_1 n_2)
\]

where \( \hat{\Omega}_{11} \) is the estimate of the covariance matrix of \( \tilde{e}_{1t} \) without restrictions—which is the covariance matrix of the residuals of Group 1; and \( \hat{\Omega}_{11}(0) \), also covariance of residuals, is the estimate of the covariance matrix of \( \tilde{e}_{1t} \) under restriction of the null hypothesis \( A_2 = 0 \); and the degree of freedom is the number of restrictions since \( A_2 \) is an \( n_1 \times n_2 p \) matrix.

As we are mainly concerned with the systematic causes in Group 1, we are not interested in the Granger causality from \( \tilde{y}_{1t} \) to \( \tilde{y}_{2t} \). Nevertheless, if we want, the Granger causalities in both directions can be tested based on the estimated covariance matrix of each group with or
without the restrictions of the null hypothesis. To test the Granger causality from Group 1 to Group 2, the null hypothesis is $B_1 = 0$ where the likelihood ratio test is based on the residual covariance $\hat{\Omega}_{22}$ and $\hat{\Omega}_{22}(0)$.

Besides, of Granger causality that is due to historical influence, there are also instantaneous feedbacks between Group 1 and Group 2. As he, total covariance matrix of the system is $\Omega = \begin{pmatrix} \Omega_{11} & \Omega_{12} \\ \Omega_{21} & \Omega_{22} \end{pmatrix} = \begin{pmatrix} E(\bar{e}_{1t}\bar{e}_{1t}') & E(\bar{e}_{1t}\bar{e}_{2t}') \\ E(\bar{e}_{2t}\bar{e}_{1t}') & E(\bar{e}_{2t}\bar{e}_{2t}') \end{pmatrix}$, the instantaneous feedback between the two groups is due to $\Omega_{12} = \Omega_{21} \neq 0$. A joint test for $A_2 = 0, B_1 = 0$, and $\Omega_{12} = 0$ is given by

$$2(L_1 - L_0) = T\left\{ \log \hat{\Omega}_{11}(0) + \log \hat{\Omega}_{22}(0) - \log \begin{vmatrix} \hat{\Omega}_{11} \\ \hat{\Omega}_{21} \\ \hat{\Omega}_{22} \end{vmatrix} \right\} \sim \chi^2(n_1n_2(2p + 1)) \quad (26)$$

Hamilton (1994) gave detailed discussions on the vector autoregression and the above tests. Following the same idea, we can go further for only one directional feedback to our main interest in Group 1. If only the hypothesis $\Omega_{12} = \Omega_{21} \neq 0$ is wanted to test, an alternative VAR equation is needed,

$$\tilde{y}_{2t} = \tilde{d} + D_0 \tilde{y}_{1t} + D_1 \tilde{x}_{1t} + D_2 \tilde{x}_{2t} + \tilde{v}_{2t}. \quad (24b)$$

where $D_0 = \Omega_{21} \Omega_{11}^{-1}$, $\tilde{d} = \bar{e}_2 - \Omega_{21} \bar{e}_1$, $D_1 = B_1 - \Omega_{21} \Omega_{11}^{-1} A_2$, and $D_2 = B_2 - \Omega_{21} \Omega_{11}^{-1} A_2$. In this way, $\tilde{v}_{2t}$ is uncorrelated with $\bar{e}_{2t}$. Therefore, to test instantaneous feedback between groups is to test the $n_1 \times n_2$ matrix $D_0 = 0$. And the likelihood ratio is acquired from the residual covariance with and without restriction of $D_0 = 0$ in (24b),

$$2(L_1 - L_0) = T\left\{ \log \left| \hat{\Omega}_{22}(D_0 = 0) \right| - \log \left| \hat{\Omega}_{22}(D_0 \neq 0) \right| \right\} \sim \chi^2(n_1n_2) \quad (27)$$

Note the residual covariance with restriction of $D_0 = 0$ in (3-2-7c) is just the residual covariance without restriction of $B_1 = 0$ in (3-2-7b). If the joint test of $A_2 = 0$ and $\Omega_{12} = 0$ is wanted, in which whether $B_1 = 0$ is not concerned, the likelihood ratio becomes
\begin{equation}
2(L_1 - L_0) = T\{\log|\hat{\Omega}_{11}(0)| + \log|\hat{\Omega}_{22}(D_0 = 0)| - \log|\hat{\Omega}_{11}| - \log|\hat{\Omega}_{22}(D_0 \neq 0)|\}
\end{equation}

which follows $\chi^2(n_1n_2(\text{p}+1))$.

With the above tests about VAR and the appropriate variable group of interest, the VAR of main interest for monitoring can be defined. At the beginning, some trial-error checks are necessary when suitable order and groups are unknown. Experience and physical or economic principle help in the selection. If the true systematic pattern is really generated by variables of Group 1 that are block-exogenous with respect to the other variables, the above test should give desired results.

