Dynamic Analysis of Variance Methods for Multivariate Control System Data

T.J. Harris*, C.T. Seppala and D.W. Bacon

Department of Chemical Engineering, Queen’s University
Kingston, ON Canada
August, 1999

* harrist@post.queensu.ca

Abstract:

Most control system data from chemical processes are multivariate in nature and characterized by serial correlation which arises from the presence of inertial elements such as tanks, reactors and recycle streams. The analysis and interpretation of this data are further complicated by the presence of process upsets, production rate changes or other interventions. Vector autoregressive (VAR) models provide a convenient framework in which to model and analyze these types of behavior. VAR models are easily estimated using ordinary least squares, well-understood, and suitable for including vectors of exogenous inputs or deterministic time trends. Estimated VAR models can be used as a basis for the computation of: i) impulse response functions and multi-step ahead forecast error variance decompositions, ii) the setpoint or target component of the variance of controlled variables, so that this component may be distinguished from the variability arising from other types of disturbances, and iii) graphical and numerical statistics for assessing the interactions between variables.

Other dynamic analysis methods have been developed in econometrics that can be applied and/or extended to aid in the analysis of control system data. Cointegration analysis has evolved into an important theoretical and applied research topic in the econometrics literature during the last fifteen years. A set of non-stationary, stochastic variables are said to be cointegrated if a linear combination of these variables can be found that produces a stationary series. Non-stationary behavior in multivariate time series has traditionally been handled by separately differencing each integrated variable, a practice that can result in information loss and inferior model performance when cointegration is present. Cointegration methods can be directly incorporated into VAR model-based methods, extending their flexibility to analyze non-stationary disturbances.

1 Introduction:

Time series methods have found wide-spread applications in process identification, process control, process monitoring, assessment of process control performance and general process analysis. Most applications have used univariate time-series models, popularized
by Box and Jenkins (1970), and Astrom (1967). The general form of a univariate time series model can be written as:

\[ w_t = \alpha + \phi_1 w_{t-1} + \ldots + \phi_p w_{t-p} + a_t - \theta_1 a_{t-1} - \ldots - \theta_q a_{t-q} \]

where

\[ w_t = \nabla^d Y_t \equiv (1-q^{-1})^d Y_t \]

and

\[ Y_t = y_t - y_{ref} \]

In Equation (1.1) \( y_t \) is the observed value of the process variable at time \( t \) and \( y_{ref} \) is a reference value or target value, which might be the mean value of the process. \( \alpha \) is a constant term. \( q^1 \) is the shift operator defined as \( q^1 Y_t = Y_{t-1} \). The sequence \( \{a_t\} \) is the driving force for the stochastic difference equation (1.1). \( \{a_t\} \) is usually modelled as a set of independently and identically distributed (iid) random variables with constant mean zero and variance \( \sigma_a^2 \). The difference operator \( d \) allows for non-stationary, or drifting behavior in the process variable whenever \( d > 0 \). Equation (1.1) is known as an Autoregressive-Integrated-Moving Average (ARIMA) model of order \( (p,d,q) \). The time series is stationary if it does not exhibit explosive growth. It is invertible if the \( \{a_t\} \) sequence can be recovered from observed values of the process using the time series model. Both of these conditions imply restrictions on the parameters.

Time series models are very flexible and can readily represent a wide variety of observed process behavior. In particular, by changing the distributional assumptions of \( \{a_t\} \), one can represent both stochastic and deterministic disturbances. An example of the latter would be steps, ramps or exponential rises. The basic time series methods proposed by Box and Jenkins and Astrom in the late 1960’s have been expanded considerably to accommodate interventions or special events, outliers, and inclusion of explanatory or endogenous variables, which may be stochastic or deterministic (Wei, 1990). The iterative model building procedure proposed by Box and Jenkins has been enormously facilitated by the availability of personal computing and sophisticated graphical user interfaces. Many of the early approaches to specifying the model structure (i.e. choice of \( p,d,q \)) were based on recognizing patterns in the auto and cross correlation functions. Structural specification is now most often accomplished as a natural outcome of model order specification (Gustafsson and Hjalmarsson, 1995). All of the model order specification methods essentially have two components; one which ‘rewards’ a good fit, and a second which ‘penalizes’ model complexity.

