Chapter 43
The VARMAX Procedure

Contents

Overview: VARMAX Procedure ....................................................... 2980
Getting Started: VARMAX Procedure ........................................... 2982
  Vector Autoregressive Model ......................................................... 2982
  Bayesian Vector Autoregressive Model .............................. 2989
  Vector Error Correction Model ..................................................... 2990
  Bayesian Vector Error Correction Model ......................... 2996
  Vector Autoregressive Fractionally Integrated Moving Average Model . 2997
  Vector Autoregressive Model with Exogenous Variables . 3001
  Parameter Estimation and Testing on Restrictions ............. 3004
  Causality Testing ................................................................. 3006
  Multivariate GARCH Models ..................................................... 3007
Syntax: VARMAX Procedure ........................................................ 3018
  Functional Summary ............................................................... 3018
  PROC VARMAX Statement ......................................................... 3021
  BOUND Statement ................................................................. 3024
  BY Statement ...................................................................... 3025
  CAUSAL Statement ................................................................. 3025
  COINTEG Statement ............................................................... 3026
  GARCH Statement ................................................................. 3030
  ID Statement ....................................................................... 3031
  INITIAL Statement ............................................................... 3032
  MODEL Statement ................................................................. 3034
  NLOPTIONS Statement .......................................................... 3049
  OUTPUT Statement ............................................................... 3049
  RESTRICT Statement ............................................................ 3050
  TEST Statement ................................................................. 3062
Details: VARMAX Procedure ......................................................... 3064
  Missing Values ................................................................. 3064
  VARMAX Model ................................................................. 3064
  Dynamic Simultaneous Equations Modeling ..................... 3068
  Impulse Response Function ..................................................... 3071
  Forecasting ................................................................. 3082
  Tentative Order Selection ...................................................... 3087
  VAR and VARX Modeling ....................................................... 3092
  Seasonal Dummies and Time Trends .............................. 3098
  Bayesian VAR and VARX Modeling ........................................... 3099
Overview: VARMAX Procedure

Given a multivariate time series, the VARMAX procedure estimates the model parameters and generates forecasts that are associated with vector autoregressive moving average processes with exogenous regressors (VARMAX) models. Often, economic or financial variables are not only contemporaneously correlated with each other, but also correlated with each other’s past values. You can use the VARMAX procedure to model these types of time relationships. In many economic and financial applications, the variables of interest (dependent, response, or endogenous variables) are influenced by variables external to the system under consideration (independent, input, predictor, regressor, or exogenous variables). The VARMAX procedure enables you to model the dynamic relationships both among the dependent variables and between the dependent and independent variables.

A VARMAX model is defined in terms of the orders of the autoregressive or moving average processes (or both). When you use the VARMAX procedure, these orders can be specified by options or they can be automatically determined according to the information criteria. The VARMAX procedure supports the following information criteria: Akaike’s information criterion (AIC), the corrected AIC (AICC), the Hannan-Quinn criterion (HQC), the final prediction error (FPE), and the Schwarz Bayesian criterion (SBC), which is
also known as the Bayesian information criterion (BIC). For the definitions and usages of the information criteria, see the section “The Minimum Information Criterion (MINIC) Method” on page 3091.

If you do not want to use automatic order selection, the VARMAX procedure provides the following autoregressive order identification aids: partial cross-correlations, partial autoregressive coefficients, partial canonical correlations, and Yule-Walker estimates.

For situations where the stationarity of the time series is in question, the VARMAX procedure provides the following tests to aid in determining the presence of unit roots and cointegration: Dickey-Fuller tests, the Stock-Watson common trends test for the possibility of cointegration among nonstationary vector processes of integrated order one, and Johansen cointegration tests for nonstationary vector processes of integrated order one and order two.

For stationary vector times series or nonstationary series that are made stationary by appropriate differencing or cointegration, the VARMAX procedure provides the vector autoregressive and moving average (VARMA) model and the vector error correction model (VECM). The vector error correction model can be in both autoregressive (AR) and autoregressive and moving average (ARMA) forms.

To cope with the problem of high dimensionality in the parameters of the VAR model and the VECM, the VARMAX procedure provides both the Bayesian vector autoregressive (BVAR) model and the Bayesian vector error correction model (BVECM). Bayesian models are used when prior information about the model parameters is available.

The VARMAX procedure also allows independent (exogenous) variables and their distributed lags to influence dependent (endogenous) variables in various models. These models are identified by an X suffix added to the original model name; for example, VARMAX, VECMX, BVARX, and BVECMX.

Correlations in the second moments of the vector time series might exist; this is called conditional heteroscedasticity. The VARMAX procedure supports three forms of multivariate generalized autoregressive conditional heteroscedasticity (GARCH) models to model the conditional heteroscedasticity: the Baba-Engle-Kroner-Kraft (BEKK) GARCH model, the constant conditional correlation (CCC) GARCH model, and the dynamic conditional correlation (DCC) GARCH model. For CCC and DCC GARCH models, five subforms of univariate GARCH models are supported: the GARCH model, the exponential GARCH (EGARCH) model, the quadratic GARCH (QGARCH) model, the threshold GARCH (TGARCH) model, and the power GARCH (PGARCH) model.

You can use the VARMAX-GARCH model or the VEC-ARMAX-GARCH model to simultaneously model both the first and second moments of the time series.

Finally, for stationary time series exhibiting long-range dependence (also known as long memory or persistence), that is series with a slowly decaying sample autocorrelation function, the VARMAX procedure supports the VARFIMA (vector autoregressive fractionally integrated moving average) and VARFIMAX models.

Forecasting is one of the main objectives of multivariate time series analysis. After successfully fitting the VARMAX, BVARX, VECMX, BVECMX, VARFIMAX and multivariate GARCH models, the VARMAX procedure computes predicted values and conditional heteroscedasticity based on the parameter estimates and the past values of the vector time series. Out-of-sample multistep-ahead forecasts are also supported.

The following model parameter estimation methods are supported:

- the least squares (LS) method, which can be applied to VARX models
• the maximum likelihood (ML) method, which can be applied to all types of models and is used by default for VARFIMAX models,

• the conditional maximum likelihood (CML) method, which can be applied to VARMAX models

When you use the ML or CML method, you can start your optimization with the default or with different initial parameter values.

The VARMAX procedure supports the estimation of the restricted model when you impose linear constraints on the parameters of interest. The VARMAX procedure also supports various hypothesis tests of long-run effects and adjustment coefficients by using the likelihood ratio test that is based on Johansen cointegration analysis. It also supports the likelihood ratio test of weak exogeneity for each variable. In fact, because the VARMAX procedure outputs log-likelihood values for all models, you can always use the likelihood ratio test to check any linear hypothesis on parameters that are estimated in the models by estimating the restricted and unrestricted models separately. The VARMAX procedure also supports another alternative test, the Wald test.

After fitting the model parameters, the VARMAX procedure uses the following tests to provide model checks and residual analysis: Durbin-Watson (DW) statistics, the $F$ test for autoregressive conditional heteroscedastic (ARCH) disturbance, the $F$ test for AR disturbance, the Jarque-Bera normality test, and the portmanteau test.

The VARMAX procedure supports several modeling features, including seasonal deterministic terms, linear and quadratic time trends, subset models, multiple regression with distributed lags, the dead-start model (which does not have present values of the exogenous variables), and so on.

The VARMAX procedure provides a Granger causality test to determine the Granger-causal relationships between two distinct groups of variables. It also provides the following: the infinite order AR representation, the impulse response function (also called infinite order MA representation), the decomposition of the predicted error covariances, roots of the characteristic functions for both the AR and MA parts to evaluate the proximity of the roots to the unit circle, and contemporaneous relationships among the components of the vector time series.

---

**Getting Started: VARMAX Procedure**

This section provides several examples of the types of models that the VARMAX procedure supports.

**Vector Autoregressive Model**

Let $y_t = (y_{1t}, \ldots, y_{kt})'$, $t = 1, 2, \ldots$, denote a $k$-dimensional time series vector of random variables of interest. The $p$th-order VAR process is written as

$$y_t = \delta + \Phi_1 y_{t-1} + \cdots + \Phi_p y_{t-p} + \epsilon_t$$

where $\epsilon_t = (\epsilon_{1t}, \ldots, \epsilon_{kt})'$ is a vector white noise process such that $E(\epsilon_t) = 0$, $E(\epsilon_t \epsilon'_s) = \Sigma$, and $E(\epsilon_t \epsilon'_s) = 0$ for $t \neq s$; $\delta = (\delta_1, \ldots, \delta_k)'$ is a constant vector; and $\Phi_i$ is a $k \times k$ matrix.
Analyzing and modeling the series jointly enables you to understand the dynamic relationships over time among the series and to improve the accuracy of forecasts for individual series by using the additional information available from the related series and their forecasts.

Consider the first-order stationary bivariate vector autoregressive model:

\[ y_t = \begin{pmatrix} 1.2 & -0.5 \\ 0.6 & 0.3 \end{pmatrix} y_{t-1} + \epsilon_t, \quad \text{with } \Sigma = \begin{pmatrix} 1.0 & 0.5 \\ 0.5 & 1.25 \end{pmatrix} \]  

The following IML procedure statements simulate a bivariate vector time series from this model to provide test data for the VARMAX procedure:

```plaintext
proc iml;
   sig = {1.0 0.5, 0.5 1.25};
   phi = {1.2 -0.5, 0.6 0.3};
   /* simulate the vector time series */
   call varmasim(y,phi) sigma = sig n = 100 seed = 34657;
   cn = {'y1' 'y2'};
   create simull from y[colname=cn];
   append from y;
quit;
```

The following statements plot the simulated vector time series \( y_t \), which is shown in Figure 43.1:

```plaintext
data simull;
   set simull;
   date = intnx( 'year', '01jan1900'd, _n_-1 );
   format date year4.;
run;

proc sgplot data=simull;
   series x=date y=y1 / lineattrs=(pattern=solid);
   series x=date y=y2 / lineattrs=(pattern=dash);
   yaxis label="Series";
run;
```
The following statements fit a VAR(1) model to the simulated data:

```plaintext
/*--- Vector Autoregressive Model ---*/

proc varmax data=simul1;
   id date interval=year;
   model y1 y2 / p=1 noint lagmax=3
      print=(estimates diagnose);
   output out=for lead=5;
run;
```
First, you specify the input data set in the PROC VARMAX statement. Then, you use the MODEL statement to designate the dependent variables, $y_1$ and $y_2$. To estimate a zero-mean VAR model, you specify the order of the autoregressive model in the $P=$ option and the NOINT option. The MODEL statement fits the model to the data and prints parameter estimates and their significance. The PRINT=ESTIMATES option prints the matrix form of parameter estimates, and the PRINT=DIAGNOSE option prints various diagnostic tests. The LAGMAX=3 option prints the output for the residual diagnostic checks.

To output the forecasts to a data set, you specify the OUT= option in the OUTPUT statement. If you want to forecast five steps ahead, you use the LEAD=5 option. The ID statement specifies the yearly interval between observations and provides the Time column in the forecast output.

The VARMAX procedure output is shown in Figure 43.2 through Figure 43.10. The VARMAX procedure first displays descriptive statistics, as shown in Figure 43.2. The Type column indicates that the variables are dependent variables. The N column indicates the number of nonmissing observations.

**Figure 43.2** Descriptive Statistics

<table>
<thead>
<tr>
<th>The VARMAX Procedure</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of Observations 100</td>
</tr>
<tr>
<td>Number of Pairwise Missing 0</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Simple Summary Statistics</th>
</tr>
</thead>
<tbody>
<tr>
<td>variable</td>
</tr>
<tr>
<td>----------</td>
</tr>
<tr>
<td>y1</td>
</tr>
<tr>
<td>y2</td>
</tr>
</tbody>
</table>

**Figure 43.3** shows the model type and the estimation method that is used to fit the model to the simulated data. It also shows the AR coefficient matrix in terms of lag 1, the schematic representation, and the parameter estimates and their significance that can indicate how well the model fits the data.

The “AR” table shows the AR coefficient matrix. The “Schematic Representation” table schematically represents the parameter estimates and enables you to easily verify their significance in matrix form.

In the “Model Parameter Estimates” table, the first column shows the variable on the left side of the equation; the second column is the parameter name $AR_{i,j}$, which indicates the $(i,j)$ element of the lag $l$ autoregressive coefficient; the next four columns provide the estimate, standard error, $t$ value, and $p$-value for the parameter; and the last column is the regressor that corresponds to the displayed parameter.
The fitted VAR(1) model with estimated standard errors in parentheses is given as

\[
y_t = \begin{pmatrix} 1.160 & -0.511 \\ (0.055) & (0.059) \\ 0.546 & 0.385 \\ (0.058) & (0.062) \end{pmatrix} y_{t-1} + \epsilon_t
\]

Clearly, all parameter estimates in the coefficient matrix $\Phi_1$ are significant.

The model can also be written as two univariate regression equations:

\[
y_{1t} = 1.160 y_{1,t-1} - 0.511 y_{2,t-1} + \epsilon_{1t} \\
y_{2t} = 0.546 y_{1,t-1} + 0.385 y_{2,t-1} + \epsilon_{2t}
\]

The table in Figure 43.4 shows the innovation covariance matrix estimates, the log likelihood, and the various information criteria results. The variable names in the table for the innovation covariance matrix estimates $\Sigma$ are printed for convenience: $y_1$ means the innovation for $y_1$, and $y_2$ means the innovation.
for $y_2$. The log likelihood for a VAR model that is estimated through least squares method is defined as $-T(\log(\hat{\Sigma}_{ML}) + k)/2$, where $T(= 100 - 1 = 99)$ is the sample size except the presample being skipped because of the AR lag order, $k(= 2)$ is the number of dependent variables, and $\hat{\Sigma}_{ML}$ is the maximum likelihood estimate (MLE) of innovation covariance matrix. The matrix $\hat{\Sigma}_{ML}$ is computed from the reported least squares estimate of the innovation covariance matrix, $\hat{\Sigma}$, by adjusting the degrees of freedom. $\hat{\Sigma}_{ML} = \frac{T-r_b}{T} \hat{\Sigma}$, where $r_b(= 2)$ is the number of parameters in each equation. You can use the information criteria to compare the fit of competing models to a set of data. The model that has a smaller value of the information criterion is preferred when it is compared to other models. For more information about how to calculate the information criteria, see the section “Multivariate Model Diagnostic Checks” on page 3110.