As the properties of the estimates of the parameters are asymptotic only, and in reality, the sample size for estimating the parameters is limited, the VAR residuals are not i.i.d. normal. Note the residuals (Figure 1) are asymptotically multivariate normal with mean of zero,

\begin{equation}
\hat{\epsilon}_t = \tilde{y}_t - \hat{\tilde{y}}_t = \Pi'\tilde{x}_t + \hat{\epsilon}_t - \hat{\Pi}'\tilde{x}_t = \hat{\epsilon}_t + (\Pi' - \hat{\Pi}')\tilde{x}_t = \hat{\epsilon}_t + \hat{O}(1/\sqrt{T}).
\end{equation}

The autocovariance matrix of the residual for lag $l$ is found as

\begin{equation}
\hat{\Gamma}_e(l) = E(\hat{\epsilon}_t, \hat{\epsilon}_{t-l}') = E(\tilde{y}_t - \hat{\tilde{y}}_t, (\tilde{y}_{t-l} - \hat{\tilde{y}}_{t-l}))
= E(\hat{\epsilon}_t, \hat{\epsilon}_{t-l}') - E[\epsilon_t, ((\Pi' - \hat{\Pi}')\tilde{x}_{t-l}')] - E[((\Pi' - \hat{\Pi}')\tilde{x}_t, \hat{\epsilon}_{t-l}')] + E[((\Pi' - \hat{\Pi}')\tilde{x}_t, \hat{\epsilon}_{t-l}')(\Pi - \hat{\Pi})]
\end{equation}

which is also asymptotically zero for $l \neq 0$. Hence, with large samples the residuals are approximately uncorrelated along time. Baillie (1979) and Reinsel (1980)'s result for VAR(p) one-step-ahead forecasting error covariance with estimated parameters is

\begin{equation}
\hat{\Omega} = (1 + pn/T)\Omega
\end{equation}

Therefore, after the processes are estimated it is more appropriate and reliable to test if the residuals are vector white noises or more restrictively for VAR chart, i.i.d. normal sequence. If the residuals are not white noises, it implies the estimated VAR(p) model does not fit the processes well. This can be done through testing each of the elements of the autocorrelation matrix for comparing with $1/\sqrt{T}$. That is
\[
\hat{\rho}_e(l) = V_e^{-1/2} \hat{\Gamma}_e(l) V_e^{-1/2}. \text{ and } \text{var}(\hat{\rho}_v(l)) \sim 1/\sqrt{T}
\] (32)

for \(l \neq 0, \) and \(i, j = 1, \ldots, n, \) where \(\hat{\rho}_v(l)\) is the elements of cross-correlation matrix \(\hat{\rho}_v(l)\)

and \(V_e = \text{diag}(\hat{\sigma}_{11}, \ldots, \hat{\sigma}_{nn}).\) In forecasting practice, we employ two standard error limits, 
\(\pm 2/\sqrt{T}.\)

However, this way is not efficient since there are totally \(n(n+1)/2\) of these elements to be 
tested, and each for sufficient number of lags. A better way to achieve the white noise test is to 
check the over all model adequacy. The multivariate portmanteau tests have been studied by 
Hosking (1980) and Poskitt and Tremane (1982), etc. An overall "goodness-of-fit" statistic 
(Reinsel, 1993) can be

\[
Q_s = T^2 \sum_{l=1}^{s} (T - l)^{-1} \text{tr}[\hat{\Gamma}_e(l)\hat{\sigma}_e^{-1}\hat{\Gamma}_e(-l)\hat{\sigma}_e^{-1}]
\]

\[
= T^2 \sum_{l=1}^{s} (T - l)^{-1} \text{tr}[\hat{\rho}_v(l)\hat{\rho}_v(0)^{-1}\hat{\Gamma}_e(-l)\hat{\rho}_v(0)^{-1}]
\] (33)

which follows approximately chi-square of \(n^2(s - p - q)\) degrees of freedom.

If we confirm the vector white noises of the residuals, with the assumption of normality 
of the error terms, we treat the residuals as i.i.d., a normal sequence. We construct the VAR chart 
by charting the ordinary Hotelling \(T^2\) chart of the vector residuals.

The Upper Control Limit (UCL) is determined through the consideration of average run 
length of in-control and out-of-control processes. Referring to simple Shewhart chart with known 
parameters, the 3-sigma limits correspond to a Type I error of 0.0027 or an average run length of 
370 for in-control process. However, in our VAR chart the statistics for different observations are 
dependent. Since we not only base them on the same estimated \(\hat{\sigma}\) but also the processes are 
autocorrelated and the estimated error terms are only asymptotically i.i.d. sequence. Hence the 
average run length is different from 370 if the Type I error \(\alpha\) is set at 0.0027. As the practical
processes differ in the parameters, the choices of UCL will be also different for concordance of ARL. The actual setting of UCL may depend on the experience of practical run trials. For instance, by presetting UCL at $\chi^2_{0.005}(n)$ we test the practical average run length (ARL). Then we adjust the in-control and out-of-control average run lengths. For estimated process parameters, we acquire the ARL and UCL by experimental simulations and examination. Preset Type I error does not determine average run length. Later, we will simulate an example process and show the coordination between UCL and ARL for that specific process. In fact, all the autocorrelated processes, no matter whether univariate or multivariate, the ideas of control limit setting are in the same way. The estimation of coefficients produces average run lengths for the specific estimated parameters based on the observations of the process. Hence, the way to get a chart’s ARL is not general but related to certain processes.