Although knowledge of time series methods has been widespread in the control community for thirty years, there are new important uses for these methods. In particular, the application of time series methods for assessing and monitoring the performance of control loops has emerged as an important topic of both theoretical and practical interest (see Harris et al (1998) and Qin (1998) for recent reviews). Engineering control systems are widely used to regulate process variables at their target values in the presence of measured and unmeasured process disturbances. The effectiveness of these control
schemes often changes due to changes in the disturbance structure or disturbance/process variable interactions. Most applications of statistical methods for characterizing the performance of control systems have involved one output variable (y) and possibly several explanatory or predictive variables. Direct applications of performance assessment techniques to multivariable systems are more involved, and require considerable a priori knowledge about what is known as the time-delay structure of the multivariate process (Harris et al, 1996; Huang et al 1997a,1997b).

While much effort has been directed towards deriving, developing, and implementing such control performance assessment methods, considerably less work has emerged on complementary analysis and diagnosis techniques. Performance assessment statistics are good for detecting changes in variability, but are less useful for uncovering the underlying causes for output variance inflation. Many of the opportunities for understanding process variability arise from understanding the multivariate behavior of the process. In the remainder of the paper, we shall outline the use of multivariate time series methods for process analysis. In particular, we shall focus on a number of dynamic analysis methods recently developed in the field of econometrics.

2 Multivariate Time Series:

The extension to multivariate systems follows intuitively from (1.1):

\[
\begin{align*}
0 \phi_0 w_t = \alpha + \phi_1 w_{t-1} + \ldots + \phi_p w_{t-p} + \theta_0 a_t - \theta_1 a_{t-1} - \ldots - \theta_q a_{t-q} \\
\text{where} \\
w_t = \nabla^d Y_t \equiv \text{diag}(\nabla^{d_1} Y_{i_1}, \nabla^{d_2} Y_{i_2}, \ldots, \nabla^{d_n} Y_{i_n}) \\
\text{and} \\
Y_t = y_t - y_{\text{ref}}
\end{align*}
\]

(2.1)

y is now a vector of n variables, and the parameters \{\phi_i, \theta_i\} are n x n matrices. a is an n-dimensional stochastic process with the properties:

\[
E[a_t] = 0, E[a_t a_{t-\tau}^T] = 0, t \neq \tau, \text{and } E[a_t a_{t-\tau}^T] = \Sigma_a
\]

(2.2)

\Sigma_a is the covariance matrix of the driving forces. Equation (2.2) is known as a vector autoregressive-integrated-moving-average (VARIMA) model.

The representation in Equation (2.2) is not unique and the coefficients are not identifiable unless certain restrictions are imposed (Hannan, 1976; Kashyap and Rao, 1976). Just as in the univariate case, conditions exists for ascertaining stability and invertibility.

As in the univariate case, it is convenient in the multivariate case to use a purely vector autoregressive (VAR) representation. The Final Prediction Error (FPE) representation of a VAR model is obtained by setting \(\phi_0=I_n, \theta_0=I_n, \text{ and } \theta_i=0, i>0, \text{ i.e.}:
\]

\[
E[a_t] = 0, E[a_t a_{t-\tau}^T] = 0, t \neq \tau, \text{and } E[a_t a_{t-\tau}^T] = \Sigma_a
\]

(2.2)
\[ w_t = \alpha + \phi_1 w_{t-1} + \ldots + \phi_p w_{t-p} + a_t \]  

(2.3)

The term FPE arise from the observation that the \( a_t \) can be interpreted as the error in making a one step-ahead forecast.

The VAR representation in Equation (2.3) is not unique. A second representation, known as an orthogonalized innovations model, is of the form:

\[ \phi_0 w_t = \bar{\alpha} + \phi_0 w_{t-1} + \ldots + \phi_p w_{t-p} + v_t \]  

(2.4)

The n-dimensional stochastic driving force \( \{v_t\} \) has diagonal covariance matrix \( \Sigma_v \). The orthogonalized innovations form is readily obtained form the FPE representation via the transformation \( a_t = \phi_0^{-1} v_t, \phi_i = \phi_0^{-1} v_i, i = 1 \ldots p \), with \( \phi_0 \) defined as the inverse of the lower diagonal Cholesky factor of \( \Sigma_a \), i.e. \( \Sigma_a = PP^T, \phi_0 = P^{-1} \). With this convention \( \phi_0 \) is lower triangular. Although the two models can be obtained from each other via a simple transformation, each accounts for instantaneous correlation in a fundamentally different fashion. In the FPE form, instantaneous correlation, i.e. correlation between variables at the same instant in time, is represented through the covariance matrix \( \Sigma_a \), whereas in the orthogonalized innovations form, instantaneous correlation is directly achieved through the \( \phi_0 \) coefficient matrix. Equation (2.4) is referred to as a structural form model.