**Figure 43.4** Innovation Covariance Estimates, Log Likelihood, and Information Criteria

<table>
<thead>
<tr>
<th>Covariances of Innovations</th>
<th>Variable</th>
<th>$y_1$</th>
<th>$y_2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$y_1$</td>
<td>1.28875</td>
<td>0.39751</td>
<td></td>
</tr>
<tr>
<td>$y_2$</td>
<td>0.39751</td>
<td>1.41839</td>
<td></td>
</tr>
</tbody>
</table>

Log-likelihood -122.362

<table>
<thead>
<tr>
<th>Information Criteria</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>AICC</td>
<td>259.9557</td>
</tr>
<tr>
<td>HQC</td>
<td>266.0748</td>
</tr>
<tr>
<td>AIC</td>
<td>258.7249</td>
</tr>
<tr>
<td>SBC</td>
<td>276.8908</td>
</tr>
<tr>
<td>FPEC</td>
<td>1.738092</td>
</tr>
</tbody>
</table>

Figure 43.5 shows the cross covariances of the residuals. The values of the lag 0 are slightly different from Figure 43.4 because of the different degrees of freedom.

**Figure 43.5** Multivariate Diagnostic Checks

<table>
<thead>
<tr>
<th>Cross Covariances of Residuals</th>
<th>Lag</th>
<th>Variable</th>
<th>$y_1$</th>
<th>$y_2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 y1</td>
<td>1.26271</td>
<td>0.38948</td>
<td></td>
<td></td>
</tr>
<tr>
<td>y2</td>
<td>0.38948</td>
<td>1.38974</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 y1</td>
<td>0.03121</td>
<td>0.05675</td>
<td></td>
<td></td>
</tr>
<tr>
<td>y2</td>
<td>-0.04646</td>
<td>-0.05398</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2 y1</td>
<td>0.08134</td>
<td>0.10599</td>
<td></td>
<td></td>
</tr>
<tr>
<td>y2</td>
<td>0.03482</td>
<td>-0.01549</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3 y1</td>
<td>0.01644</td>
<td>0.11734</td>
<td></td>
<td></td>
</tr>
<tr>
<td>y2</td>
<td>0.00609</td>
<td>0.11414</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Figure 43.6 and Figure 43.7 show tests for white noise residuals that are based on the cross correlations of the residuals. The output shows that you cannot reject the null hypothesis that the residuals are uncorrelated.
The VARMAX procedure provides diagnostic checks for the univariate form of the equations. The table in Figure 43.8 describes how well each univariate equation fits the data. From the two univariate regression equations shown in Figure 43.3, the values of $R^2$ in the second column of Figure 43.8 are 0.84 and 0.79. The standard deviations in the third column are the square roots of the diagonal elements of the covariance matrix from Figure 43.4. The $F$ statistics in the fourth column test the null hypotheses $H_{11} = H_{12} = 0$ and $H_{21} = H_{22} = 0$, where $H_{ij}$ is the $(i,j)$ element of the matrix $H_1$. The last column shows the $p$-values of the $F$ statistics. The results show that each univariate model is significant.

The check for white noise residuals in terms of the univariate equation is shown in Figure 43.9. This output contains information that indicates whether the residuals are correlated and heteroscedastic. In the first table, the second column contains the Durbin-Watson test statistics to test the null hypothesis that the residuals are
The Bayesian vector autoregressive (BVAR) model avoids problems of collinearity and overparameterization that often occur with the use of VAR models. BVAR models avoid these problems by imposing priors on the AR parameters.

The following statements fit a BVAR(1) model to the simulated data:

```plaintext
```
/*--- Bayesian Vector Autoregressive Process ---*/

proc varmax data=simul1;
  model y1 y2 / p=1 noint
       prior=(lambda=0.9 theta=0.1);
run;

The hyperparameters, LAMBDA=0.9 and THETA=0.1, in the PRIOR= option control the prior covariance. Part of the VARMAX procedure output is shown in Figure 43.11, whose parameter estimates are slightly different from those in Figure 43.3. By choosing the appropriate priors, you might be able to obtain more accurate forecasts by using a BVAR model instead of an unconstrained VAR model. For more information, see the section “Bayesian VAR and VARX Modeling” on page 3099.

Figure 43.11  Parameter Estimates for the BVAR(1) Model

<table>
<thead>
<tr>
<th>Type of Model</th>
<th>BVAR(1)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Estimation Method</td>
<td>Maximum Likelihood Estimation</td>
</tr>
<tr>
<td>Prior Lambda</td>
<td>0.9</td>
</tr>
<tr>
<td>Prior Theta</td>
<td>0.1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Model Parameter Estimates</th>
<th>Standard Error</th>
<th>t Value</th>
<th>Pr &gt;</th>
<th>Variable</th>
</tr>
</thead>
<tbody>
<tr>
<td>Equation</td>
<td>Parameter</td>
<td>Estimate</td>
<td></td>
<td></td>
</tr>
<tr>
<td>y1</td>
<td>AR1_1_1</td>
<td>1.05623</td>
<td>0.04999</td>
<td>21.13</td>
</tr>
<tr>
<td></td>
<td>AR1_1_2</td>
<td>-0.34707</td>
<td>0.04807</td>
<td>-7.22</td>
</tr>
<tr>
<td>y2</td>
<td>AR1_2_1</td>
<td>0.40068</td>
<td>0.04867</td>
<td>8.23</td>
</tr>
<tr>
<td></td>
<td>AR1_2_2</td>
<td>0.48728</td>
<td>0.05670</td>
<td>8.59</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Covariances of Innovations</th>
<th>y1</th>
<th>y2</th>
</tr>
</thead>
<tbody>
<tr>
<td>y1</td>
<td>1.36278</td>
<td>0.45343</td>
</tr>
<tr>
<td>y2</td>
<td>0.45343</td>
<td>1.48077</td>
</tr>
</tbody>
</table>

Vector Error Correction Model

A vector error correction model (VECM) can lead to a better understanding of the nature of any nonstationarity among the different component series and can also improve longer-term forecasting compared to an unconstrained model.

The VECM($p$) form with the cointegration rank, $r(\leq k)$, is written as

$$
\Delta y_t = \delta + \Pi y_{t-1} + \sum_{i=1}^{p-1} \Phi_i^* \Delta y_{t-i} + \epsilon_t
$$
where \( \Delta \) is the differencing operator, such that \( \Delta y_t = y_t - y_{t-1} \); \( \Pi = \alpha \beta' \), where \( \alpha \) and \( \beta \) are \( k \times r \) matrices; and \( \Phi_i^* \) is a \( k \times k \) matrix.

The VECM(\( p \)) form has an equivalent VAR(\( p \)) representation as described in the section “Vector Autoregressive Model” on page 2982.

\[
y_t = \delta + (I_k + \Pi + \Phi_1^*)y_{t-1} + \sum_{i=2}^{p-1} (\Phi_i^* - \Phi_{i-1}^*)y_{t-i} - \Phi_{p-1}^*y_{t-p} + \epsilon_t
\]

where \( I_k \) is a \( k \times k \) identity matrix.

An example of the second-order nonstationary vector autoregressive model is

\[
y_t = \begin{pmatrix} -0.2 & 0.1 \\ 0.5 & 0.2 \end{pmatrix} y_{t-1} + \begin{pmatrix} 0.8 & 0.7 \\ -0.4 & 0.6 \end{pmatrix} y_{t-2} + \epsilon_t
\]

with

\[
\Sigma = \begin{pmatrix} 100 & 0 \\ 0 & 100 \end{pmatrix} \text{ and } y_{-1} = y_0 = \begin{pmatrix} 0 \\ 0 \end{pmatrix}
\]

This process can be given the following VECM(2) representation with the cointegration rank one:

\[
\Delta y_t = \begin{pmatrix} -0.4 \\ 0.1 \end{pmatrix} (1, -2)y_{t-1} - \begin{pmatrix} 0.8 & 0.7 \\ -0.4 & 0.6 \end{pmatrix} \Delta y_{t-1} + \epsilon_t
\]

The following PROC IML statements generate simulated data for this VECM(2) form and the PROC SGPLOT statements plot the data, as shown in Figure 43.12:

```plaintext
proc iml;
   sig = 100*i(2);
   phi = {-0.2 0.1, 0.5 0.2, 0.8 0.7, -0.4 0.6};
   call varmasim(y,phi) sigma=sig n=100 initial=0
      seed=45876;
   cn = {'y1' 'y2'};
   create simul2 from y[colname=cn];
   append from y;
   quit;

data simul2;
```
Cointegration Testing

The following statements use the Johansen cointegration rank test. The COINTTEST=(JOHANSEN) option performs the Johansen trace test and is equivalent to specifying the COINTTEST option with no additional suboptions or specifying the COINTTEST=(JOHANSEN=(TYPE=TRACE)) option.
/*--- Cointegration Test ---*/

proc varmax data=simul2;
   model y1 y2 / p=2 noint dftest cointtest=(johansen);
run;

Figure 43.13 shows the output for Dickey-Fuller tests for the nonstationarity of each series and the Johansen cointegration rank test between series.

### Figure 43.13 Dickey-Fuller Tests and Cointegration Rank Test

#### The VARMAX Procedure

<table>
<thead>
<tr>
<th>Variable</th>
<th>Type</th>
<th>Rho</th>
<th>Pr &lt; Rho</th>
<th>Tau</th>
<th>Pr &lt; Tau</th>
</tr>
</thead>
<tbody>
<tr>
<td>y1</td>
<td>Zero Mean</td>
<td>1.47</td>
<td>0.9628</td>
<td>1.65</td>
<td>0.9755</td>
</tr>
<tr>
<td></td>
<td>Single Mean</td>
<td>-0.80</td>
<td>0.9016</td>
<td>-0.47</td>
<td>0.8916</td>
</tr>
<tr>
<td></td>
<td>Trend</td>
<td>-10.88</td>
<td>0.3573</td>
<td>-2.20</td>
<td>0.4815</td>
</tr>
<tr>
<td>y2</td>
<td>Zero Mean</td>
<td>-0.05</td>
<td>0.6692</td>
<td>-0.03</td>
<td>0.6707</td>
</tr>
<tr>
<td></td>
<td>Single Mean</td>
<td>-6.03</td>
<td>0.3358</td>
<td>-1.72</td>
<td>0.4204</td>
</tr>
<tr>
<td></td>
<td>Trend</td>
<td>-50.49</td>
<td>0.0003</td>
<td>-4.92</td>
<td>0.0006</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>H0: Rank=0</th>
<th>H1: Rank&gt;0</th>
<th>Eigenvalue</th>
<th>Trace</th>
<th>Pr &gt; Trace</th>
<th>Drift in ECM</th>
<th>Drift in Process</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>0.5086</td>
<td>70.7279</td>
<td>&lt;.0001</td>
<td>NOINT</td>
<td>Constant</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>0.0111</td>
<td>1.0921</td>
<td>0.3441</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

In Dickey-Fuller tests, the second column specifies three types of models, which are zero mean, single mean, or trend. The third column (Rho) and the fifth column (Tau) are the test statistics that are used to test the null hypothesis that the series has a unit root. Other columns are their p-values. You can see that both series have unit roots. For a description of Dickey-Fuller tests, see the section “PROBDF Function for Dickey-Fuller Tests” on page 167 in Chapter 5, “SAS Macros and Functions.”

In the “Cointegration Rank Test Using Trace” table, the last two columns explain the drift in the model or process. Because the NOINT option is specified, the model is

\[
\Delta y_t = \Pi y_{t-1} + \Phi^*_1 \Delta y_{t-1} + \epsilon_t
\]

The column Drift in ECM indicates that there is no separate drift in the error correction model, and the column Drift in Process indicates that the process has a constant drift before differencing.

H0 is the null hypothesis, and H1 is the alternative hypothesis. The first row tests the cointegration rank \( r = 0 \) against \( r > 0 \), and the second row tests \( r = 1 \) against \( r > 1 \). The trace test statistics in the fourth column are computed by \(-T \sum_{i=r+1}^{k} \log(1 - \lambda_i)\), where \( T \) is the available number of observations and \( \lambda_i \) is the eigenvalue in the third column. The p-values for these statistics are output in the fifth column. If you compare the p-value in each row to the significance level of interest (such as 5%), the null hypothesis that there is no cointegrated process (H0: \( r = 0 \)) is rejected, whereas the null hypothesis that there is at most one cointegrated process (H0: \( r = 1 \)) cannot be rejected.

The following statements fit a VECM(2) form to the simulated data:
The VARMAX procedure output is shown in Figure 43.14 through Figure 43.17. In Figure 43.14, “1” indicates the first column of the $\alpha$ and $\beta$ matrices. Because the cointegration rank is 1 in the bivariate system, $\alpha$ and $\beta$ are two-dimensional vectors. The estimated cointegrating vector is $\hat{\beta} = (1, -1.96)'$. Therefore, the long-run relationship between $y_{1t}$ and $y_{2t}$ is $y_{1t} = 1.96y_{2t}$. The first element of $\hat{\beta}$ is 1 because $y_1$ is specified as the normalized variable. Asymptotically, $\alpha$ follows a normal distribution, and the $t$ values and $p$-values of its elements are shown in the “Alpha and Beta Parameter Estimates” table; however, because $\beta$ follows a nonnormal distribution, the corresponding standard errors, $t$ values, and $p$-values are missing. The Variable column shows the variables that correspond to the coefficients. For example, for the coefficient $\alpha_{ij}$ (the $i$th element in the $j$th column of $\alpha$), $\text{ALPHA}_{i-j}$, the variable is the inner product of the transpose of the $j$th column of $\beta$ ($\text{Beta}_{j'}$) and the vector of lag 1 dependent variables $y_{t-1}$. 