As just mentioned, $\chi^2(n)$ is just the asymptotic distribution of the Hotelling T^2 statistic in the VAR model, it has narrower limits than the corresponding $F$ distribution. The inaccuracy of estimation may make the average run length more skewed (Quesenberry, 1993). Therefore, even Type I error at 0.0027 may be still too small for acquiring an in-control ARL of 370. The in-control ARL of the VAR chart consistent to that of traditional 3-sigma Shewhart chart may correspond to an even smaller significance level for the chi-square distribution than 0.0027, (for example $\alpha=0.002$).

Finally, we point out that although we emphasized a lot the reduction from multivariate ARMA model to multivariate AR model, it is not a necessary condition for constructing VAR chart. If computation of multivariate maximum likelihood estimation is available and reliable, the MLE estimation acquires for us the residuals or one-step-ahead forecasting errors, $\hat{e}_t = \hat{y}_t - \bar{y}_t$. The multivariate residual chart--in this case, it may be called VARMA chart--can then be constructed the same way as the above VAR chart. The main advantage of VAR chart with the recommended multivariate AR model is the easiness of implementation. Ordinary least squares
and matrix operation are available even in usual spreadsheet software so that VAR chart can be popularized for practitioners. On contrary, as we know, maximum likelihood estimation for moving average is a non-linear optimization approach where local optimal points may appear especially for multivariate time series. Reliable results are sometimes subject to constraint of computational difficulties.

THE EFFECTS OF SHIFTS FOR VAR CHART

There are three classes of parameters in VAR model: the process mean \( \mu \), the covariance matrix of error term, \( \Omega \), and the autoregression structure of the model, \( \Phi \). All these three kinds of parameters can be subject to change due to the assignable causes in the processes. We discuss the effects of these shifts separately.

For a mean shift, the systematic change only happens on mean \( \hat{\mu} \) (or \( \bar{c} \)), not on other parameters such as \( \Omega \) and \( \Phi \). The relationship between the mean of a stationary time series vector, \( \bar{\mu} \), and the constant term vector, \( \bar{c} \), is, taking expectation of equation (3-2-3),

\[
\mu = \bar{c} + \Phi(L)\hat{\mu} \quad \text{or} \quad \bar{c} = (I_n - \Phi(L))\hat{\mu}.
\] (34)

If a shift of \( \delta \) on the system mean \( \hat{\mu} \) occurs, the time series vector is still stationary and the constant term in the model (24) should finally change to be

\[
(I_n - \Phi(L))(\hat{\mu} + \delta) = \bar{c} + (I_n - \Phi(L))\delta = \bar{c} + \bar{\eta}
\] (35)

The effects of mean shift in the \( T^2 \) statistic of the VAR chart are given in the following lemma.

**Lemma 1.** For a process model in (24) and VAR chart in (15), a shift \( \delta \) in the process mean makes the Hotelling \( T^2 \) a noncentral chi-square distribution. That is, it adds a normal distribution with mean of \( \hat{\eta}_k^T \hat{\Omega}^{-1}\hat{\eta}_k \) and variance of \( 4\hat{\eta}_k^T \hat{\Omega}^{-1}\hat{\eta}_k \) on the in-control \( T^2 \). The effects depend on the process shift \( \delta \) and the estimates of VAR parameter matrices, \( \hat{\Phi} \) and \( \hat{\Omega} \).
The proof of this lemma is in Appendix II. As a result, the effect on VAR chart statistic is, for \( t \geq t_0 \),

\[
T^2_{t} = \text{shift} T^2_{t} = -2\hat{\eta}_k' \hat{\Omega}^{-1} \hat{e}_t + \hat{\eta}_k' \hat{\Omega}^{-1} \hat{\eta}_k
\]

(36)

where \( k = t - t_0 = 0, 1, 2, \ldots \), and \( \hat{\eta}_k = (I_n - \Phi_1 - \Phi_2 - \cdots - \Phi_k)\hat{\delta} \). For \( t > t_0 + p \), \( \hat{\eta}_k = \hat{\eta} = \hat{\eta}_p \). The first term in the right hand side of (3-2-10) is the in-control statistic \( \text{no shift} T^2_{t} \sim \chi^2(n) \) with mean of \( n \) and variance of \( 2n \). The second term \( -2\hat{\eta}_k' \hat{\Omega}^{-1} \hat{e}_i \) is a linear combination of normal distributions \( \hat{e}_i \sim N(0, \Omega) \) so that it is normal with mean of 0 and variance of \( 4\hat{\eta}_k' \hat{\Omega}^{-1} \hat{\eta}_k \). Its effect on the first term is approximately to increase the variation but without moving the mean. The third term is a positive definite scalar so it adds a positive shift on the mean of the second term. Therefore, the total effect of the process shift \( \hat{\eta}_k \) is to add a normal distribution with mean of \( \hat{\eta}_k' \hat{\Omega}^{-1} \hat{\eta}_k \) and variance of \( 4\hat{\eta}_k' \hat{\Omega}^{-1} \hat{\eta}_k \) on the in-control \( T^2 \). We measure the effects with \( \hat{\eta}_k' \hat{\Omega}^{-1} \hat{\eta}_k \). When the in-control chi-square is with high degree of freedom, the effect is approximately a positive shift on the mean of the \( T^2 \) statistic and a positive shift of on the variance of the \( T^2 \) statistic. The VAR will signal the shift since the statistic violates the in-control chi-square distribution. The out-of-control statistic will be at its steady-state distribution after \( p \) periods from the shift occurrence, while it changes for \( 0 \leq k = t - t_0 \leq p \).