The VAR model has some advantages over the VARIMA model that make it suitable for applied work, including: 1) the structural identification stage reduces to a numerical problem of autoregressive order selection, 2) under reasonable assumptions, ordinary least squares (OLS) regression can be used to perform the estimation, and 3) many methods have been developed for interpretation of the estimated coefficients.

The VAR model building process proceeds according to the following steps: 1) stationarity testing, 2) order selection, 3) parameter estimation, 4) residual diagnostics. Steps 2) to 4) are repeated until the lowest order model with the best residual properties is found. Note the similarity to the well-known univariate model building procedure.

There are many ways to estimate VAR models: multivariate least squares, generalized least squares (GLS), the discrete Kalman filter, and ordinary least squares (OLS) (Hamilton, 1994; Lütkepohl, 1991; Davidson and MacKinnon, 1993; Suoranta, 1990). OLS is the most convenient method from an implementation and ease of estimation perspective. If it is assumed that the system dynamics can be approximated by a single model order applied to all n equations in the vector process, then the right hand side in Equation (2.3) will have exactly the same regressors across all n equations. This allows the same results to be derived using equation-by-equation ordinary least squares (OLS) in place of the more complex generalized least squares approach. Thus, the \( \phi_i \) coefficient matrices can be easily estimated by performing n multi-input single-output regressions on each of the system variables one by one. This regression can easily be computed using any standard time
series analysis or regression software. Algorithms for VAR estimation that make use of the assumption that the regressors are the same in each equation are especially efficient.

The use of multivariable methods to analyze control systems and chemical processes is not new. Devries and Wu (1978) used VARIMA models and multivariate spectral analysis to assess the effectiveness of control on basis weight variation in a paper machine. Although they attempted to use the full VARIMA structure, only the autoregressive components were found to be significant. Tade and Bacon (1984) apply multivariate time series methods to analyze a petroleum fractionation unit. Multivariate autoregressive models have been used in at least one proprietary software package to analyze variation in control system data, i.e., KCL-Wedge from KCL Development Oy in Finland (Ritala, 1993; Ihalainen and Ritala, 1996). The background theory for these applications can be found in Suoranta (1990).

Before discussing a number of post-estimation diagnostics and analysis methods, it is important to note that a stationary VARIMA model can also be expressed in Vector Moving Average (VMA) form:

\[ w_t = \beta + a_t + \Psi_t a_{t-1} + \Psi_2 a_{t-2} + \ldots \]

or

\[ w_t = \beta + \overline{\Psi}_0 v_t + \overline{\Psi}_1 v_{t-1} + \overline{\Psi}_2 v_{t-2} + \ldots \]

The second representation in Equation (2.5) arises when a structural form VAR model has been used to represent the process. In this representation \( \overline{\Psi}_0 \) is lower triangular. Although the FPE model is most useful for estimation, post-estimation diagnostics and analysis methods make use of both representations.

Before introducing the topic of cointegration, we will briefly review some post diagnostics and analysis methods. These methods can provide tremendous insight into the dynamic behavior and interactions of multivariate processes.

3.0 Post Diagnostics and Analysis Methods:

3.1 Assessment of Predictive Power / Granger Causality

Granger causality is the name given to the correlation test proposed by Granger (1969). While the term causality figures in the name of this test, it is acknowledged that in practice the test cannot distinguish between true causality and correlation. With the understanding that Granger causality is not a test for causal relationships, the original nomenclature will be used throughout this paper.

According to Granger’s (1969) definition, a causal relationship exists between two variables if one variable can help improve the mean square error of the forecast of the other variable. Equivalently, variable \( y_1 \) Granger-causes variable \( y_2 \) if removing \( y_1 \) from the information set available to predict \( y_2 \) inflates the mean square error of the optimal
prediction of $y_2$. While this is a very intuitive definition, in practice it is strictly correlation-based and must be interpreted like any other correlation-based statistic. The mathematical statement of Granger causality can be found in Granger (1969), Hamilton (1994) and Lütkepohl (1991).