**Figure 43.14** Parameter Estimates for the VECM(2) Form

<table>
<thead>
<tr>
<th>Type of Model</th>
<th>VECM(2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Estimation Method</td>
<td>Maximum Likelihood Estimation</td>
</tr>
<tr>
<td>Cointegrated Rank</td>
<td>1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Beta</th>
</tr>
</thead>
<tbody>
<tr>
<td>Variable</td>
</tr>
<tr>
<td>$y_1$</td>
</tr>
<tr>
<td>$y_2$</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Alpha</th>
</tr>
</thead>
<tbody>
<tr>
<td>Variable</td>
</tr>
<tr>
<td>$y_1$</td>
</tr>
<tr>
<td>$y_2$</td>
</tr>
</tbody>
</table>

**Alpha and Beta Parameter Estimates**

<table>
<thead>
<tr>
<th>Equation</th>
<th>Parameter</th>
<th>Estimate</th>
<th>Standard Error</th>
<th>$t$ Value</th>
<th>Pr &gt;</th>
<th>Variable</th>
</tr>
</thead>
<tbody>
<tr>
<td>$D_y1$</td>
<td>ALPHAR1</td>
<td>-0.46680</td>
<td>0.04786</td>
<td>-9.75</td>
<td>&lt;0.001</td>
<td>Beta[1]_* <em>DEP</em>(t-1)</td>
</tr>
<tr>
<td></td>
<td>BETA1</td>
<td>1.00000</td>
<td></td>
<td></td>
<td></td>
<td>y1(t-1)</td>
</tr>
<tr>
<td>$D_y2$</td>
<td>ALPHAR2</td>
<td>0.10667</td>
<td>0.05146</td>
<td>2.07</td>
<td>0.0409</td>
<td>Beta[1]_* <em>DEP</em>(t-1)</td>
</tr>
<tr>
<td></td>
<td>BETA2</td>
<td>-1.95575</td>
<td></td>
<td></td>
<td></td>
<td>y2(t-1)</td>
</tr>
</tbody>
</table>
Figure 43.15 shows the parameter estimates in terms of lag 1 coefficients, $y_{t-1}$, and lag 1 first-differenced coefficients, $\Delta y_{t-1}$, and their significance. “Alpha * Beta” indicates the coefficients of $y_{t-1}$ and is obtained by multiplying the Alpha and Beta estimates in Figure 43.14. The parameter $AR_{1-i-j}$ (which is shown in the “Model Parameter Estimates” table) corresponds to the elements in the “Alpha * Beta” matrix. The parameter $AR_{2-i-j}$ corresponds to the elements in the differenced lagged AR coefficient matrix. The “D_” prefixed to a variable name in Figure 43.15 implies differencing.

**Figure 43.15** Parameter Estimates for the VECM(2) Form, Continued

<table>
<thead>
<tr>
<th>Parameter Alpha * Beta Estimates</th>
<th>y1</th>
<th>y2</th>
</tr>
</thead>
<tbody>
<tr>
<td>y1</td>
<td>-0.46680</td>
<td>0.91295</td>
</tr>
<tr>
<td>y2</td>
<td>0.10667</td>
<td>-0.20862</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>AR Coefficients of Differenced Lag</th>
<th>y1</th>
<th>y2</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 y1</td>
<td>-0.74332</td>
<td>-0.74621</td>
</tr>
<tr>
<td>2 y2</td>
<td>0.40493</td>
<td>-0.57157</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Model Parameter Estimates</th>
<th>Standard Error</th>
<th>t Value</th>
<th>Pr &gt;</th>
<th>Variable</th>
</tr>
</thead>
<tbody>
<tr>
<td>$D_y1$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AR1_1_1</td>
<td>-0.46680</td>
<td>-9.75</td>
<td>&lt;.0001</td>
<td>$y_1(t-1)$</td>
</tr>
<tr>
<td>AR1_1_2</td>
<td>0.91295</td>
<td>9.75</td>
<td>&lt;.0001</td>
<td>$y_2(t-1)$</td>
</tr>
<tr>
<td>AR2_1_1</td>
<td>-0.74332</td>
<td>-16.42</td>
<td>&lt;.0001</td>
<td>$D_y1(t-1)$</td>
</tr>
<tr>
<td>AR2_1_2</td>
<td>-0.74621</td>
<td>-15.65</td>
<td>&lt;.0001</td>
<td>$D_y2(t-1)$</td>
</tr>
<tr>
<td>$D_y2$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AR1_2_1</td>
<td>0.10667</td>
<td>2.07</td>
<td>0.0409</td>
<td>$y_1(t-1)$</td>
</tr>
<tr>
<td>AR1_2_2</td>
<td>-0.20862</td>
<td>-2.07</td>
<td>0.0409</td>
<td>$y_2(t-1)$</td>
</tr>
<tr>
<td>AR2_2_1</td>
<td>0.40493</td>
<td>8.32</td>
<td>&lt;.0001</td>
<td>$D_y1(t-1)$</td>
</tr>
<tr>
<td>AR2_2_2</td>
<td>-0.57157</td>
<td>-11.15</td>
<td>&lt;.0001</td>
<td>$D_y2(t-1)$</td>
</tr>
</tbody>
</table>

Figure 43.16 shows the parameter estimates of the innovations covariance matrix and their significance.

**Figure 43.16** Parameter Estimates for the VECM(2) Form, Continued

<table>
<thead>
<tr>
<th>Covariance Parameter Estimates</th>
<th>Standard Error</th>
<th>t Value</th>
<th>Pr &gt;</th>
<th>Variable</th>
</tr>
</thead>
<tbody>
<tr>
<td>COV1_1</td>
<td>94.75575</td>
<td>7.00</td>
<td>&lt;.0001</td>
<td></td>
</tr>
<tr>
<td>COV1_2</td>
<td>4.52684</td>
<td>0.44</td>
<td>0.6614</td>
<td></td>
</tr>
<tr>
<td>COV2_2</td>
<td>109.57038</td>
<td>7.00</td>
<td>&lt;.0001</td>
<td></td>
</tr>
</tbody>
</table>
The fitted model is represented as

\[
\Delta y_t = \begin{pmatrix}
-0.467 & 0.913 \\
0.048 & 0.094 \\
0.107 & -0.209 \\
0.051 & 0.100
\end{pmatrix}
\begin{pmatrix}
(0.048) & (0.094) \\
(0.107) & (-0.209) \\
(0.051) & (0.100)
\end{pmatrix}
\begin{pmatrix}
y_{t-1} \\
y_t
\end{pmatrix}
+ \begin{pmatrix}
-0.743 & -0.746 \\
0.045 & 0.048 \\
0.405 & -0.572 \\
0.049 & (0.051)
\end{pmatrix}
\begin{pmatrix}
\Delta y_{t-1} \\
\epsilon_t
\end{pmatrix}
\]

Figure 43.17 Change the VECM(2) Form to the VAR(2) Model

<table>
<thead>
<tr>
<th>Lag</th>
<th>Variable</th>
<th>(y_1)</th>
<th>(y_2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>(y_1)</td>
<td>-0.21013</td>
<td>0.16674</td>
</tr>
<tr>
<td>1</td>
<td>(y_2)</td>
<td>0.51160</td>
<td>0.21980</td>
</tr>
<tr>
<td>2</td>
<td>(y_1)</td>
<td>0.74332</td>
<td>0.74621</td>
</tr>
<tr>
<td>2</td>
<td>(y_2)</td>
<td>-0.40493</td>
<td>0.57157</td>
</tr>
<tr>
<td>3</td>
<td>(y_1)</td>
<td>0.00000</td>
<td>0.00000</td>
</tr>
<tr>
<td>3</td>
<td>(y_2)</td>
<td>0.00000</td>
<td>0.00000</td>
</tr>
</tbody>
</table>

The PRINT=(IARR) option in the previous SAS statements prints the reparameterized coefficient estimates. Because LAGMAX=3 in those statements, the coefficient matrix of lag 3 is zero.

The VECM(2) form in Figure 43.17 can be rewritten as the following second-order vector autoregressive model:

\[
y_t = \begin{pmatrix}
-0.210 & 0.167 \\
0.512 & 0.220
\end{pmatrix}
\begin{pmatrix}
y_{t-1} \\
y_{t-2}
\end{pmatrix}
+ \begin{pmatrix}
0.743 & 0.746 \\
-0.405 & 0.572
\end{pmatrix}
\begin{pmatrix}
y_{t-2} \\
y_{t-3}
\end{pmatrix}
+ \epsilon_t
\]

Bayesian Vector Error Correction Model

Bayesian inference on a cointegrated system begins by using the priors of \(\beta\), which are obtained from the VECM(\(p\)) form. Bayesian vector error correction models can improve forecast accuracy for cointegrated processes.

To use a Bayesian vector error correction model, you specify both the PRIOR= option in the MODEL statement and the COINTEG statement. The following statements fit a BVECM(2) form to the simulated data:

```sas
/*--- Bayesian Vector Error Correction Model ---*/
proc varmax data=simul2;
  model y1 y2 / p=2 noint
        prior=( lambda=0.5 theta=0.2 )
        print=(estimates);
  cointeg rank=1 normalize=y1;
run;
```
The VARMAX procedure output in Figure 43.18 shows the model type fitted to the data, the estimates of the adjustment coefficient (α), the parameter estimates in terms of lag 1 coefficients (y_{t-1}), and lag 1 first-differenced coefficients (∆y_{t-1}).

**Figure 43.18** Parameter Estimates for the BVECM(2) Form

**The VARMAX Procedure**

<table>
<thead>
<tr>
<th>Type of Model</th>
<th>BVECM(2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Estimation Method</td>
<td>Maximum Likelihood Estimation</td>
</tr>
<tr>
<td>Cointegrated Rank</td>
<td>1</td>
</tr>
<tr>
<td>Prior Lambda</td>
<td>0.5</td>
</tr>
<tr>
<td>Prior Theta</td>
<td>0.2</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Alpha Variable</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>y1</td>
<td>-0.34392</td>
</tr>
<tr>
<td>y2</td>
<td>0.16659</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Parameter Alpha + Beta Estimates</th>
</tr>
</thead>
<tbody>
<tr>
<td>Variable</td>
</tr>
<tr>
<td>y1</td>
</tr>
<tr>
<td>y2</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>AR Coefficients of Differenced Lag</th>
</tr>
</thead>
<tbody>
<tr>
<td>DIF Lag</td>
</tr>
<tr>
<td>1 y1</td>
</tr>
<tr>
<td>y2</td>
</tr>
</tbody>
</table>

**Vector Autoregressive Fractionally Integrated Moving Average Model**

Fractionally integrated models can be used to model stationary time series whose sample autocorrelation function decays slowly at large positive and negative lags. This behavior is often referred to as long-range dependence (LRD), long memory, or persistence; series that exhibit such behavior are called long-range dependent (LRD).

A typical parametric model for a k-dimensional series \( y_t = (y_{1t}, \ldots, y_{kt})' \), \( t = 1, \ldots, T \), whose individual components are LRD is the VARFIMA (vector autoregressive fractionally integrated moving average) model. It is obtained as a natural extension of the well-known class of ARFIMA models by fractionally integrating the individual components of a k-dimensional white noise series. For example, a bivariate VARFIMA(0, D, 0) series with no intercept term is given by

\[
    y_t = \begin{pmatrix} y_{1t} \\ y_{2t} \end{pmatrix} = \begin{pmatrix} (I - B)^{-d_1} \\ 0 \end{pmatrix} \begin{pmatrix} 0 \\ (I - B)^{-d_2} \end{pmatrix} \begin{pmatrix} \epsilon_{1t} \\ \epsilon_{2t} \end{pmatrix} = (I - B)^{-D} \epsilon_t
\]

where \( B \) is the backshift operator; \( I = B^0 \) is the identity operator; \( d_1, d_2 \in (-1/2, 1/2) \) are the LRD parameters of the component series \( \{y_{1t}\}_{t \in \mathbb{Z}} \) and \( \{y_{2t}\}_{t \in \mathbb{Z}} \), respectively; \( D = \text{diag}(d_1, d_2) \); and \( \{\epsilon_t\}_{t \in \mathbb{Z}} = \ldots \)
Chapter 43: The VARMAX Procedure

\{(ε_{1t}, ε_{2t})\}_{t \in \mathbb{Z}} is a bivariate white noise series indexed by the set of integers \(\mathbb{Z}\) with zero mean \(\mathbb{E}\epsilon_t = 0\) and covariance \(\mathbb{E}\epsilon_t\epsilon'_t = \Sigma\).

The multivariate VARFIMA model is defined analogously. The matrix \(\Sigma\) is in general nondiagonal, which enables the VARFIMA model to capture dependence between the individual series.

The following statements plot a simulated bivariate VARFIMA(0, D, 0) series with \(d_1 = 0.2, d_2 = 0.4\), and Gaussian errors with \(\Sigma_{11} = \Sigma_{22} = 3\) and \(\Sigma_{12} = 0.5\):

```plaintext
data VARFIMA0D0;
  time = _N_;  
  input y1 y2;
  datalines;
  1.6380971 1.877144 
  ... more lines ...
  0.3482938 4.8601886  
  1.5320803 2.8687495
;

proc sgplot data = VARFIMA0D0;
  series x = time y=y1 / lineattrs=(pattern=solid);
  series x = time y=y2 / lineattrs=(pattern=dash);
  yaxis label="Series";
run;
```
Before fitting a VARFIMA model to a data set, you should plot the series’ sample autocorrelation function to confirm its slow decay. It is also instructive to plot the periodogram of the series. In the presence of long memory, the periodogram explodes at frequencies near 0.