Therefore, the shift vector measured in terms of the covariance matrix is the non-centrality parameter that determines the effect of the shift on the VAR chart. \( \Delta_0 = -\delta' \Omega^{-1} \delta \) can be viewed as the initial effect and \( \Delta = \delta' \Omega^{-1} \delta \) can be viewed as the steady state effect. The transient effect can be measured as \( \Delta_k = \hat{\eta}_k' \Omega^{-1} \hat{\eta}_k, \ 0 \leq k = t - t_0 \leq p \). Since the change in sign of
the mean shift results with the same non-central parameter, the direction of the mean shift does not influence the magnitude of the effect.

We note that the variable mean shift $\delta_r$ consists of both systematic causes and special causes. On the contrary, the error term represents the special cause only. We view, however, that its in-control covariance matrix as part of the parameters determining the normal status of the system. Hence, we can classify it as a representative of the systematic causes. Therefore, in the sense of detecting special causes, one considers the shift in the mean of the error term. In this case, for a step mean shift $\bar{\eta}_r$ in the error term, which is also in the constant term of the VAR model, the effect on VAR chart does not have transient pattern.

Other systematic shifts, such as the shift on the covariance matrix of error term, $\Omega$, or the shift on the autocorrelation structure of the model, $\Phi$, will also violate the distributions in (15), hence out of control signals will be observed after the shifts. Like in the mean shift, the amount of the effect of a shift depends on the scale of the shift and the Phase I estimated covariance matrix $\hat{\Omega}$. The larger the shift scale, the more serious is the violation of the chi-square distribution.

**Lemma 2.** If the process covariance $\Omega$ shifts to $\Omega_1$, then $\Omega_1 = D \Omega D'$, and the effect on VAR chart is $T^2_{shift} = \hat{e}_t^T D' \hat{\Omega}^{-1} D \hat{e}_t$.

That is, the out-of-control statistic is a Hotelling that we construct from amplified residuals $\hat{\hat{e}}_t = D \hat{e}_t$, and the original covariance estimate $\hat{\Omega}$. Consequently, if the shift is only in the diagonal elements of the covariance, an increase in the variance will result to an increase in Hotelling $T^2$. The proof of this lemma is in Appendix III. This result tells us the out-of-control statistic encounters a step shift itself immediately after the occurrence of the shift in the process variance. The sequence of the Hotelling $T^2$ is in dependent without time series pattern except the step shift.
We give the effect of coefficient shifts in the following lemma. The proof of this lemma is in Appendix III.

**Lemma 3.** For of a VAP(p) process, the effect on the Hotelling $T^2$ of coefficient shift from $\Phi(L) = \Phi_0(L) + \Delta \Phi(L)$ to $\Phi_0(L)$, is

$$T^2_{shift} = T^2_{noshift} + 2(\Delta \Phi \tilde{x})' \Omega^{-1} \tilde{e} + (\Delta \Phi \tilde{x})' \Omega^{-1} \Delta \Phi \tilde{x}, \quad (37a)$$

and it is also

$$T^2_{shift} = [\Phi(L)^{-1} \Phi_0(L) \tilde{e}]' \Omega^{-1} [\Phi(L)^{-1} \Phi_0(L) \tilde{e}], \quad (37b)$$

With this result, since the fundamental innovations at different time points are independent each other, the above quadratic form statistic in (37) is composed of two types of distributions. One is $\tilde{e}_{t-j}' \Omega^{-1} \tilde{e}_{t-j}$, which is a chi-square with degree of freedom of $n$. The other is $\tilde{e}_{t-j}' \Omega^{-1} \tilde{e}_{t-j}, j \neq i$, a symmetric distribution with mean of zero, variance of $n$, and decreasing kurtosis for $n$. If we call $n$ as the degree of freedom, the largest kurtosis is 9, corresponding to the univariate case when $n=1$. When $n$ is large, the distribution converges to normal according to central limit theorem and kurtosis reduces to 3. The Simulation methods will verify the results on this distribution.

The control limit and the scale of the shift determine the out-of-control ARL, which determines how soon the VAR chart can, on average, detect the shift. In particular, certain shifts in the coefficients give time variant ARL due to the property of the out-of-control statistic.