Granger-noncausality can be characterized either from VAR model coefficients or from the equivalent moving average form. A necessary and sufficient condition for $y_1$ not to be Granger-caused by $y_2$, based on the VAR model in Equation (2.3), is:

$$\phi_{i,2}^{(i)} = 0, \quad i = 1, 2, \ldots, p$$

(3.1)

where $\phi_{i,2}^{(i)}$ is the $i$th order autoregressive coefficient for $y_1$ in the equation for $y_1$ and $p$ is the VAR model order. A necessary and sufficient condition for $y_1$ not to be Granger-caused by $y_2$, based on the model in Equation (2.5), is:

$$\Psi_{i,2}^{(i)} = 0, \quad i = 1, 2, \ldots$$

(3.2)

where $\Psi_{i,2}^{(i)}$ is the $i$th order moving average coefficient for driving force $a_2$ in the equation for $y_1$. In Lütkepohl (1991) it is shown that these results are a consequence of the fact that equality of the 1-step ahead predictors in a VMA model implies the equality of the h-step ahead predictors. In practice, a number of hypothesis tests on the estimated coefficients are used to establish Granger causality.

The results of the Granger tests can conveniently be presented in graphical form as shown in Figure 1. In this figure, the arrows represent the case where the null hypothesis of no predictive power was rejected at a given significance level. Where there is no arrow, the null hypothesis of no predictive power was not rejected. The graphical representation facilitates the interpretation of the test results, especially when $n>2$. Graphical representations of correlation are often used in multivariate time series analysis (Tiao and Box, 1981). For example, Tade and Bacon (1984) used diagrams similar to those shown in Figure 1 to represent significant cross-correlation between process variables in a multivariate analysis of petroleum fractionation unit. Granger causality tests can serve to highlight some the bulk dynamic interaction properties of process variables, such as independence, one-way interactions, and feedback relationships.

### 3.2 Multivariate Impulse Response Analysis

In econometrics, impulse response analysis is routinely used as an analysis tool for investigating the properties of estimated models. The use of impulse response analysis for analyzing univariate control system data has proven to be quite useful (Tyler and Morari, 1996; Kozub, 1997; Harris et al 1998). The multivariate impulse responses for $w_t$ are the $\Psi_i$ coefficient matrices from Equation (2.5), for $i=0, \ldots, \infty$. Examination of the numerical contents of these matrices is not very helpful for visualization of the dynamics of $w_t$. It is more efficient to display the values of the coefficient matrices in the form of an $n \times n$
matrix of impulse response plots. The \((i,j)\)th impulse response plot shows the response of variable \(j\) to a single shock in the innovations process of variable \(i\) at time \(t = 0\). The size of the innovations shock is up to the user; the impulse responses are simply interpreted as being scaled (Lütkepohl, 1991). The most common shock sizes are unit shocks (because this produces the actual moving average coefficient values) and standard deviation shocks (representing the average size of expected deviations in the innovations process). The impulse response functions will allow qualitative observations to be made regarding the dynamic interactions of the process variables. Important features such as oscillations, settling time, damping and interaction between outputs will be directly visible. We note that the impulse responses can also be calculated directly from the VAR representation.

Computation of the asymptotic distributions of estimated impulse responses in order that approximate standard error limits may be computed is a tedious task. See, for example, the discussion in Lütkepohl (1991). One alternative method, based on Monte Carlo integration, has been developed, and is included in the time series analysis software package RATS (Doan, 1995). The procedure uses brute force computing power to take draws from the posterior distribution of the VAR coefficients and measure the resulting changes in the impulse responses to estimate their moments. For more details on this procedure see Kloek and VanDijk (1978) and Zellner (1971). Experience with this method of computing the standard error limits has shown it to be a reliable approximation.

3.3 Forecast Error Decomposition:

In the application of univariate models for performance assessment and analysis, the forecast error decomposition of a process variable is extremely valuable (Harris et al, 1998). To illustrate these concepts, consider the moving-average representation of a univariate process:

\[
w_t = \alpha + a_t + \varphi_1 a_{t-1} + \varphi_2 a_{t-2} + \ldots
\]  

(3.3)

where \(\{a_t\}\) is an iid process with mean 0 and variance \(\sigma_a^2\). It is well known that the \(k\)-step ahead minimum-mean-square-error forecast error for \(w_{t+k}\), is given by:

\[
e_{t+k/t} \equiv w_{t+k} - \hat{w}_{t+k/t} = \sum_{j=0}^{k-1} \psi_j a_{t+k-j}
\]  

(3.4)

where the forecast is denoted by \(\hat{w}_{t+k/t}\). When \(y_t\) is nonstationary, one can readily construct the minimum-mean-square forecast for \(y_{t+k}\). The \(k\)-step ahead forecast variance for \(w_t\) is given by:

\[
\sigma_w^2(k) = \sum_{i=0}^{k} \psi_i^2 \sigma_a^2
\]  

(3.5)
As $k \to \infty$, $\sigma_w^2(k) \to \sigma_w^2$. Equation (3.5) clearly emphasizes the importance of the impulse weights in analysis of process variability.