The following statements produce the periodogram and the sample autocorrelation function for the specified data:

```latex
ods graphics on;
proc timeseries data=VARFIMA0D0 plots = (periodogram acf);
   var y1 y2;
   spectra freq / adjmean;
   corr / NLAG = 30;
run;
```
The magnitude of the LRD parameters $d_1$ and $d_2$ controls the memory of the two series. Series $y_2$ has a larger LRD parameter than series $y_1$ and hence is expected to exhibit longer memory. In the time domain, this effect is illustrated in Figure 43.20, where the autocorrelation function of series $y_2$ (right plot in Figure 43.20) decays more slowly than the autocorrelation function of series $y_1$ (left plot in Figure 43.20) with the increasing lag.

Figure 43.21 is the frequency domain analogue of Figure 43.20. In this case, the longer memory of series $y_2$ is reflected by its periodogram (right plot in Figure 43.21), which blows up higher than the periodogram of series $y_1$ (left plot in Figure 43.21) at frequencies near 0. Note the different scales used in the two plots.

The following statements fit the VARFIMA($0, D, 0$) model with no intercept term to the data. The FI option in the MODEL statement specifies fractional integration.

```plaintext
proc varmax data = VARFIMA0D0;
  model y1 y2 / fi noint method = ML;
run;
```
The estimation method that PROC VARMAX uses by default for the VARFIMA series is maximum likelihood (for more information, see the section “VARFIMA and VARFIMAX Modeling” on page 3146). All five parameter are estimated close to their true value and are significant.

The VARMAX Procedure

<table>
<thead>
<tr>
<th>Type of Model</th>
<th>VARFIMA(0,D,0)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Estimation Method</td>
<td>Maximum Likelihood Estimation</td>
</tr>
</tbody>
</table>

Model Parameter Estimates

<table>
<thead>
<tr>
<th>Equation</th>
<th>Parameter</th>
<th>Estimate</th>
<th>Standard Error</th>
<th>t Value</th>
<th>Pr &gt;</th>
<th>Variable</th>
</tr>
</thead>
<tbody>
<tr>
<td>y1</td>
<td>D1</td>
<td>0.20250</td>
<td>0.03555</td>
<td>5.70</td>
<td>0.0001</td>
<td></td>
</tr>
<tr>
<td>y2</td>
<td>D2</td>
<td>0.38839</td>
<td>0.03053</td>
<td>12.72</td>
<td>0.0001</td>
<td></td>
</tr>
</tbody>
</table>

Covariances of Innovations

<table>
<thead>
<tr>
<th>Variable</th>
<th>y1</th>
<th>y2</th>
</tr>
</thead>
<tbody>
<tr>
<td>y1</td>
<td>3.20607</td>
<td>0.48068</td>
</tr>
<tr>
<td>y2</td>
<td>0.48068</td>
<td>3.15651</td>
</tr>
</tbody>
</table>

The following statements fit the VARX(1,0) model to the given data:

Vector Autoregressive Model with Exogenous Variables

A VAR process can be affected by other observable variables that are determined outside the system of interest. Such variables are called exogenous (independent) variables. Exogenous variables can be stochastic or nonstochastic. The process can also be affected by the lags of exogenous variables. A model used to describe this process is called a VARX(p,s) model.

The VARX(p,s) model is written as

\[ y_t = \delta + \sum_{i=1}^{p} \Phi_i y_{t-i} + \sum_{i=0}^{s} \Theta_i^* x_{t-i} + \epsilon_t \]

where \( x_t = (x_{1t}, \ldots, x_{rt})' \) is an \( r \)-dimensional time series vector and \( \Theta_i^* \) is a \( k \times r \) matrix.

For example, a VARX(1,0) model is

\[ y_t = \delta + \Phi_1 y_{t-1} + \Theta_0^* x_t + \epsilon_t \]

where \( y_t = (y_{1t}, y_{2t}, y_{3t})' \) and \( x_t = (x_{1t}, x_{2t})' \).

The following statements fit the VARX(1,0) model to the given data:
data grunfeld;
    input year y1 y2 y3 x1 x2 x3;
    label y1='Gross Investment GE'
         y2='Capital Stock Lagged GE'
         y3='Value of Outstanding Shares GE Lagged'
         x1='Gross Investment W'
         x2='Capital Stock Lagged W'
         x3='Value of Outstanding Shares Lagged W';
datalines;
1935 33.1 1170.6 97.8 12.93 191.5 1.8
1936 45.0 2015.8 104.4 25.90 516.0 .8
1937 77.2 2803.3 118.0 35.05 729.0 7.4
1938 44.6 2039.7 156.2 22.89 560.4 18.1
... more lines ...

    /*---- Vector Autoregressive Process with Exogenous Variables ----*/

proc varmax data=grunfeld;
    model y1-y3 = x1 x2 / p=1 lagmax=5
         printform=univariate
         print=(impulsx=(all) estimates);
run;
The VARMAX procedure output is shown in Figure 43.23 through Figure 43.25.

Figure 43.23 shows the descriptive statistics for the dependent (endogenous) and independent (exogenous) variables with labels.

Figure 43.23  Descriptive Statistics for the VARX(1, 0) Model

The VARMAX Procedure

<table>
<thead>
<tr>
<th>Simple Summary Statistics</th>
</tr>
</thead>
<tbody>
<tr>
<td>Variable</td>
</tr>
<tr>
<td>----------</td>
</tr>
<tr>
<td>y1</td>
</tr>
<tr>
<td>y2</td>
</tr>
<tr>
<td>y3</td>
</tr>
<tr>
<td>x1</td>
</tr>
<tr>
<td>x2</td>
</tr>
</tbody>
</table>
Figure 43.24 shows the parameter estimates for the constant, the lag zero coefficients of exogenous variables, and the lag one AR coefficients. From the schematic representation of parameter estimates, the significance of the parameter estimates can be easily verified. The symbol “C” means the constant and “XL0” means the lag zero coefficients of exogenous variables.

**Figure 43.24** Parameter Estimates for the VARX(1, 0) Model

### The VARMAX Procedure

<table>
<thead>
<tr>
<th>Type of Model</th>
<th>VARX(1,0)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Estimation Method</td>
<td>Least Squares Estimation</td>
</tr>
</tbody>
</table>

#### Constant

<table>
<thead>
<tr>
<th>Variable</th>
<th>Constant</th>
</tr>
</thead>
<tbody>
<tr>
<td>y1</td>
<td>-12.01279</td>
</tr>
<tr>
<td>y2</td>
<td>702.08673</td>
</tr>
<tr>
<td>y3</td>
<td>-22.42110</td>
</tr>
</tbody>
</table>

#### XLag

<table>
<thead>
<tr>
<th>Lag</th>
<th>Variable</th>
<th>x1</th>
<th>x2</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>y1</td>
<td>1.69281</td>
<td>-0.00859</td>
</tr>
<tr>
<td></td>
<td>y2</td>
<td>-6.09850</td>
<td>2.57980</td>
</tr>
<tr>
<td></td>
<td>y3</td>
<td>-0.02317</td>
<td>-0.01274</td>
</tr>
</tbody>
</table>

#### AR

<table>
<thead>
<tr>
<th>Lag</th>
<th>Variable</th>
<th>y1</th>
<th>y2</th>
<th>y3</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>y1</td>
<td>0.23699</td>
<td>0.00763</td>
<td>0.02941</td>
</tr>
<tr>
<td></td>
<td>y2</td>
<td>-2.46656</td>
<td>0.16379</td>
<td>-0.84090</td>
</tr>
<tr>
<td></td>
<td>y3</td>
<td>0.95116</td>
<td>0.00224</td>
<td>0.93801</td>
</tr>
</tbody>
</table>

#### Schematic Representation

<table>
<thead>
<tr>
<th>Variable/Lag</th>
<th>C</th>
<th>XL0</th>
<th>AR1</th>
</tr>
</thead>
<tbody>
<tr>
<td>y1</td>
<td>.</td>
<td>.</td>
<td>..</td>
</tr>
<tr>
<td>y2</td>
<td>+</td>
<td>.</td>
<td>..</td>
</tr>
<tr>
<td>y3</td>
<td>-</td>
<td>..</td>
<td>++</td>
</tr>
</tbody>
</table>

+ is > 2*std error, - is < -2*std error, . is between, * is N/A
Figure 43.25 shows the parameter estimates and their significance.

**Figure 43.25 Parameter Estimates for the VARX(1, 0) Model Continued**

| Equation | Parameter | Estimate  | Standard Error | t Value | Pr > |t| | Variable |
|----------|-----------|-----------|----------------|---------|-------|---|---------|
| y1       | CONST1    | -12.01279 | 27.47108       | -0.44   | 0.6691| 1  |         |
|          | XLO_1_1   | 1.69281   | 0.54395        | 3.11    | 0.0083| x1(t)|         |
|          | XLO_1_2   | -0.00859  | 0.05361        | -0.16   | 0.8752| x2(t)|         |
|          | AR1_1_1   | 0.23699   | 0.20668        | 1.15    | 0.2722| y1(t-1)|       |
|          | AR1_1_2   | 0.00763   | 0.01627        | 0.47    | 0.6470| y2(t-1)|       |
|          | AR1_1_3   | 0.02941   | 0.04852        | 0.61    | 0.5548| y3(t-1)|       |
| y2       | CONST2    | 702.08673 | 256.48046      | 2.74    | 0.0169| 1  |         |
|          | XLO_2_1   | -6.09850  | 5.07849        | -1.20   | 0.2512| x1(t)|         |
|          | XLO_2_2   | 2.57980   | 0.50056        | 5.15    | 0.0002| x2(t)|         |
|          | AR1_2_1   | -2.46656  | 1.92967        | -1.28   | 0.2235| y1(t-1)|       |
|          | AR1_2_2   | 0.16379   | 0.15193        | 1.08    | 0.3006| y2(t-1)|       |
|          | AR1_2_3   | -0.84090  | 0.45304        | -1.86   | 0.0862| y3(t-1)|       |
| y3       | CONST3    | -22.42110 | 10.31166       | -2.17   | 0.0487| 1  |         |
|          | XLO_3_1   | -0.02317  | 0.20418        | -0.11   | 0.9114| x1(t)|         |
|          | XLO_3_2   | -0.01274  | 0.02012        | -0.63   | 0.5377| x2(t)|         |
|          | AR1_3_1   | 0.95116   | 0.07758        | 12.26   | 0.0001| y1(t-1)|       |
|          | AR1_3_2   | 0.00224   | 0.00611        | 0.37    | 0.7201| y2(t-1)|       |
|          | AR1_3_3   | 0.93801   | 0.01821        | 51.50   | 0.0001| y3(t-1)|       |

The fitted model is given as

\[
\begin{pmatrix}
  y_{1t} \\
  y_{2t} \\
  y_{3t}
\end{pmatrix}
= \begin{pmatrix}
-12.013 \\
27.471
\end{pmatrix}
+ \begin{pmatrix}
1.693 & -0.009 \\
(0.544) & (0.054)
\end{pmatrix}
\begin{pmatrix}
x_{1t} \\
x_{2t}
\end{pmatrix}
+ \begin{pmatrix}
0.237 & 0.008 & 0.029 \\
(0.207) & (0.016) & (0.049)
\end{pmatrix}
\begin{pmatrix}
y_{1,t-1} \\
y_{2,t-1} \\
y_{3,t-1}
\end{pmatrix}
+ \begin{pmatrix}
-2.467 & 0.164 & -0.841 \\
(1.930) & (0.152) & (0.453)
\end{pmatrix}
\begin{pmatrix}
\epsilon_{1t} \\
\epsilon_{2t} \\
\epsilon_{3t}
\end{pmatrix}
+ \begin{pmatrix}
0.951 \\
0.002 \\
(0.078)
\end{pmatrix}
\begin{pmatrix}
y_{1,t-1} \\
y_{2,t-1} \\
y_{3,t-1}
\end{pmatrix}
+ \begin{pmatrix}
0.93801 \\
0.01821
\end{pmatrix}
\begin{pmatrix}
y_{1,t-1} \\
y_{2,t-1} \\
y_{3,t-1}
\end{pmatrix}
\]

**Parameter Estimation and Testing on Restrictions**

In the previous example, the VARX(1,0) model is written as

\[
y_t = \delta + \Theta_0' x_t + \Phi_1 y_{t-1} + \epsilon_t
\]
with

$$
\Theta_0^* = \begin{pmatrix}
\theta_{11}^* & \theta_{12}^* \\
\theta_{21}^* & \theta_{22}^* \\
\theta_{31}^* & \theta_{32}^*
\end{pmatrix}
\quad\Phi_1 = \begin{pmatrix}
\phi_{11} & \phi_{12} & \phi_{13} \\
\phi_{21} & \phi_{22} & \phi_{23} \\
\phi_{31} & \phi_{32} & \phi_{33}
\end{pmatrix}
$$

In Figure 43.25 of the preceding section, you can see several insignificant parameters. For example, the coefficients \(XL0_1\_2\), \(AR1_1\_2\), and \(AR1_3\_2\) are insignificant.

The following statements restrict the coefficients of \(\theta_{12}^* = \phi_{12} = \phi_{32} = 0\) for the VARX(1,0) model:

```plaintext
/*--- Models with Restrictions and Tests ---*/
proc varmax data=grunfeld;
  model y1-y3 = x1 x2 / p=1 print=(estimates);
  restrict XL(0,1,2)=0, AR(1,1,2)=0, AR(1,3,2)=0;
run;
```

The output in Figure 43.26 shows that three parameters \(\theta_{12}^*, \phi_{12}, \text{and } \phi_{32}\) are replaced by the restricted values, zeros, and their standard errors are also zeros to indicate that the parameters are fixed to these values.