**ILLUSTRATIONS OF VAR CONTROL CHARTS**

Weekly production schedule billing figures of a company for $T=100$ weeks was previously studied by authors (Makridakis and Wheelright, 1978; Reinsel, 1993) to build
forecasting models. From the context of the data, we anticipate that the production schedule figures ($Y_1$) influence or drive the billing figures ($Y_2$) so that $Y_2$ should have feedback from $Y_1$. Using model building techniques with transfer function model (Box and Jenkins, 1976) it has been found in literature that $Y_1$ is an MA(1) and it transfers to $Y_2$ in a complicated function (see Reinsel, 1993):

$$y_{2t} = 71.764 + \frac{2.106L^2 - 1.965L^4}{1-1.541L + 0.793L^2} y_{1t} + \frac{1-0.544L^2}{1-1.423L + 0.763L^2} e_{2t}$$

Reinsel (1993) showed how to model these two processes as vector AR(3) or preferably, vector ARMA(2,1). Choosing VAR(3) is for smallest AIC, and individual residual correlation values are within $\pm 2/\sqrt{T}$, although likelihood ratio test indicating $p=4$. However, Reinsel (1993) did not report portmanteau results. Actually, with our computing program, VAR(3) residual give significant portmanteau testing results for all lags. VAR(4) is actually the more appropriate one, though less parsimonious. The results are as Part (a) of Table 3.

<table>
<thead>
<tr>
<th>Table 3 Production Schedule and Billing Figures</th>
</tr>
</thead>
<tbody>
<tr>
<td>(a) (T=100), (\Rightarrow) Accept VAR(4)</td>
</tr>
<tr>
<td>(H_0: p_0=3); vs. (H_1: p_1=4); (p = 3\ 4)</td>
</tr>
<tr>
<td>LR = 48.9646 &gt; (\chi^2_{0.05}(9) = 9.4877); AIC = 2.9564 2.5361</td>
</tr>
<tr>
<td>(sequentially, (H_0: p_0=4) was confirmed) BIC = 3.2832 2.9747</td>
</tr>
<tr>
<td>(s): 5 6 7 8 9 10 12</td>
</tr>
<tr>
<td>(Q(s):) 14.8107 17.1218 19.4239 23.3610 30.4240 33.2375 39.8869</td>
</tr>
<tr>
<td>(\chi^2_{0.05}(s):) 9.4877 15.0737 21.0261 26.2962 31.4104 36.4150 46.1943</td>
</tr>
<tr>
<td>d.f.: 4.0000 8.0000 12.0000 16.0000 20.0000 24.0000 32.0000</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>(b) (T=70), Phase I (\Rightarrow) Accept VAR(4)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(H_0: p_0=3); vs. (H_1: p_1=4); (p = 3\ 4)</td>
</tr>
<tr>
<td>LR = 24.6771 &gt; (\chi^2_{0.05}(9) = 9.4877); AIC = 2.7061 2.4674</td>
</tr>
<tr>
<td>(sequentially, (H_0: p_0=4) was confirmed) BIC = 3.1143 3.0163</td>
</tr>
<tr>
<td>(s): 5 6 7 8 9 10 12</td>
</tr>
<tr>
<td>(Q(s):) 7.1584 10.8711 13.1005 15.1571 22.0304 28.6714 41.6591</td>
</tr>
<tr>
<td>(\chi^2_{0.05}(s):) 9.4877 15.0737 21.0261 26.2962 31.4104 36.4150 46.1943</td>
</tr>
</tbody>
</table>
To show VAR chart application with this process, we use the first 70 observations in Phase I. The model from this Phase I data is checked in Part (b) of Table 1 where VAR(4) is well justified. The Phase I chart on VAR residuals was checked at control limit of $\alpha=0.01$. In the Phase II chart two control limits were used, respectively at $\alpha=0.01$ and $\alpha=0.0027$. Monitoring the observations from the 71st to the 100th on this chart, Obs. #78 and #97 were found signaling for control limit of $\alpha=0.0027$, while if control limit is set at $\alpha=0.01$, more observations, Obs. #76, #78, #90, #97, and #98, signaled alarms. Comparing Part (c) and (d) of Table 1, the estimates are similar. And the residual chart of Part (d) is also in control, but actually, some of the observations are out of the routine processes. These observations need detailed diagnostics to find out the possible special causes. The recursive one-step-ahead forecasting errors have mean standard deviation of 1.2936 and 1.3293 respectively for the two variables, and the mean absolute deviation of 1.9048 and 1.5516, respectively. Figure 1-2-5 shows the charts in Phase I and II.
In another example, we find plastic production data of five temperature variables at different positions. After differencing for a 5 period season, we identified a VAR(5) model to approximate the process in Phase I. A trial run for the first 250 observations showed some outliers. Cutting off these seemingly out of control part of data, 165 observations were used for Phase I estimation (for estimation reasons, the first 20 observation of the 185 were also cut off to delete the initial effects). Presetting UCL at $\alpha=0.0027$, a shift was found at the 13th, 14th, 24th and 30th observations in Phase II. Detailed study showed the processes might have moving average terms, since the portmanteau tests do always give significant results for many lags. But VAR(5) may give acceptable approximation for filtering out part of the systematic patterns.

Like in the independent multivariate chart, the false signal of the chart does not inform the details about the assignable causes. VAR chart cannot tell which parameter makes the false signal. To find out the real cause of the false signal, other univariate charts can be included to help. Or, decomposition similar to Mason et al. (1995) may be used. The VAR chart should be especially effective when the small changes in all the parameters of the autocorrelated multivariate process make a jointly large effect on the process while the single changes are so small that the univariate chart cannot detect effectively. Therefore, when VAR chart detected a signal, we employ a univariate chart to help detect whether it is a single cause or joint causes of several parameters.