In multivariate processes, the $k$-step ahead forecast error variance can conveniently be expressed in terms of VMA matrix coefficients. This is true whether the VMA model is in FPE form or orthogonalized innovations form. Although we will only illustrate the results for the FPE form, it can be shown that the $k$-step ahead forecast error is invariant to the structural form chosen (Seppala, 1999). The vector moving average form for $w_1$ can be written as:

$$w_{1,t} = a_{1,t} + \Psi^{(1)}_{1,1}a_{1,t-1} + \Psi^{(2)}_{1,1}a_{1,t-2} + \ldots + \Psi^{(q)}_{1,1}a_{1,t-q} + \ldots + a_{n,t} + \Psi^{(1)}_{1,n}a_{n,t-1} + \Psi^{(2)}_{2,n}a_{n,t-2} + \ldots + \Psi^{(q)}_{r,n}a_{n,t-q}$$

where $\Psi^{(i)}_{1,j}$ are the coefficients relating $w_1$ to $\{a_{j,t}, t=1,\ldots\}$. These coefficients are obtained by inspection from the VMA model given in Equation (2.5). In a manner analogous to the univariate process, the $k$-step ahead minimum-mean-square-error forecast error for $w_{1,t+k}$ is given by:

$$e_{1,t+k} = w_{1,t+k} - \hat{w}_{1,t+k}$$

$$= \sum_{j=0}^{k-1} \Psi^{(j)}_{1,1}a_{1,t+k-j} + \sum_{j=0}^{k-1} \Psi^{(j)}_{1,2}a_{1,t+k-j} + \ldots + \sum_{j=0}^{k-1} \Psi^{(j)}_{1,n}a_{1,t+k-j}$$

The $k$-step ahead forecast variance for $w_1$ is given by:

$$\sigma_{w,1}^2(k) = \sum_{j=1}^{k} [\Psi^{(j)}_{1,1}]^2 \sigma_i^2 + \sum_{j=1}^{k} [\Psi^{(j)}_{1,2}]^2 \sigma_i^2 + \ldots + \sum_{j=1}^{k} [\Psi^{(j)}_{1,n}]^2 \sigma_i^2 + 2 \sum_{j=0}^{k-1} \Psi^{(j)}_{1,1}\Psi^{(j)}_{1,2}\sigma_i\sigma_j + \ldots + 2 \sum_{j=0}^{k-1} \Psi^{(j)}_{1,1}\Psi^{(j)}_{1,n}\sigma_i\sigma_n$$

In Equation (3.8) $\sigma_i^2$ refers to the variance of $\{a_{i,t}\}$ and $\sigma_i\sigma_m$ refers to the covariance between $\{a_{i,t}\}$ and $\{a_{m,t}\}$. The nomenclature in Equation (3.8) is very cumbersome, and it is convenient to express $\sigma_{w,1}^2(k)$ as:

$$\sigma_{w,1}^2(k) = \sum_{i=0}^{n} \langle \Psi_{1,i}^2 \rangle_{k} \sigma_i^2 + 2 \sum_{i=0}^{n-1} \sum_{m=i+1}^{n} \langle \Psi_{1,i} \Psi_{1,m} \rangle_{k} \sigma_i \sigma_m$$

where:
\[
\left\langle \Psi_{1,i} \sigma_{1,m} \right\rangle_k = \sum_{j=0}^{k} \Psi^{(j)}_{1,i} \Psi^{(j)}_{1,m}
\]  
(3.10)

As \( k \to \infty, \sigma^{2}_{w}(k) \to \sigma^{2}_{w} \). The interpretation, calculation and display of the results in Equation (3.9) is less cumbersome than the notation! The first term on the right hand side of Equation (3.9) gives the direct contribution to the variance of the prediction error from the presence of the driving forces. Each of the terms inside of the summation on this first term is positive. The second term gives the contribution to the variance which arises from the off-diagonal elements of \( \Sigma_{a} \). These covariance terms need not be positive. Equation (3.9) effectively describes an analysis of variance with respect to the driving forces. The results can be conveniently presented in a stacked bar plot (Seppala, 1999). The effect of removing a source of variation can be readily ascertained from Equation (3.9).