**Figure 43.26** Parameter Estimation with Restrictions

**The VARMAX Procedure**

<table>
<thead>
<tr>
<th>Equation</th>
<th>Parameter</th>
<th>Estimate</th>
<th>Standard Error</th>
<th>t Value</th>
<th>Pr &gt;</th>
<th>Variable</th>
</tr>
</thead>
<tbody>
<tr>
<td>y1</td>
<td>CONST1</td>
<td>-2.16781</td>
<td>13.13755</td>
<td>-0.17</td>
<td>0.8715</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>XL0_1_1</td>
<td>1.67592</td>
<td>0.40792</td>
<td>4.11</td>
<td>0.0012</td>
<td>x1(t)</td>
</tr>
<tr>
<td></td>
<td>XL0_1_2</td>
<td>-0.00000</td>
<td>0.00000</td>
<td>x2(t)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>AR1_1_1</td>
<td>0.27671</td>
<td>0.17606</td>
<td>1.57</td>
<td>0.1401</td>
<td>y1(t-1)</td>
</tr>
<tr>
<td></td>
<td>AR1_1_2</td>
<td>-0.00000</td>
<td>0.00000</td>
<td>y2(t-1)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>AR1_1_3</td>
<td>0.01747</td>
<td>0.03519</td>
<td>0.50</td>
<td>0.6279</td>
<td>y3(t-1)</td>
</tr>
<tr>
<td>y2</td>
<td>CONST2</td>
<td>768.14598</td>
<td>224.12735</td>
<td>3.43</td>
<td>0.0045</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>XL0_2_1</td>
<td>-6.30880</td>
<td>4.85729</td>
<td>-1.30</td>
<td>0.2166</td>
<td>x1(t)</td>
</tr>
<tr>
<td></td>
<td>XL0_2_2</td>
<td>2.65308</td>
<td>0.43840</td>
<td>6.05</td>
<td>0.0001</td>
<td>x2(t)</td>
</tr>
<tr>
<td></td>
<td>AR1_2_1</td>
<td>-2.16968</td>
<td>1.83550</td>
<td>-1.18</td>
<td>0.2584</td>
<td>y1(t-1)</td>
</tr>
<tr>
<td></td>
<td>AR1_2_2</td>
<td>0.10945</td>
<td>0.11751</td>
<td>0.93</td>
<td>0.3686</td>
<td>y2(t-1)</td>
</tr>
<tr>
<td></td>
<td>AR1_2_3</td>
<td>-0.93053</td>
<td>0.41478</td>
<td>-2.24</td>
<td>0.0429</td>
<td>y3(t-1)</td>
</tr>
<tr>
<td>y3</td>
<td>CONST3</td>
<td>-19.88165</td>
<td>7.69575</td>
<td>-2.58</td>
<td>0.0227</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>XL0_3_1</td>
<td>-0.03576</td>
<td>0.20079</td>
<td>-0.18</td>
<td>0.8614</td>
<td>x1(t)</td>
</tr>
<tr>
<td></td>
<td>XL0_3_2</td>
<td>-0.00919</td>
<td>0.01747</td>
<td>-0.53</td>
<td>0.6076</td>
<td>x2(t)</td>
</tr>
<tr>
<td></td>
<td>AR1_3_1</td>
<td>0.96398</td>
<td>0.06907</td>
<td>13.96</td>
<td>0.0001</td>
<td>y1(t-1)</td>
</tr>
<tr>
<td></td>
<td>AR1_3_2</td>
<td>0.00000</td>
<td>0.00000</td>
<td>y2(t-1)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>AR1_3_3</td>
<td>0.93412</td>
<td>0.01473</td>
<td>63.41</td>
<td>0.0001</td>
<td>y3(t-1)</td>
</tr>
</tbody>
</table>

The output in Figure 43.27 shows the estimates of the Lagrangian parameters and their significance. Based on the \(p\)-values associated with the Lagrangian parameters, you cannot reject the null hypotheses \(\theta_{12}^* = 0\), \(\phi_{12} = 0\), and \(\phi_{32} = 0\) with the 0.05 significance level.
The TEST statement in the following example tests $\phi_{31} = 0$ and $\phi_{12} = \phi_{32} = 0$ for the VARX(1,0) model:

```
proc varmax data=grunfeld;
   model y1-y3 = x1 x2 / p=1;
   test AR(1,3,1)=0;
   test XL(0,1,2)=0, AR(1,1,2)=0, AR(1,3,2)=0;
run;
```

The output in Figure 43.28 shows that the first column in the output is the index corresponding to each TEST statement. You can reject the hypothesis test $\phi_{31} = 0$ at the 0.05 significance level, but you cannot reject the joint hypothesis test $\phi_{12} = \phi_{32} = 0$ at the 0.05 significance level.

### Figure 43.28 TEST Statement Results

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Estimate</th>
<th>Standard Error</th>
<th>t Value</th>
<th>Pr &gt;</th>
<th>Equation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Restrict0</td>
<td>1.74969</td>
<td>21.44026</td>
<td>0.08</td>
<td>0.9389</td>
<td>XL0_1_2 = 0</td>
</tr>
<tr>
<td>Restrict1</td>
<td>30.36254</td>
<td>70.74347</td>
<td>0.43</td>
<td>0.6899</td>
<td>AR1_1_2 = 0</td>
</tr>
<tr>
<td>Restrict2</td>
<td>55.42191</td>
<td>164.03075</td>
<td>0.34</td>
<td>0.7524</td>
<td>AR1_3_2 = 0</td>
</tr>
</tbody>
</table>

### Causality Testing

The following statements use the CAUSAL statement to compute the Granger causality test for a VAR(1) model. For the Granger causality tests, the autoregressive order should be defined by the P= option in the MODEL statement. The variable groups are defined in the CAUSAL statement as well. Regardless of whether the variables specified in the GROUP1= and GROUP2= options are designated as dependent or exogenous (independent) variables in the MODEL statement, the CAUSAL statement fits the VAR($p$) model by considering the variables in the two groups as dependent variables.

```
/*--- Causality Testing ---*/
proc varmax data=grunfeld;
   model y1-y3 = x1 x2 / p=1 noprint;
   causal group1=(x1) group2=(y1-y3);
   causal group1=(y3) group2=(y1 y2);
run;
```
The output in Figure 43.29 is associated with the CAUSAL statement. The first CAUSAL statement fits the VAR(1) model by using the variables \(y_1\), \(y_2\), \(y_3\), and \(x_1\). The second CAUSAL statement fits the VAR(1) model by using the variables \(y_1\), \(y_3\), and \(y_2\).

**Figure 43.29** CAUSAL Statement Results

<table>
<thead>
<tr>
<th>Test</th>
<th>DF</th>
<th>Chi-Square</th>
<th>Pr &gt; ChiSq</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>3</td>
<td>2.40</td>
<td>0.4946</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>262.88</td>
<td>&lt;.0001</td>
</tr>
</tbody>
</table>

Test 1: Group 1 Variables: \(x_1\)

Group 2 Variables: \(y_1\) \(y_2\) \(y_3\)

Test 2: Group 1 Variables: \(y_3\)

Group 2 Variables: \(y_1\) \(y_2\)

The null hypothesis of the Granger causality test is that GROUP1 is influenced only by itself, and not by GROUP2.

The first column in the output is the index corresponding to each CAUSAL statement. The output shows that you cannot reject that \(x_1\) is influenced by itself and not by \((y_1, y_2, y_3)\) at the 0.05 significance level for Test 1. You can reject that \(y_3\) is influenced by itself and not by \((y_1, y_2)\) for Test 2. For more information, see the section “VAR and VARX Modeling” on page 3092.

**Multivariate GARCH Models**

Modeling and forecasting the volatility of time series has been the focus of many researchers and practitioners, especially in the fields of risk management, portfolio optimization, and asset pricing. One of the most powerful tools for volatility modeling is the autoregressive conditional heteroscedasticity (ARCH) model proposed by Engle (1982) and extended by Bollerslev (1986) to the generalized autoregressive conditional heteroscedasticity (GARCH) model. The VARMAX procedure supports three forms of multivariate GARCH models: BEKK, CCC, and DCC. This section shows some examples of how to specify, estimate, and compare various forms of multivariate GARCH models.

Data about two indices, the Dow Jones Industrial Average index and the Standard & Poor’s 500 index, are obtained from Yahoo Finance and used in this section. The sample contains daily data from February 16, 2005, to February 13, 2015. The following statements input the daily prices and then generate the daily returns:

```sas
data indices;
  input date : MMDDYY10. DJIA SP500;
  logDJIA = log(DJIA); logSP500 = log(SP500);
  rdJIA = (logDJIA-lag(logDJIA))*100;
  rSP500 = (logSP500-lag(logSP500))*100;
  datalines;
  2/16/2005 10834.88 1210.34
  2/17/2005 10754.26 1200.75
  2/18/2005 10785.22 1201.59
```

```sas
```
... more lines ...

To model the volatility of bivariate returns, rDJIA and rSP500, you can start with the BEKK GARCH(1,1) model. The following equations describe the bivariate BEKK GARCH(1,1) model:

\[
\begin{align*}
    r_t &= H_t^{\frac{1}{2}} \epsilon_t \\
    H_t &= C + A_1^t r_{t-1} r_{t-1}' + G_1 H_{t-1} G_1 \\
         &= \begin{bmatrix} c_{11} & c_{12} \\ c_{12} & c_{22} \end{bmatrix} + \begin{bmatrix} a_{11,1} & a_{12,1} \\ a_{21,1} & a_{22,1} \end{bmatrix}' \begin{bmatrix} r_{1,t-1} \\ r_{2,t-1} \end{bmatrix} \begin{bmatrix} a_{11,1} & a_{12,1} \\ a_{21,1} & a_{22,1} \end{bmatrix} \\
         &+ \begin{bmatrix} g_{11,1} & g_{12,1} \\ g_{21,1} & g_{22,1} \end{bmatrix}' \begin{bmatrix} h_{11,t-1} & h_{12,t-1} \\ h_{12,t-1} & h_{22,t-1} \end{bmatrix} \begin{bmatrix} g_{11,1} & g_{12,1} \\ g_{21,1} & g_{22,1} \end{bmatrix}
\end{align*}
\]

In these equations, \( r_t \) is the vector of returns at time \( t \), \( H_t \) is the conditional covariance matrix of \( r_t \), \( H_t^{\frac{1}{2}} \) denotes the square root of \( H_t \) such that the square of matrix \( H_t^{\frac{1}{2}} \) is \( H_t \), \( \epsilon_t \) is the innovation at time \( t \) and follows an iid bivariate standard normal distribution, \( C \) is a \( 2 \times 2 \) symmetric parameter matrix, \( A_1 \) is a \( 2 \times 2 \) full parameter matrix for the first lag of the ARCH term, and \( G_1 \) is a \( 2 \times 2 \) full parameter matrix for the first lag of the GARCH term. Hence, there are 11 parameters in total for a bivariate BEKK GARCH(1,1) model; that is, a vector \( \{ c_{11}, c_{12}, c_{22}, a_{11,1}, a_{21,1}, a_{12,1}, a_{22,1}, g_{11,1}, g_{12,1}, g_{21,1}, g_{22,1} \} \).

You can use the FORM=BEKK option in the GARCH statement to specify the BEKK GARCH form, or you can omit this option because BEKK is the default value for the FORM= option. The Q= option in the GARCH statement specifies the lags of the ARCH terms, and the P= option in the GARCH statement specifies the lags of the GARCH terms. The forecasts of conditional covariance matrices are output to a SAS data set when you specify the OUTHT= option in the GARCH statement. The parameter estimates and their covariance matrix are output to a SAS data set when you specify the OUTEST= option together with the OUTCOV option in the PROC VARMAX statement.

The following statement specifies the BEKK GARCH(1,1) model:

```plaintext
/*---- BEKK ----*/
proc varmax data=indices outest=oebekk outcov;
    model rDJIA rSP500 / noint;
    garch p=1 q=1 form=bekk outht=ohbekk;
run;
```

Figure 43.30 shows the log likelihood and the information criteria. They are used later in the model comparison.
Figure 43.30 BEKK GARCH Log Likelihood and Information Criteria

The VARMAX Procedure

<table>
<thead>
<tr>
<th>Information Criteria</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>AICC</td>
<td>-2699.85</td>
</tr>
<tr>
<td>HQC</td>
<td>-2676.68</td>
</tr>
<tr>
<td>AIC</td>
<td>-2699.95</td>
</tr>
<tr>
<td>SBC</td>
<td>-2635.82</td>
</tr>
<tr>
<td>FPEC</td>
<td>0.080617</td>
</tr>
</tbody>
</table>

Figure 43.31 shows the parameters estimates for the BEKK GARCH(1,1) model. For the constant term $C$, $GCHC_{i,j}, i, j = 1, 2$, correspond to parameters $c_{ij}$, respectively. Because $C$ is symmetric, $GCHC_{2,1}$ is omitted. For the ARCH and GARCH terms, $ACH_{i,j,l}, l = 1, i, j = 1, 2$, correspond to $a_{ij,l}$, respectively, and $GCH_{i,j,l}, l = 1, i, j = 1, 2$, correspond to $g_{ij,l}$, respectively.

Figure 43.31 BEKK GARCH Parameter Estimates

| GARCH Model Parameter Estimates | Standard Error | t Value | Pr > |t| |
|-------------------------------|---------------|---------|-----|---|
| GCHC1_1                        | 0.19101       | 0.00000 | 2.04 | 0.0417 |
| GCHC1_2                        | 0.09343       | 0.00000 | 1.85 | 0.0639 |
| GCHC2_2                        | 0.00000       | 0.01807 | 0.00 | 1.0000 |
| ACH1_1_1                       | 0.27518       | 0.13503 | 2.04 | 0.0417 |
| ACH1_2_1                       | 0.20619       | 0.11122 | 1.85 | 0.0639 |
| ACH1_1_2                       | 0.24907       | 0.11982 | 2.08 | 0.0377 |
| ACH1_2_2                       | 0.23448       | 0.09739 | 2.41 | 0.0161 |
| GCH1_1_1                       | 0.11391       | 0.10984 | 1.04 | 0.2998 |
| GCH1_2_1                       | 0.64841       | 0.11363 | 5.71 | 0.00001 |
| GCH1_1_2                       | 0.75455       | 0.11263 | 6.70 | 0.00001 |
| GCH1_2_2                       | 0.20598       | 0.11520 | 1.79 | 0.0739 |

As shown in Figure 43.31, the standard errors of $GCHC1_1$ and $GCHC1_2$ are both zeros. It might be a sign that the numerical optimization for the BEKK GARCH model converges to a local minimum instead of the global minimum, which often happens for nonlinear optimization of complex models that have many parameters. A possible way to solve this problem is to try different initial values. To obtain reasonable initial values, the following statements fit a diagonal BEKK GARCH model (that is, a restricted BEKK GARCH model in which the ARCH and GARCH parameter matrices are diagonal):

```plaintext
/*--- Diagonal BEKK ---*/

proc varmax data=indices outest=oebkk outcov;
  model rDJIA rSP500 / noint;
  garch p=1 q=1 form=bekk;
  restrict ach(1,1,2), ach(1,2,1), gch(1,1,2), gch(1,2,1);
run;
```
The parameter estimates of the diagonal BEKK GARCH model are shown in Figure 43.32. As expected, the standard errors of the off-diagonal elements of the ARCH and GARCH parameter matrices (namely $ACH1_{1,2}$, $ACH1_{2,1}$, $GCH1_{1,2}$, and $GCH1_{2,1}$) are all zeros because they are restricted in the RESTRICT statement. All other parameters have valid standard errors.