In summery, the VAR chart showed its potential usefulness. Even if the true processes may be ARMA(p,q), VAR chart gives simple yet good approximation to detect the special causes. We note that although OLS is easy to implement, sufficiently accurate estimates still need large sample sizes. We need and desire computer-aided calculations. A potential development in the chart is to combine with adjustment of engineering process control.
APPENDIX A : The Proof on Lemma 1

Lemma 1. For a VAR(p) process, a shift $\delta$ in the process mean vectors makes the Hotelling $T^2$ on VAR residuals a noncentral chi-square distribution. That is, when $t_0 \leq t \leq p$, and $k = t - t_0$, the transient state form is

$$T^2_{\text{shift}} = T^2_{\text{noshift}} - 2\hat{\eta}_k'\hat{\Omega}^{-1}\hat{c}_t + \hat{\eta}_k'\hat{\Omega}^{-1}\hat{\eta}_k,$$

with $\hat{\eta}_k = (I_n - \hat{\Phi}_1 - \hat{\Phi}_2 - \cdots - \hat{\Phi}_k)\delta$. When $t \geq t_0 + p$, the steady state form is

$$T^2_{\text{shift}} = T^2_{\text{noshift}} - 2\hat{\eta}'\hat{\Omega}^{-1}\hat{c}_t + \hat{\eta}'\hat{\Omega}^{-1}\hat{\eta}$$

with $\hat{\eta} = (I_n - \hat{\Phi}_1 - \cdots - \hat{\Phi}_p)\delta$.

Proof:

Since the relationship between the mean of a stationary time series vector, $\mu$, and the constant term vector, $\bar{c}$, is $\bar{c} = (I_n - \Phi_1 - \cdots - \Phi_p)\bar{\mu}$, if a shift of $\delta$ on system mean $\bar{\mu}$ occurs, the constant term in (3) has a shift $(I_n - \Phi_1 - \cdots - \Phi_p)\delta = \bar{\eta}$. The forecasting error without shift occurrence is $\hat{e}_t = \bar{y}_t - \bar{\hat{y}}_t$. After the shift, the process status becomes $\bar{z}_t$ instead of $\bar{y}_t$. A sudden shift at time just before $t_0$ on mean for some of the variables implies

$$\bar{z}_t = \bar{y}_t + \delta, \quad \text{for } t \geq t_0.$$

Since the shift is unknown before it is detected, the one-step-ahead forecasting is still based on the estimated coefficient matrix $\hat{\Pi} = (\hat{c} \hat{\Phi})$. Thus, for a VAR(p) process,

$$\hat{z}_t = \hat{c} + (\hat{\Phi}_1L + \cdots + \hat{\Phi}_pL^p)\bar{z}_t, \quad \text{when } t \geq t_0 + p.$$

Then the forecasting error after the shift is
\[
\tilde{y}_t = \tilde{c} + (\Phi_1 L + \Phi_2 L^2 + \cdots + \Phi_p L^p)\tilde{y}_t + \tilde{e}_t
\]
\[
\hat{e}_t = z_t - \tilde{z}_t = \tilde{z}_t - \tilde{c} - (\tilde{\Phi}_1 L + \cdots + \tilde{\Phi}_p L^p)\tilde{z}_t
\]
\[
= \tilde{y}_t - (\tilde{\Phi}_1 L + \cdots + \tilde{\Phi}_p L^p)\tilde{\delta} - \hat{y}_t + \tilde{\delta} = \hat{e}_t + \tilde{\eta}.
\]

Here the \( \hat{e}_t \) would have been the forecasting error at time \( t \), had the shift not occurred.

With the shifted \( \hat{e}_t = \hat{e}_t - \tilde{\eta} \), the \( T^2 \) statistic is a non central chi-square:
\[
T^2_{shift} = \hat{e}_t' \hat{\Omega}^{-1} \hat{e}_t = (\hat{e}_t - \tilde{\eta})' \hat{\Omega}^{-1} (\hat{e}_t - \tilde{\eta}) = \hat{e}_t' \hat{\Omega}^{-1} \hat{e}_t - 2\tilde{\eta} \hat{\Omega}^{-1} \hat{e}_t + \tilde{\eta} \hat{\Omega}^{-1} \tilde{\eta}
\]
\[
= T^2_{no\ shift} - 2\tilde{\eta} \hat{\Omega}^{-1} \hat{e}_t + \tilde{\eta} \hat{\Omega}^{-1} \tilde{\eta}.
\]