Desborough and Harris (1992) and Kozub (1997) describe applications in process analysis and assessment of control loop performance for the case where \( \Sigma_{a} \) is diagonal. In this latter case, the analysis of variance as \( k \to \infty \) can be expressed in terms of the driving forces or equivalently in terms of the other variables, \( \{w_{j}, j=2\ldots n\} \).

4 Cointegrated Systems:

In many processes, some variables will meander or drift considerably. It might be anticipated that that certain linear combinations of these nonstationary variables might be stationary. The original variables are not ‘tied’ to each other over short time frames. However, over a larger time-frame these variables are connected or ‘tied’ together. In the econometrics literature, an equilibrium relationship is said to exist between these variables. Linear combinations of nonstationary variables which are stationary are known as cointegrated variables.

To introduce the concepts of cointegration, we will consider the following example:

\[
y_{1,t} = \gamma y_{2,t} + e_{1,t}
y_{2,t} = y_{2,t-1} + a_{2,t}
\]

where
\[
e_{1,t} = \phi e_{1,t-1} + a_{1,t}
\]

\( \{a_{1,t}, a_{2,t}\} \) are iid normal random variables with mean zero and covariance

\[
\Sigma = \begin{bmatrix}
\sigma^{2}_{a1} & 0 \\
0 & \sigma^{2}_{a2}
\end{bmatrix}
\]

(4.2)

We will consider the following cases.
Case A:  \( \{\gamma = 0, \phi = 1\} \): \( y_1 \) and \( y_2 \) are both nonstationary and independent.

Case B:  \( \{\gamma = -.5, \phi = 1\} \): \( y_1 \) and \( y_2 \) are both nonstationary. The nonstationary behavior in \( y_1 \) arises from its own source of integration plus a contribution from \( y_2 \). This is a not a cointegrated system. We shall refer to this as mixed since there are two sources of integration in \( y_1 \).

Case C:  \( \{\gamma = -.5, \phi = .8\} \): \( y_1 \) and \( y_2 \) are both non-stationary. The nonstationary behavior in \( y_1 \) arises solely from \( y_2 \). \( y_1 \) and \( y_2 \) are co-integrated.

Figure 2 shows simulations of these processes with \( N = 1000 \). \( \Sigma_a \) has been adjusted so that the sample variances of \( e_1 \) and \( y_2 \) are both 1. The results of a regressing \( y_1 \) on \( y_2 \) are shown in Table 1 for each of the three cases, as well as the results of regressing \( \nabla y_1 \) on \( \nabla y_2 \).

<table>
<thead>
<tr>
<th>Case</th>
<th>Independent Source Integration?</th>
<th>( \hat{\gamma} ) (std error)</th>
<th>( \hat{\gamma} ) (std error)</th>
</tr>
</thead>
<tbody>
<tr>
<td>A: ( \gamma = 0, \phi = 1 )</td>
<td>Yes</td>
<td>0.571 (0.026)</td>
<td>0.010 (0.116)</td>
</tr>
<tr>
<td>B: ( \gamma = -.5, \phi = 1 )</td>
<td>No (Mixed)</td>
<td>0.071 (0.026)</td>
<td>-0.491 (0.116)</td>
</tr>
<tr>
<td>C: ( \gamma = -.5, \phi = .8 )</td>
<td>No</td>
<td>-0.442 (0.032)</td>
<td>-.626 (0.374)</td>
</tr>
</tbody>
</table>

Table 1:  Regression Analysis results for Case A-C:

There are a number of interesting observations that can be made from Table 1.

- The estimate of the regression coefficient in Case A is statistically significant, but misleading. This is known as a spurious regression (Granger and Newbold, 1974). A spurious regression results when the regressor, explanatory variable and additive error are each nonstationary. A spurious regression arises when a nonstationary variable is regressed upon another independent nonstationary variable.

- For Case B, a reasonable value for the regression coefficient is obtained when \( \Delta y_1 \) is regressed on \( \Delta y_2 \). This is not unexpected since we can express the true process as:

\[
\nabla y_1 = \gamma \nabla y_2 + a_{1,t}
\]

The regression variable and explanatory variable are both stationary. The additive error satisfies the assumptions of independence required to ensure that the least squares
estimates are consistent. When $y_1$ is regressed on $y_2$, the additive error is nonstationary. We recall that this is not a cointegrated system since $y_1$ has its own independent source of integration.