**Figure 43.32** Diagonal BEKK GARCH Parameter Estimates

| GARCH Model Parameter Estimates | Standard Error | t Value | Pr > |t| |
|-------------------------------|---------------|--------|------|
| Parameter                   | Estimate      |        |      | |
| GCHC1_1   | 0.01407       | 0.00254 | 5.53 | 0.0001 |
| GCHC1_2   | 0.01446       | 0.00262 | 5.51 | 0.0001 |
| GCHC2_2   | 0.01598       | 0.00299 | 5.34 | 0.0001 |
| ACH1_1_1  | 0.25702       | 0.01251 | 20.54| 0.0001 |
| ACH1_2_1  | 0.00000       | 0.00000 |      |      |
| ACH1_1_2  | 0.00000       | 0.00000 |      |      |
| ACH1_2_2  | 0.26061       | 0.01302 | 20.02| 0.0001 |
| GCH1_1_1  | -0.95794      | 0.00413 | -231.85 | 0.0001 |
| GCH1_2_1  | 0.00000       | 0.00000 |      |      |
| GCH1_1_2  | 0.00000       | 0.00000 |      |      |
| GCH1_2_2  | -0.95694      | 0.00443 | -216.10 | 0.0001 |

Figure 43.33 shows the log likelihood and the information criteria. The log likelihood for the diagonal BEKK GARCH model is larger than that of the previous estimated BEKK GARCH model (which is shown in Figure 43.30). The larger value confirms that the previous BEKK GARCH model does not converge to the global minimum.

**Figure 43.33** Diagonal BEKK GARCH Log Likelihood and Information Criteria

<table>
<thead>
<tr>
<th>Log-likelihood</th>
<th>1520.235</th>
</tr>
</thead>
<tbody>
<tr>
<td>Information Criteria</td>
<td></td>
</tr>
<tr>
<td>AICC</td>
<td>-3026.43</td>
</tr>
<tr>
<td>HQC</td>
<td>-3011.66</td>
</tr>
<tr>
<td>AIC</td>
<td>-3026.47</td>
</tr>
<tr>
<td>SBC</td>
<td>-2985.66</td>
</tr>
<tr>
<td>FPEC</td>
<td>0.080617</td>
</tr>
</tbody>
</table>

The following statements reestimate the BEKK GARCH model whose initial values are parameter estimates of the diagonal BEKK GARCH model (which are shown in Figure 43.32):

```latex
/**** BEKK with Initial Values ****/

proc varmax data=indices outest=oebekk outcov;
  model rDJIA rSP500 / noint;
  garch p=1 q=1 form=bekk;
  initial gchc(1,1)=0.01407, gchc(1,2)=0.01446, gchc(2,2)=0.01598,
```
The parameter estimates of the reestimated BEKK GARCH models are shown in Figure 43.34. The standard errors of all parameters are valid.

**Figure 43.34** Reestimated BEKK GARCH Parameter Estimates

The number of parameters for a BEKK GARCH model increases very quickly as the number of dependent variables increases. There are \((p + q)k^2 + k(k + 1)/2\) parameters for a \(k\)-variate BEKK GARCH\((p, q)\) model. For example, a 16-variate BEKK GARCH(1,1) model has 648 parameters to be estimated.

Compared with BEKK GARCH models, CCC GARCH models are much more parsimonious. In a CCC GARCH model, each series follows a GARCH process and their composition through the constant conditional
correlation matrix constructs the conditional covariance matrices. A bivariate CCC GARCH(1,1) has the form

\[
\begin{align*}
    r_t &= H_t^{1/2} \epsilon_t \\
    H_t &= D_t S D_t \\
    D_t &= \begin{bmatrix} \sqrt{h_{11,t}} & 0 \\
    0 & \sqrt{h_{22,t}} \end{bmatrix} \\
    S &= \begin{bmatrix} 1 & s_{12} \\
    s_{12} & 1 \end{bmatrix} \\
    h_{11,t} &= c_{11} + a_{11,1} r_{1,t-1}^2 + g_{11,1} h_{11,t-1} \\
    h_{22,t} &= c_{22} + a_{22,1} r_{2,t-1}^2 + g_{22,1} h_{22,t-1}
\end{align*}
\]

where \( D_t \) is the diagonal matrix with conditional standard deviations and \( S \) is the time-invariant conditional correlation matrix. Hence, there are seven parameters to be estimated; that is, a vector \( \left( s_{12}, c_{11}, c_{22}, a_{11,1}, a_{22,1}, g_{11,1}, g_{22,1} \right) \). A \( k \)-variate CCC GARCH(\( p, q \)) model has \( (p + q + 1)k + k(k - 1)/2 \) parameters. For example, a 16-variate CCC GARCH(1,1) model has 168 parameters to be estimated, about 1/4 of the number that a BEKK GARCH(1,1) model has.

The following statements estimate a CCC GARCH(1,1) model:

```plaintext
/*--- CCC ---*/
proc varmax data=indices outest=oeccc outcov;
    model rDJIA rSP500 / noint;
    garch p=1 q=1 form=ccc outht=ohccc;
run;
```

Figure 43.36 shows the parameter estimates for the CCC GARCH(1,1) model. For the constant conditional correlation matrix \( S \), \( CCC1_2 \) corresponds to the parameter \( s_{12} \).

**Figure 43.36** CCC GARCH Parameter Estimates

| Parameter  | Estimate | Standard Error | t Value | Pr > |t| |
|------------|----------|----------------|---------|------|---|
| CCC1_2     | 0.97294  | 0.00109        | 890.75  | 0.0001 |
| GCHC1_1    | 0.03713  | 0.00504        | 7.37    | 0.0001 |
| GCHC2_2    | 0.04004  | 0.00536        | 7.47    | 0.0001 |
| ACH1_1_1   | 0.06862  | 0.00737        | 9.31    | 0.0001 |
| ACH1_2_2   | 0.06684  | 0.00690        | 9.68    | 0.0001 |
| GCH1_1_1   | 0.88472  | 0.01183        | 74.81   | 0.0001 |
| GCH1_2_2   | 0.88916  | 0.01099        | 80.92   | 0.0001 |
Figure 43.37 shows the log likelihood and the information criteria. Compared to the SBC for the BEKK GARCH model (shown in Figure 43.35), the SBC for the CCC GARCH model is much larger, which means the CCC GARCH model should not be preferred.

The CCC GARCH model is not preferred over the BEKK GARCH model in this case because the basic assumption in the CCC GARCH model—that the conditional correlation matrix is time-invariant—might not hold. A DCC GARCH model relaxes this assumption and models the time-varying conditional correlation matrix in an ARMA form. A bivariate DCC GARCH(1,1) has the form

\[
\begin{align*}
    r_t &= H_t^{1/2} \epsilon_t \\
    H_t &= D_t S_t D_t \\
    D_t &= \begin{bmatrix}
        \sqrt{h_{11,t}} & 0 \\
        0 & \sqrt{h_{22,t}}
    \end{bmatrix} \\
    h_{11,t} &= c_{11} + a_{11,1} r_{1,t-1}^2 + g_{11,1} h_{11,t-1} \\
    h_{22,t} &= c_{11} + a_{22,1} r_{2,t-1}^2 + g_{22,1} h_{22,t-1} \\
    S_t &= \begin{bmatrix}
        1 & s_{12,t} \\
        s_{12,t} & 1
    \end{bmatrix} \\
    s_{12,t} &= \frac{q_{12,t}}{\sqrt{q_{11,t} q_{22,t}}} \\
    q_{12,t} &= (1 - \alpha - \beta) s_{12} + \alpha \frac{r_{1,t-1}^2}{\sqrt{h_{11,t-1} h_{11,t}}} \frac{r_{2,t-1}^2}{\sqrt{h_{22,t-1} h_{22,t}}} + \beta q_{12,t-1} \\
    q_{11,t} &= (1 - \alpha - \beta) + \alpha \frac{r_{1,t-1}^2}{h_{11,t-1} h_{11,t}} + \beta q_{11,t-1} \\
    q_{22,t} &= (1 - \alpha - \beta) + \alpha \frac{r_{2,t-1}^2}{h_{22,t-1} h_{22,t}} + \beta q_{22,t-1}
\end{align*}
\]

where \( S_t \) is the time-varying conditional correlation matrix at time \( t \). Compared to the CCC GARCH model, two more parameters, \( \alpha \) and \( \beta \), are added into the DCC GARCH model. There are nine parameters in total; that is, a vector \( (\alpha, \beta, s_{12}, c_{11}, c_{22}, a_{11,1}, a_{22,1}, g_{11,1}, g_{22,1})^T \).
The following statements estimate a DCC GARCH model:

```plaintext
/*--- DCC ---*/
proc varmax data=indices outest=oedcc outcov;
   model rDJIA rSP500 / noint;
   garch p=1 q=1 form=dcc outht=ohdcc;
run;
```

Figure 43.38 shows the parameter estimates for the DCC GARCH(1,1) model. DCCA corresponds to the parameter $\alpha$, DCCB corresponds to the parameter $\beta$, and DCCS1_2 corresponds to the parameter $s_{12}$, the off-diagonal element in the unconditional correlation matrix. The standard errors of many parameter estimates are zeros, which might be a sign that the model does not converge to the global minimum.

**Figure 43.38** DCC GARCH Parameter Estimates

<table>
<thead>
<tr>
<th>GARCH Model Parameter Estimates</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Parameter</td>
<td>Estimate</td>
<td>Standard Error</td>
<td>t Value</td>
<td>Pr &gt;</td>
<td>t</td>
</tr>
<tr>
<td>DCCA</td>
<td>0.01540</td>
<td>0.00000</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DCCB</td>
<td>0.00000</td>
<td>0.00000</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DCCS1_2</td>
<td>0.98743</td>
<td>0.00040</td>
<td>999.00</td>
<td>0.0001</td>
<td></td>
</tr>
<tr>
<td>GCHC1_1</td>
<td>1.28530</td>
<td>0.00000</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GCHC2_2</td>
<td>1.50117</td>
<td>0.00000</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ACH1_1_1</td>
<td>0.03378</td>
<td>0.00216</td>
<td>15.62</td>
<td>0.0001</td>
<td></td>
</tr>
<tr>
<td>ACH1_2_2</td>
<td>0.02694</td>
<td>0.00084</td>
<td>32.07</td>
<td>0.0001</td>
<td></td>
</tr>
<tr>
<td>GCH1_1_1</td>
<td>0.07596</td>
<td>0.00000</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GCH1_2_2</td>
<td>0.09939</td>
<td>0.00000</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Figure 43.39 shows the log likelihood and the information criteria.

**Figure 43.39** DCC GARCH Log Likelihood and Information Criteria

<table>
<thead>
<tr>
<th>Log-likelihood</th>
<th>700.3131</th>
</tr>
</thead>
<tbody>
<tr>
<td>Information Criteria</td>
<td></td>
</tr>
<tr>
<td>AICC</td>
<td>-1382.55</td>
</tr>
<tr>
<td>HQC</td>
<td>-1363.58</td>
</tr>
<tr>
<td>AIC</td>
<td>-1382.63</td>
</tr>
<tr>
<td>SBC</td>
<td>-1330.16</td>
</tr>
<tr>
<td>FPEC</td>
<td>0.080617</td>
</tr>
</tbody>
</table>

Because a CCC GARCH model can be regarded as a restricted DCC GARCH model in which $\alpha$ and $\beta$ in the conditional correlation equations are restricted to zeros, it is expected that the log likelihood of the “unrestricted” DCC GARCH model should always be larger than (or at least equal to) the log likelihood of the
corresponding CCC GARCH model, even though DCC might have a larger information criterion and not be chosen. Hence, it is unexpected that the log likelihood of the DCC GARCH model (shown in Figure 43.39) is smaller than that of the CCC GARCH model (shown in Figure 43.37). This unexpected phenomenon confirms that the numerical optimization for the DCC GARCH model converges to a local minimum instead of the global minimum. Different initial values should be tried. In addition to some reasonable values for parameters DCCA and DCCB, the INITIAL statement specifies the initial values for the DCC GARCH model parameters in the following statements; these values are based on the corresponding parameter estimates of the CCC GARCH model (shown in Figure 43.36):

/**** DCC with Initial Values ****/

proc varmax data=indices outest=oedcc outcov;
  model rDJIA rSP500 / noint;
  garch p=1 q=1 form=dcc outht=ohdccc;
    initial DCCA=0.01, DCCB=0.98, DCCS(1,2) = 0.97294,
    GCHC(1,1) = 0.03713, GCHC(2,2) = 0.04004,
    ACH(1,1,1) = 0.06862, ACH(1,2,2) = 0.06684,
    GCH(1,1,1) = 0.88472, GCH(1,2,2) = 0.88916;
  run;

Figure 43.40 shows the parameter estimates for the reestimated DCC GARCH(1,1) model. All standard errors of parameter estimates are valid.