When \( t_0 \leq t < t_0 + p \), let \( k = t - t_0 \). We have \( z_t = \tilde{y}_t + \tilde{\delta} \) for \( t_0 \leq t < k \) and we shifted \( k \) lags while \( p-k \) lags we did not shift. The one-step-ahead prediction is
\[
\hat{z}_t = \hat{c} + \hat{\Phi}_1 \hat{z}_{t-1} + \cdots + \hat{\Phi}_{k-1} \hat{z}_{t-k+1} + \cdots + \hat{\Phi}_p \hat{y}_{t-p}
\]
\[
= \hat{c} + (\hat{\Phi}_1 L + \cdots + \hat{\Phi}_{k-1} L^k)\tilde{z}_t + \cdots + \hat{\Phi}_p \hat{y}_{t-p}
\]
Thus
\[
\hat{e}_t = z_t - \hat{z}_t
\]
\[
= \tilde{z}_t - (\hat{c} + (\hat{\Phi}_1 L + \cdots + \hat{\Phi}_{k-1} L^k)\tilde{z}_t + \cdots + \hat{\Phi}_p \hat{y}_{t-p})
\]
\[
= \tilde{y}_t + \tilde{\delta} - (\hat{c} + (\hat{\Phi}_1 L + \cdots + \hat{\Phi}_{k-1} L^k)\tilde{y}_t + \cdots + \hat{\Phi}_p \hat{y}_{t-p}) -
\]
\[
- (\hat{\Phi}_1 L + \cdots + \hat{\Phi}_{k-1} L^k)\tilde{\delta}
\]
\[
= \hat{e}_t + \tilde{\eta}_k
\]
where \( \tilde{\eta}_k = (I_n - \hat{\Phi}_1 - \cdots - \hat{\Phi}_k)\tilde{\delta} \), \( k = 0, 1, 2, \ldots, (t-t_0) \).

Therefore the effect on VAR residual chart statistic is
\[ T_{t}^{2} = \text{shift}T_{t}^{2} = 2\hat{\eta}_{k}'\hat{\Omega}^{-1}\hat{\eta}_{k} + \hat{\eta}_{k}'\hat{\Omega}^{-1}\hat{\eta}_{k} \quad \text{for } t \leq t \leq p, \text{ and } k = t - t_{0}. \]

This is still a positive shift on the mean of \( T^{2} \) and on the variance of \( T^{2} \). From the definition of \( \hat{\eta} \) and \( \hat{\eta}_{k} \), the effects during \( t_{0} \leq t \leq p \) can be either larger or smaller than the effects during \( t \geq p \), depending on the specific VAR structure \( \Phi(L) \) and \( \hat{\Omega} \).

In the above expressions, we used the estimated Q.E.D.

**APPENDIX B: The Proof on Lemma 2**

**Lemma 2.** If the process covariance \( \Omega \) shifts to \( \Omega_{1} \), then \( \Omega_{1} = D \Omega D' \), and the effect on \( \text{VAR chart} \) is \( \text{shift}T_{t}^{2} = \hat{\epsilon}_{t}' D^{-1} \hat{\Omega}^{-1} D^{-1} \hat{\epsilon}_{t} \).

**Proof:**

If the process covariance \( \Omega \) shifts to \( \Omega_{1} \), then since the covariance matrices are always positive definite symmetric matrices, there must exist a nonsingular matrix \( D \) so that \( \Omega_{1} = D \Omega D' \). This can be easily seen from the decomposition \( \Omega = PA\Lambda P' \) and \( \Omega_{1} = P_{1}\Lambda_{1}P_{1}' \), where \( I = PP' = P_{1}P_{1}' \), and \( \Lambda_{1} = C^{1/2} \Lambda C^{1/2} \) or \( C^{1/2} \Lambda^{1/2} = \Lambda_{1}^{1/2} \) for the positive diagonal matrices \( \Lambda \), \( \Lambda_{1} \), \( C \). That is, \( D = P_{1}C^{1/2}P' \) and \( D^{-1} = P'C^{-1/2}P_{1}' \).

From Cholesky factorization \( \Omega = (PA^{1/2})(\Lambda^{1/2}P') \), \( \Omega_{1} = (P_{1}\Lambda_{1}^{1/2})(\Lambda_{1}^{1/2}P_{1}') \), with estimated in-control covariance matrix \( \hat{\Omega} = (\hat{P}\hat{\Lambda}_{1}^{1/2})(\hat{\Lambda}_{1}^{1/2}\hat{P}') \), we have the residuals before the covariance shift, \( \hat{\epsilon}_{t} = (PA^{1/2})\tilde{u}_{t} \). And after the covariance shift the residuals can be also expressed as \( \hat{\epsilon}_{t} = (P_{1}\Lambda_{1}^{1/2})\tilde{u}_{t} \), where \( \tilde{u}_{t} \sim N(\tilde{0},I) \) is normalized sequence.
Without knowing the covariance shift, the Hotelling statistic is based on the in-control covariance estimate and with the updated one-step-ahead forecasting errors. Therefore the effect of the covariance shift on VAR chart is

\[ T^2_{\text{shift}} = \hat{e}_t^* \hat{\Omega}^{-1} \hat{e}_t = \tilde{u}_t^* \Lambda_1^{1/2} P_1 \Omega^{-1} P_1 \Lambda_1^{1/2} \tilde{u}_t \]

\[ = \tilde{u}_t^* \Lambda_1^{1/2} P' P \Lambda^{1/2} \Lambda_1^{1/2} P_1 \Omega^{-1} P_1 \Lambda_1^{1/2} \Lambda^{-1/2} P' P \Lambda^{1/2} \tilde{u}_t \]

\[ = \hat{e}_t^* D' \hat{\Omega}^{-1} D \hat{e}_t \]

where \( D = P_1 \Lambda_1^{1/2} \Lambda^{-1/2} P' = P_1 C^{1/2} P' \) and \( \hat{e}_t = D \hat{e}_t \). Hence the out-of-control statistic is Hotelling-style that is constructed from amplified residuals and the original covariance.