- For Case C, a reasonable value for the regression coefficient is obtained when $y_1$ is regressed on $y_2$ and when $\Delta y_1$ is regressed on $\Delta y_2$. For the regression of $y_1$ on $y_2$, the model is given by:

$$y_1 = \gamma y_2 + e_{1,t}$$

(4.4)

The regressor and explanatory variables are non-stationary. The additive error is stationary, but not independent. In this situation, the regression coefficients are consistent, but they are not asymptotically normal (Granger and Newbold, 1974; Davidson and MacKinnon, 1993). $\hat{\gamma}$ converges to $\gamma$ at a rate proportional to $1/N$, and not $1/\sqrt{N}$ as might be expected.

For the regression of $\nabla y_1$ on $\nabla y_2$, the model can be expressed as:

$$\nabla y_1 = \gamma \nabla y_2 + s_t$$

where

$$s_t = \nabla e_t$$

and

$$\nabla e_t = \phi \nabla e_{t-1} + \nabla a_{1,t}$$

(4.5)

$\nabla e_t$ is a stationary (but noninvertible). The regression and explanatory variable are both stationary. The additive error however, is serially correlated.

Any attempt to directly use a time series model on the differenced data will result in incorrect estimates of the parameters. It is also obvious that any analysis using this model will also be in error.

In general, y is said to exhibit cointegrated behavior if each of the series exhibits non-stationary behavior (i.e. $d_i > 0$) and some linear combination of the variables $b^T y$ is stationary (Granger, 1981; Lütkepohl, 1991; Hamilton, 1994). In Case C, $b = (1, -\gamma)^T$.

Recently, Abadir and Taylor (1999) have proposed an alternate definition which allows one to consider fractional orders of integration (i.e. where $d > 0$, but is not integer valued).

In modelling the cointegrated time series in Case C, the VARIMA model is modified to include lagged values of the y’s in the time series model for the differenced y’s, i.e.:
Although $y_1$ and $y_2$ are nonstationary, the linear combination $\mathbf{b}^T \mathbf{y}$ is stationary. A Granger-causality diagram for Case C is shown in Figure 3.

When $n > 2$, there may be more than one linear combination which is nonstationary. This situation is readily accommodated (Engle and Granger, 1987; Hamilton, 1994; Johansen, 1991; Phillips and Ouliaris, 1988).

For a process which admits a VAR representation, it is convenient to model the process using the Error Correction Representation (Engle and Granger, 1987):

$$
\nabla y_t = \phi_1 \nabla y_{t-1} + \ldots + \phi_p \nabla y_{t-p} + c - B \mathbf{A} \mathbf{y}_t + \mathbf{e}_t
$$

where $c$ is a constant, $\mathbf{A} \mathbf{y}_t$ denotes the vector of linear combinations of $\mathbf{y}$'s which are stationary, $\mathbf{B}$ is a matrix of parameters to be estimated, and $\mathbf{e}_t$ is an $n$-dimensional stochastic process with the properties:

$$
E[\mathbf{e}_t] = 0, E[\mathbf{ee}^T] = 0, t \neq \tau, \text{ and } E[\mathbf{ee}^T] = \Sigma_e
$$

An overview of tests to detect the presence of cointegration and computational procedures to estimate the parameters in Equation (4.8) can be found in Hamilton (1994) and Lütkepohl (1991). The presence of cointegrating variables introduces restrictions on the moving average representation and autoregressive parameters (Engle and Granger, 1987).

As a final point, Engle and Yoo (1987) have made the following observations regarding cointegration and forecasting in multivariate time series. Given a vector of integrated variables from a cointegrated system:

- If all variables are differenced as would appear from their appropriate univariate properties, then the system no longer has a multivariate time series representation with an invertible moving average representation.

- Multi-step forecasts generated from cointegrated systems have a property not shared by general integrated systems: linear combinations of forecasts are identically zero for large horizons, regardless of the forecast origin and the forecast error variance for this linear combination goes to infinity as the horizon goes to infinity.

Elementary results from cointegration analysis have been applied in Seppala (1999) to some basic problems encountered in the analysis of control system data. Using plant data from a variable setpoint single-input-single-output loop and a simulated multiple-input-multiple-output process, it has been shown that cointegrated behavior can arise in process control monitoring applications. The presence of cointegration has been exploited to
estimate some important process parameters without external testing or intervention, and using only simple calculations. Other potential applications of cointegration are discussed in the next sections.

4.1 Analysis of Process Disturbances:

Often in process analysis, there are variables that are measured, but not directly controlled. Examples of these might include the concentration and temperature of feed material to a reactor or the flowrate of an intermediate product to a distillation column. Many times these variables exhibit nonstationary behavior. Often, material, energy and momentum balance constraints do not permit totally independent variation among the variables. That is, the variables will be ‘tied’ together. Cointegration analysis enables a more sophisticated and insightful analysis of the disturbances by decomposing the multivariate time series into integrated and stationary components. This information can be used to attempt to reduce variation due to integration nearer its source, rather than trying to handle all the variables that have inherited their integrated behavior.