**Figure 43.40**  Reestimated DCC GARCH Parameter Estimates

| GARCH Model Parameter Estimates | Standard Error | t Value | Pr > |t| |
|---------------------------------|----------------|---------|------|---|
| Parameter                      | Estimate       |          |      |   |
| DCCA                           | 0.03802        | 0.00634  | 6.00 | 0.0001 |
| DCCB                           | 0.93782        | 0.01084  | 86.49 | 0.0001 |
| DCCS1,2                        | 0.97401        | 0.00247  | 394.22 | 0.0001 |
| GCHC1,1                        | 0.02193        | 0.00370  | 5.93  | 0.0001 |
| GCHC2,2                        | 0.02395        | 0.00401  | 5.97  | 0.0001 |
| ACH1,1,1                       | 0.07842        | 0.00787  | 9.97  | 0.0001 |
| ACH1,2,2                       | 0.07758        | 0.00770  | 10.07 | 0.0001 |
| GCH1,1,1                       | 0.89540        | 0.01046  | 85.58 | 0.0001 |
| GCH1,2,2                       | 0.89738        | 0.01012  | 86.64 | 0.0001 |

As shown in Figure 43.41, the log likelihood of the DCC GARCH model increases dramatically. Compared to the SBC of the CCC GARCH model (shown in Figure 43.37), the SBC for the DCC GARCH model is much smaller, and the DCC GARCH model is chosen. However, compared to the SBC for the BEKK GARCH model (shown in Figure 43.35), the SBC for the DCC GARCH model is a little larger. The BEKK GARCH model should be chosen although it has two more parameters than the DCC GARCH model.
Compared to the BEKK GARCH model, in addition to parsimony, another advantage of the DCC (and also the CCC) GARCH model is that you can use subforms other than GARCH to model the conditional covariance of each series. For example, you can use the threshold GARCH (TGARCH) model for modeling the conditional covariances of rDJIA and rSP500. A bivariate DCC TGARCH(1,1) has the same form as the bivariate DCC GARCH(1,1) except that the conditional covariance equations are replaced by

\[
\begin{align*}
    h_{11,t} &= c_{11} + a_{11,1} r_{1,t-1}^2 + 1_{r_{1,t-1} < 0} b_{11,1} r_{1,t-1}^2 + g_{11,1} h_{11,t-1} \\
    h_{22,t} &= c_{11} + a_{22,1} r_{2,t-1}^2 + 1_{r_{2,t-1} < 0} b_{22,1} r_{2,t-1}^2 + g_{22,1} h_{22,t-1}
\end{align*}
\]

where the indicator function \(1_A\) is 1 if \(A\) is true and 0 otherwise. Compared to the DCC GARCH model, two more parameters, \(b_{11,1}\) and \(b_{22,1}\), are added to the DCC TGARCH model to catch the so-called leverage effect (that is, the positive and negative returns have different impacts on future volatility).

The following statements include the SUBFORM=TARCH option to fit a bivariate DCC TGARCH(1,1) model with the same initial values that are used for the previous DCC GARCH(1,1) model:

```plaintext
proc varmax data=indices outest=oedcct outcov;
    model rDJIA rSP500 / noint;
        garch p=1 q=1 form=dcc outht=ohdcct subform=tgarch;
        initial DCCA=0.01, DCCB=0.98, DCCS(1,2) = 0.97294,
            GCHC(1,1) = 0.03713, GCHC(2,2) = 0.04004,
            ACH(1,1,1) = 0.06862, ACH(1,2,2) = 0.06684,
            GCH(1,1,1) = 0.88472, GCH(1,2,2) = 0.88916;
    output out=odcct lead=0;
run;
```

Figure 43.42 shows the parameter estimates for the DCC TGARCH(1,1) model. \(TACH1_1_1\) and \(TACH1_2_2\) correspond to the parameters \(b_{11,1}\) and \(b_{22,1}\), respectively. They are significant, which means that the leverage effect exists.
Figure 43.42  DCC TGARCH Parameter Estimates

The VARMAX Procedure

| Parameter | Estimate | Standard Error | t Value | Pr > |t| |
|-----------|----------|----------------|---------|------|---|
| DCCA      | 0.04302  | 0.00669        | 6.43    | 0.0001 |
| DCCB      | 0.92807  | 0.01142        | 81.26   | 0.0001 |
| DCCS1_2   | 0.97309  | 0.00248        | 392.01  | 0.0001 |
| GCCHC1_1  | 0.02068  | 0.00305        | 6.78    | 0.0001 |
| GCCHC2_2  | 0.02329  | 0.00346        | 6.73    | 0.0001 |
| ACH1_1_1  | 0.00104  | 0.00684        | 0.15    | 0.8787 |
| ACH1_2_2  | 0.00314  | 0.00698        | 0.45    | 0.6525 |
| TACH1_1_1 | 0.11443  | 0.01207        | 9.48    | 0.0001 |
| TACH1_2_2 | 0.10805  | 0.01166        | 9.27    | 0.0001 |
| GCCH1_1_1 | 0.91490  | 0.00956        | 95.68   | 0.0001 |
| GCCH1_2_2 | 0.91574  | 0.00964        | 95.03   | 0.0001 |

Figure 43.43 shows the log likelihood and the information criteria. The SBC for the DCC TGARCH model is smaller than the SBC for the BEKK GARCH model (which is shown in Figure 43.35). The smaller SBC means that the DCC TGARCH model is the final winner.

Other subforms of GARCH models—the exponential GARCH (EGARCH) model, the quadratic GARCH (QGARCH) model, and the power GARCH (PGARCH) model—are also supported for the CCC and DCC GARCH models. Furthermore, the multivariate GARCH models can be used together with VARMAX or vector error correction models. For more information and examples, see the sections “Multivariate GARCH Modeling” on page 3135 and “Example 43.4: Analysis of Euro Foreign Exchange Reference Rates” on page 3197.
Syntax: VARMAX Procedure

PROC VARMAX options;
    BOUND restriction, . . . , restriction;
    BY variables;
    CAUSAL GROUP1=(variables) GROUP2=(variables);
    COINTEG RANK=number < options >;
    GARCH options;
    ID variable INTERVAL=value < ALIGN=value >;
    INITIAL equation, . . . , equation;
    MODEL dependents = regressors . . . < / options >;
    NLOPTIONS options;
    OUTPUT < options >;
    RESTRICT restriction, . . . , restriction;
    TEST restriction, . . . , restriction;

Functional Summary

The statements and options available in the VARMAX procedure are summarized in Table 43.1.

Table 43.1 Functional Summary

<table>
<thead>
<tr>
<th>Description</th>
<th>Statement</th>
<th>Option</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Set Options</td>
<td>VARMAX</td>
<td>DATA=</td>
</tr>
<tr>
<td>Specifies the input data set</td>
<td>VARMAX</td>
<td>OUTTEST=</td>
</tr>
<tr>
<td>Writes parameter estimates to an output data set</td>
<td>VARMAX</td>
<td>OUTCOV</td>
</tr>
<tr>
<td>Includes covariances in the OUTTEST= data set</td>
<td>VARMAX</td>
<td>OUTSTAT=</td>
</tr>
<tr>
<td>Writes the diagnostic checking tests for a model and the cointegration test results to an output data set</td>
<td>VARMAX</td>
<td>OUT=</td>
</tr>
<tr>
<td>Writes actuals, predictions, residuals, and confidence limits to an output data set</td>
<td>OUTPUT</td>
<td>OUT=</td>
</tr>
<tr>
<td>Writes the conditional covariance matrix to an output data set</td>
<td>GARCH</td>
<td>OUTHT=</td>
</tr>
</tbody>
</table>

BY Groups

Specifies BY-group processing

ID Variable

Specifies the identifying variable

Specifies the time interval between observations

Controls the alignment of SAS date values

Options to Control the Optimization Process

Specifies the optimization options

NLOPTIONS
Table 43.1  continued

<table>
<thead>
<tr>
<th>Description</th>
<th>Statement</th>
<th>Option</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Printing Control Options</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Specifies how many lags to print results</td>
<td>MODEL</td>
<td>LAGMAX=</td>
</tr>
<tr>
<td>Suppresses the printed output</td>
<td>MODEL</td>
<td>NOPRINT</td>
</tr>
<tr>
<td>Requests all printing options</td>
<td>MODEL</td>
<td>PRINTALL</td>
</tr>
<tr>
<td>Requests the printing format</td>
<td>MODEL</td>
<td>PRINTFORM=</td>
</tr>
<tr>
<td>Controls plots produced through ODS GRAPHICS</td>
<td>VARMAX</td>
<td>PLOTS=</td>
</tr>
<tr>
<td><strong>PRINT= Option</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Prints the correlation matrix of parameter estimates</td>
<td>MODEL</td>
<td>CORRB</td>
</tr>
<tr>
<td>Prints the cross-correlation matrices of independent variables</td>
<td>MODEL</td>
<td>CORRX</td>
</tr>
<tr>
<td>Prints the cross-correlation matrices of dependent variables</td>
<td>MODEL</td>
<td>CORRY</td>
</tr>
<tr>
<td>Prints the covariance matrices of prediction errors</td>
<td>MODEL</td>
<td>COVPE</td>
</tr>
<tr>
<td>Prints the cross-covariance matrices of the independent variables</td>
<td>MODEL</td>
<td>COVX</td>
</tr>
<tr>
<td>Prints the cross-covariance matrices of the dependent variables</td>
<td>MODEL</td>
<td>COVY</td>
</tr>
<tr>
<td>Prints the covariance matrix of parameter estimates</td>
<td>MODEL</td>
<td>COVB</td>
</tr>
<tr>
<td>Prints the decomposition of the prediction error covariance matrix</td>
<td>MODEL</td>
<td>DECOMPOSE</td>
</tr>
<tr>
<td>Prints the residual diagnostics</td>
<td>MODEL</td>
<td>DIAGNOSE</td>
</tr>
<tr>
<td>Prints the contemporaneous relationships among the components of the vector time series</td>
<td>MODEL</td>
<td>DYNAMIC</td>
</tr>
<tr>
<td>Prints the parameter estimates</td>
<td>MODEL</td>
<td>ESTIMATES</td>
</tr>
<tr>
<td>Prints the infinite order AR representation</td>
<td>MODEL</td>
<td>IARR</td>
</tr>
<tr>
<td>Prints the impulse response function</td>
<td>MODEL</td>
<td>IMPULSE=</td>
</tr>
<tr>
<td>Prints the impulse response function in the transfer function</td>
<td>MODEL</td>
<td>IMPULSX=</td>
</tr>
<tr>
<td>Prints the partial autoregressive coefficient matrices</td>
<td>MODEL</td>
<td>PARCOEF</td>
</tr>
<tr>
<td>Prints the partial canonical correlation matrices</td>
<td>MODEL</td>
<td>PCANCORR</td>
</tr>
<tr>
<td>Prints the partial correlation matrices</td>
<td>MODEL</td>
<td>PCORR</td>
</tr>
<tr>
<td>Prints the eigenvalues of the companion matrix</td>
<td>MODEL</td>
<td>ROOTS</td>
</tr>
<tr>
<td>Prints the Yule-Walker estimates</td>
<td>MODEL</td>
<td>YW</td>
</tr>
<tr>
<td><strong>Model Estimation and Order Selection Options</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Specifies the initial parameter values for non-linear optimization when the model is estimated through the maximum likelihood method</td>
<td>INITIAL</td>
<td></td>
</tr>
<tr>
<td>Centers the dependent variables</td>
<td>MODEL</td>
<td>CENTER</td>
</tr>
<tr>
<td>Specifies the degrees of differencing for the specified model variables</td>
<td>MODEL</td>
<td>DIF=</td>
</tr>
<tr>
<td>Specifies the degrees of differencing for all independent variables</td>
<td>MODEL</td>
<td>DIFX=</td>
</tr>
<tr>
<td>Description</td>
<td>Statement</td>
<td>Option</td>
</tr>
<tr>
<td>-----------------------------------------------------------------------------</td>
<td>-----------</td>
<td>--------------</td>
</tr>
<tr>
<td>Specifies the degrees of differencing for all dependent variables</td>
<td>MODEL</td>
<td>DIFY=</td>
</tr>
<tr>
<td>Specifies the estimation method</td>
<td>MODEL</td>
<td>METHOD=</td>
</tr>
<tr>
<td>Selects the tentative order</td>
<td>MODEL</td>
<td>MINIC=</td>
</tr>
<tr>
<td>Suppresses the current values of independent variables</td>
<td>MODEL</td>
<td>NOCURRENTX</td>
</tr>
<tr>
<td>Suppresses the intercept parameters</td>
<td>MODEL</td>
<td>NOINT</td>
</tr>
<tr>
<td>Specifies the number of seasonal periods</td>
<td>MODEL</td>
<td>NSEASON=</td>
</tr>
<tr>
<td>Specifies the order of autoregressive polynomial</td>
<td>MODEL</td>
<td>P=</td>
</tr>
<tr>
<td>Specifies the Bayesian prior model</td>
<td>MODEL</td>
<td>PRIOR=</td>
</tr>
<tr>
<td>Specifies the order of moving-average polynomial</td>
<td>MODEL</td>
<td>Q=</td>
</tr>
<tr>
<td>Centers the seasonal dummies</td>
<td>MODEL</td>
<td>SCENTER</td>
</tr>
<tr>
<td>Specifies the degree of time trend polynomial</td>
<td>MODEL</td>
<td>TREND=</td>
</tr>
<tr>
<td>Specifies the denominator for error covariance matrix estimates</td>
<td>MODEL</td>
<td>VARDEF=</td>
</tr>
<tr>
<td>Specifies the lag order of independent variables</td>
<td>MODEL</td>
<td>XLAG=</td>
</tr>
<tr>
<td><strong>GARCH-Related Options</strong></td>
<td>GARCH</td>
<td>CORRCONSTANT=</td>
</tr>
<tr>
<td>Specifies how to calculate the constant (unconditional) correlation matrix</td>
<td>GARCH</td>
<td>FORM=</td>
</tr>
<tr>
<td>in the CCC (DCC) GARCH model</td>
<td>GARCH</td>
<td>P=</td>
</tr>
<tr>
<td>Specifies the type of the multivariate GARCH model</td>
<td>GARCH</td>
<td>Q=</td>
</tr>
<tr>
<td>Specifies the order of the GARCH polynomial</td>
<td>GARCH</td>
<td>SUBFORM=</td>
</tr>
<tr>
<td>Specifies the order of the ARCH polynomial</td>
<td>GARCH</td>
<td></td>
</tr>
<tr>
<td>Specifies the type of the univariate GARCH model for each innovation in the</td>
<td>GARCH</td>
<td></td>
</tr>
<tr>
<td>CCC or DCC GARCH model</td>
<td>GARCH</td>
<td></td>
</tr>
<tr>
<td><strong>Cointegration-Related Options</strong></td>
<td>COINTEG</td>
<td>ECTREND</td>
</tr>
<tr>
<td>Specifies the restriction on the drift in the VECM</td>
<td>COINTEG</td>
<td>EXOGENEITY</td>
</tr>
<tr>
<td>Prints the results from the weak exogeneity test of the long-run parameters</td>
<td>COINTEG</td>
<td>H=</td>
</tr>
<tr>
<td>Specifies the restriction on the cointegrated coefficient matrix</td>
<td>COINTEG</td>
<td>J=</td>
</tr>
<tr>
<td>Specifies the restriction on the adjustment coefficient matrix</td>
<td>COINTEG</td>
<td>NLC</td>
</tr>
<tr>
<td>Specifies the nonlinear constraints that the adjustment coefficient matrix and the cointegrated coefficient matrix are both full rank</td>
<td>COINTEG</td>
<td>NORMALIZE=</td>
</tr>
<tr>
<td>Specifies the variable name whose cointegrating vectors are normalized</td>
<td>COINTEG</td>
<td>RANK=</td>
</tr>
</tbody>
</table>
Table 43.1  continued