Q.E.D.

APPENDIX C: The Proof on Lemma 3

**Lemma 3.** For a VAP(p) process, if the coefficients shift from \( \Phi \) to \( \Phi + \Delta \Phi \), the effect on the Hotelling \( T^2 \) for the residuals is

\[ T^2_{\text{shift}} = T^2_{\text{noshift}} + 2 \sum_{i=1}^{p} (\Delta \Phi_i (\tilde{y}_i - \bar{\mu}))' \Omega^{-1} \hat{\varepsilon}_t + \sum_{j=1}^{p} (\sum_{i=1}^{p} (\Delta \Phi_i (\tilde{y}_i - \bar{\mu})))' \Omega^{-1} \Delta \Phi_j (\tilde{y}_i - \bar{\mu}) \]

**Proof:**

For the VAR process without coefficient shift,

\[ \tilde{y}_i = \bar{c} + (\Phi_1 L + \Phi_2 L^2 + \cdots + \Phi_p L^p) \tilde{y}_i + \hat{\varepsilon}_t \]

the one-step ahead forecasting based on non-shifted coefficients is

\[ \hat{\tilde{y}}_{it+1} = \bar{c} + (\Phi_1 L + \Phi_2 L^2 + \cdots + \Phi_p L^p) \tilde{y}_i \]

Since the means of the processes do not shift, and noticing \( \bar{c} = (I_n - \Phi_1 - \cdots - \Phi_p) \bar{\mu} \), then the processes after the coefficient shift is,
\[
\tilde{y}_i = (1 - \Phi - \Delta \Phi) \bar{\mu} + (\Phi_1 L + \Delta \Phi_1 L + \Phi_2 L^2 + \Delta \Phi_2 L^2 \cdots + \Phi_p L^p + \Delta \Phi_p L^p) \tilde{y}_i + \tilde{e}_i
\]

Hence, the residuals after the coefficient shift are
\[
\hat{\tilde{e}}_i = \tilde{y}_i - \tilde{\tilde{y}}_i = \tilde{e}_i + (\Delta \Phi_1 L + \cdots + \Delta \Phi_p L^p) \tilde{y}_i - \Delta \Phi \bar{\mu}
\]
\[
= \tilde{e}_i + (\Delta \Phi_1 L + \cdots + \Delta \Phi_p L^p)(\tilde{y}_i - \bar{\mu})
\]

Therefore, the Hotelling-style statistic will be,
\[
T^2_{\text{shift}} = \hat{\tilde{e}}_i' \Omega^{-1} \hat{\tilde{e}}_i
\]
\[
= T^2_{\text{no shift}} + 2 \sum_{i=1}^{p} (\Delta \Phi_i (\tilde{y}_i - \bar{\mu}))' \Omega^{-1} \hat{\tilde{e}}_i + \sum_{i=1}^{p} \left( \sum_{j=1}^{p} (\Delta \Phi_i (\tilde{y}_i - \bar{\mu}))' \Omega^{-1} \Delta \Phi_j (\tilde{y}_j - \bar{\mu}) \right)
\]

Note \( \Phi \) has \( p \) lags, thus the second item in this \( T^2_{\text{shift}} \) includes as many as \( p \) terms and the third item includes as many as \( p^2 \) terms. Also note the mean vector of the processes is not changed, so \( \tilde{y}_i - \bar{\mu} \) has means of zeros and can be viewed similar to \( \tilde{e}_i \) but with different covariance matrix (e.g., the variances in the diagonal are larger than those in \( \Omega \)). Hence the third item has at least \( p \) quadratic forms corresponding to \( i=j \), and the second item has \( p \) random quantities with mean of zeros.

Q.E.D.
Figure 1-2-4 (a)

Trial Run Chart for Stage I, Multivariate Five Variables

Observations, alpha=0.0027

T-square Statistic

Figure 4(b) Phase I Real Five Variables

Observations, alpha=0.0027

T-square Statistic

Figure 4(c) Phase II Real Five Variables

Observations, alpha=0.0027, run length=13
REFERENCES


Our responsibility is to provide strong academic programs that instill excellence, confidence and strong leadership skills in our graduates. Our aim is to (1) promote critical and independent thinking, (2) foster personal responsibility and (3) develop students whose performance and commitment mark them as leaders contributing to the business community and society. The College will serve as a center for business scholarship, creative research and outreach activities to the citizens and institutions of the State of Rhode Island as well as the regional, national and international communities.

The creation of this working paper series has been funded by an endowment established by William A. Orme, URI College of Business Administration, Class of 1949 and former head of the General Electric Foundation. This working paper series is intended to permit faculty members to obtain feedback on research activities before the research is submitted to academic and professional journals and professional associations for presentations. An award is presented annually for the most outstanding paper submitted.