4.2 Process Identification:

Accurate process models are required to implement advanced control strategies, such as model predictive control. The models are typically of the form:

\[ y_t = X_t + D_t \]  (4.9)

where \( X_t \) represents the dynamic effect of manipulated variables or process variables on the process output and \( D_t \) represents the effect of all unmeasured process disturbances. \( X_t \) is typically of the form

\[ X_t = \alpha_1 X_{t-1} + \ldots + \alpha_p X_{t-p} + \beta_1 U_{t-1} + \ldots + \beta_q U_{t-q} \]  (4.10)

where \( U_t \) is the process input or manipulated variable at time \( t \). \( D_t \) is typically modelled by a difference equation of the type shown in Equation (2.1) with \( w_t \) replaced by \( D_t \). Efficient estimates of the transfer function parameters \( \{\alpha's,\beta's\} \) requires that the structural form of the disturbances be correctly modelled. Many of the considerations described in the previous paragraph also apply to process identification.

4.3 Process Control:

In applications of model predictive control, (Camacho and Bordons, 1998) a model of form similar to Equation (4.9) is used to make predictions of the process behavior at time \( t+k \), i.e.:

\[ \hat{y}_{t+k/t} = X_{t+k} + \hat{D}_{t+k/t} \]  (4.11)
The manipulated variable, $U_t$, is typically chosen to minimize a quadratic objective function subject possibly to constraints on the magnitude of $U_t$ or constraints on the magnitude of the change in $U_t$. However an essential feature is the requirement to make predictions of the form shown in Equation (4.11). In most applications of model predictive control, the disturbances are represented by independent integrated disturbances. The effects on the design, tuning and performance of control algorithms when cointegrated behavior is present have not been investigated.

5 Summary:

There has been an increased use of time series methods to analyze processes control systems. Most applications use univariate time series representations. However, many insights can be obtained by using vector autoregressive models. VAR models require minimal structural information, are easy to estimate, and are compatible with a host of post-estimation statistical tests and diagnostics that aid in interpreting the estimated coefficients.

Powerful tools have been developed in the econometrics literature that allow for cointegrated behavior to be detected, modeled, and hence properly interpreted when it arises. Cointegration analysis methods also have potential beyond the problems associated with analyzing control system data. In any process modeling situation where integrated multivariate data is available, cointegration analysis methods can be used to decompose the multivariate time series into integrated and stationary components. The dimension of the integrated space can also be estimated, helping the analyst understand the underlying system. This information can be used to attempt to reduce variation due to integration nearer its source, rather than trying to handle all the variables that have inherited their integrated behavior. By providing a modeling framework for integrated and stationary behavior, cointegration analysis could also be helpful for controller design or tuning. For example, upon identification of common integrated stochastic trends, controllers nearer the source of integration could be aggressively tuned to minimize or eliminate the integrated behavior before it is propagated through the process. If such a tuning effort is successful, then it may be possible to de-tune downstream loops for potential energy savings. Much work remains to be done on the potential applications of cointegration methods in the analysis of data from physical processes.

6 References


Gustafsson F., and H. Hjalmarsson, ((1995)’Twenty-one ML estimators for model selection’, *Automatica* 31, 1377-1392


Seppala, C.T., (1999), Ph.D. Thesis, Queen’s University, Kingston, ON, Canada

Suoranta, R. (1990), *Analysis of Linear systems Applying Multivariate Autoregressive Model*, Ph.d. thesis, Technical Research Centre of Finland

\[ y_1 \quad \quad y_2 \]

\[ y_1 \quad \longrightarrow \quad y_2 \]

\[ y_1 \quad \swarrow \quad y_2 \quad \searrow \quad y_1 \]

**Figure 1: Graphical representation of Granger causality tests**

i) \( y_1 \) and \( y_2 \) are independent of each other

ii) \( y_1 \) Granger-causes \( y_2 \)

iii) mutual correlation or feedback between \( y_1 \) and \( y_2 \)
Figure 2: Simulation of Cases A-C
Figure 3: Granger causality results for Case A & C

i) VAR in Differenced Variables (Case A)

ii) Engle-Granger Approach (Case C) \[ w = y_{1,t} - \gamma y_{2,t} \]