<table>
<thead>
<tr>
<th>Description</th>
<th>Statement</th>
<th>Option</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prints the Johansen cointegration rank test</td>
<td>MODEL</td>
<td>CINTTEST=(JOHANSEN= )</td>
</tr>
<tr>
<td>Prints the Stock-Watson common trends test</td>
<td>MODEL</td>
<td>CINTTEST=(SW= )</td>
</tr>
<tr>
<td>Prints the Dickey-Fuller unit root test</td>
<td>MODEL</td>
<td>DFTEST=</td>
</tr>
<tr>
<td>Specifies the vector error correction model (obsolete)¹</td>
<td>MODEL</td>
<td>ECM=</td>
</tr>
<tr>
<td><strong>Long Memory Options</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Specifies the Vector autoregressive fractionally integrated moving average model</td>
<td>MODEL</td>
<td>FI</td>
</tr>
<tr>
<td><strong>Tests and Restrictions on Parameters</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tests the Granger causality</td>
<td>CAUSAL</td>
<td>GROUP1=</td>
</tr>
<tr>
<td>Places and tests restrictions on parameter estimates</td>
<td>BOUND</td>
<td>GROUP2=</td>
</tr>
<tr>
<td>Places and tests restrictions on parameter estimates</td>
<td>RESTRICT</td>
<td></td>
</tr>
<tr>
<td>Tests hypotheses on parameter estimates</td>
<td>TEST</td>
<td></td>
</tr>
<tr>
<td><strong>Forecasting Control Options</strong></td>
<td>OUTPUT</td>
<td>ALPHA=</td>
</tr>
<tr>
<td>Specifies the size of confidence limits for forecasting</td>
<td>OUTPUT</td>
<td>BACK=</td>
</tr>
<tr>
<td>Starts forecasting before end of the input data</td>
<td>OUTPUT</td>
<td>LEAD=</td>
</tr>
<tr>
<td>Specifies how many periods to forecast</td>
<td>OUTPUT</td>
<td>NOPRINT</td>
</tr>
<tr>
<td>Suppresses the printed forecasts</td>
<td>OUTPUT</td>
<td></td>
</tr>
</tbody>
</table>

PROC VARMAX Statement

PROC VARMAX options ;

The following options can be used in the PROC VARMAX statement:

**DATA=SAS-data-set**

specifies the input SAS data set. If the DATA= option is not specified, the PROC VARMAX statement uses the most recently created SAS data set.

**OUTEST=SAS-data-set**

writes the parameter estimates to the output data set.

**COVOUT**

**OUTCOV**

writes the covariance matrix for the parameter estimates to the OUTEST= data set. This option is valid only if the OUTEST= option is specified.

¹Starting with SAS/ETS 14.1, it is recommended that you use the COINTEG statement instead.
Chapter 43: The VARMAX Procedure

OUTSTAT=SAS-data-set
writes residual diagnostic results to an output data set. If the COINTTEST=(JOHANSEN) option is specified, the results of this option are also written to the output data set.

The following statements are the examples of these options in the PROC VARMAX statement:

```sas
proc varmax data=one outest=est outcov outstat=stat;
  model y1-y3 / p=1;
run;
```

```sas
proc varmax data=one outest=est outstat=stat;
  model y1-y3 / p=1 cointest=(johansen);
run;
```

PLOTS< (global-plot-option) >= plot-request-option <(options)>  
PLOTS< (global-plot-option) >= ( plot-request-option <(options)> ...plot-request-option <(options)> )  
controls the plots produced through ODS Graphics. When you specify only one plot, you can omit the parentheses around the plot request. Some examples follow:

```sas
plots=none
plots=all
plots(unpack)=residual(residual normal)
plots=(forecasts model)
```


```sas
proc varmax data=one plots=impulse(simple);
  model y1-y3 / p=1;
run;
```

```sas
proc varmax data=one plots=(model residual);
  model y1-y3 / p=1;
run;
```

```sas
proc varmax data=one plots=forecasts;
  model y1-y3 / p=1;
  output lead=12;
run;
```

The first VARMAX program produces the simple response impulse plots. The second VARMAX program produces the plots associated with the model and prediction errors. The plots associated with prediction errors are the ACF, PACF, IACF, distribution, white-noise, and Normal quantile plots and the prediction error plot. The third VARMAX program produces the FORECASTS and FORECASTSONLY plots.

The global-plot-option applies to the impulse and prediction error analysis plots generated by the VARMAX procedure. The following global-plot-option is available:
UNPACK displays each graph separately. (By default, some graphs can appear together in a single panel.)

The following plot-request-options are available:

**ALL** produces all plots appropriate for the particular analysis.

**FORECASTS** produces plots of the forecasts. The forecasts-only plot that shows the multistep forecasts in the forecast region is produced by default. The following forecasts-plot-options are available:

- **ALL** produces the FORECASTSONLY and the FORECASTS plots. This is the default.
- **FORECASTS** produces a plot that shows the one-step-ahead as well as the multistep forecasts.
- **FORECASTSONLY** produces a plot that shows only the multistep forecasts.

**IMPULSE** produces the plots of impulse response function and the impulse response of the transfer function.

- **ALL** produces all impulse plots. This is the default.
- **ACCUM** produces the accumulated impulse plot.
- **ORTH** produces the orthogonalized impulse plot.
- **SIMPLE** produces the simple impulse plot.

**MODEL** produces plots of dependent variables listed in the MODEL statement and plots of the one-step-ahead predicted values for each dependent variables.

**NONE** suppresses all plots.

**RESIDUAL** produces plots associated with the prediction errors obtained after modeling the data. The following residual-plot-options are available:

- **ALL** produces all plots associated with the analysis of the prediction errors. This is the default.
- **RESIDUAL** produces prediction error plot.
- **DIAGNOSTICS** produces a panel of plots useful in assessing the autocorrelations and white-noise of the prediction errors. The panel consists of the following:
  - the autocorrelation plot of the prediction errors
  - the partial autocorrelation plot of the prediction errors
  - the inverse autocorrelation plot of the prediction errors
  - the log scaled white noise plot of the prediction errors
- **NORMAL** produces a panel of plots useful in assessing normality of the prediction errors. The panel consists of the following:
  - distribution of the prediction errors with overlaid the normal curve
  - normal quantile plot of the prediction errors
Other Options

In addition, any of the following MODEL statement options can be specified in the PROC VARMAX statement, which is equivalent to specifying the option for every MODEL statement: CENTER, DFTEST=, DIF=, DIFX=, DIFY=, LAGMAX=, METHOD=, MINIC=, NOCURRENTX, NOINT, NOPRINT, NSEASON=, P=, PRINT=, PRINTALL, PRINTFORM=, Q=, SCENTER, TREND=, VARDEF=, and XLAG= options.

The following is an example of the options in the PROC VARMAX statement:

```sql
proc varmax data=one lagmax=3 method=ml;
  model y1-y3 / p=1;
run;
```

BOUND Statement

BOUND restriction, . . . , restriction;

The BOUND statement sets up linear bounds for parameters when the maximum likelihood method is applied to the estimation of VARMAX, VECM, VARMAX-GARCH, and VEC-ARMAX-GARCH models. Only one BOUND statement is allowed. If you specify more than one restriction, separate them with commas. The restrictions are specified in the same manner as the restrictions in the RESTRICT statement. For information about how to define restrictions by using matrix expressions, operators, and functions, see the section “RESTRICT Statement” on page 3050. Both equality and inequality constraints are allowed in the BOUND statement, although usually equality constraints are specified in the RESTRICT statement and inequality constraints are specified in the BOUND statement.

To use the BOUND statement, you need to know the form of the model. If you do not specify the GARCH statement, the COINTEG statement, or the ECM=, P=, Q=, or XLAG= option in the MODEL statement, then the BOUND statement is not applicable. If you specify the ECM=(NORMALIZE=), METHOD=LS, or PRIOR= option in the MODEL statement, or if you specify the EXOGENEITY, H=, J=, or NORMALIZE= option in the COINTEG statement, the BOUND statement is ignored. Nonlinear restrictions on parameters are not supported.

The following is an example of the BOUND statement for a bivariate \((k=2)\) zero-mean VARMA(1,1) model, which is by default estimated by maximum likelihood method because the MA term is present:

```sql
proc varmax data=one;
  model y1 y2 / noint p=1 q=1;
  bound -1<=AR<=1, 0<MA;
run;
```

This BOUND statement specifies that all AR parameters must be between \(-1\) and 1 and that all MA parameters must be positive.

You can use the BOUND statement together with the RESTRICT statement, as in the following bivariate \((k=2)\) zero-mean VARMA(1,1) model:

```sql
proc varmax data=one;
  model y1 y2 / noint p=1 q=1;
  bound AR+MA>=0.001;
  restrict AR(1,1,2) = 0.5;
run;
```
BY Statement

BY variables;

A BY statement can be used with PROC VARMAX to obtain separate analyses on observations in groups defined by the BY variables.

When a BY statement appears, the procedure expects the input data set to be sorted in order of the BY variables.

If your input data set is not sorted in ascending order, use one of the following alternatives:

- Sort the data using the SORT procedure with a similar BY statement.
- Specify the BY statement option NOTSORTED or DESCENDING in the BY statement for the VARMAX procedure. The NOTSORTED option does not mean that the data are unsorted but rather that the data are arranged in groups (according to values of the BY variables) and that these groups are not necessarily in alphabetical or increasing numeric order.
- Create an index on the BY variables using the DATASETS procedure.

For more information about the BY statement, see in SAS Language Reference: Concepts. For more information about the DATASETS procedure, see the discussion in the Base SAS Procedures Guide.

The following is an example of the BY statement:

```sas
proc varmax data=one;
  by region;
   model y1-y3 / p=1;
run;
```

CAUSAL Statement

CAUSAL GROUP1=(variables) GROUP2=(variables);

A CAUSAL statement prints the Granger causality test by fitting the VAR(p) model by using all variables defined in GROUP1 and GROUP2. Any number of CAUSAL statements can be specified. The CAUSAL statement proceeds with the MODEL statement and uses the variables and the autoregressive order, p, specified in the MODEL statement. Variables in the GROUP1= and GROUP2= options should be defined in the MODEL statement. If the P=0 option is specified in the MODEL statement, the CAUSAL statement is not applicable.

The null hypothesis of the Granger causality test is that GROUP1 is influenced only by itself, and not by GROUP2. If the hypothesis test fails to reject the null, then the variables listed in GROUP1 might be considered as independent variables.

For more information, see the section “VAR and VARX Modeling” on page 3092.

The following is an example of the CAUSAL statement. You specify the CAUSAL statement with the GROUP1= and GROUP2= options.
Chapter 43: The VARMAX Procedure

```plaintext
proc varmax data=one;
  model y1-y3 = x1 / p=1;
  causal group1=(x1) group2=(y1-y3);
  causal group1=(y2) group2=(y1 y3);
run;
```

The first CAUSAL statement fits the VAR(1) model by using the variables \(y_1, y_2, y_3,\) and \(x_1\) and tests the null hypothesis that \(x_1\) causes the other variables, \(y_1, y_2,\) and \(y_3,\) but the other variables do not cause \(x_1.\) The second CAUSAL statement fits the VAR(1) model by using the variables \(y_1, y_3,\) and \(y_2\) and tests the null hypothesis that \(y_2\) causes the other variables, \(y_1\) and \(y_3,\) but the other variables do not cause \(y_2.\)

### COINTEG Statement

```plaintext
COINTEG RANK=number < options > ;
```

The COINTEG statement fits the vector error correction model to the data, tests the restrictions of the long-run parameters and the adjustment parameters, and tests for weak exogeneity in the long-run parameters. The \(P=\) option in the MODEL statement specifies the autoregressive order of the VECM. Only one COINTEG statement is allowed.

The cointegrated system uses maximum likelihood estimation. If there are no moving average (MA) terms specified by the \(Q=\) option in the MODEL statement, no GARCH terms specified in the GARCH statement, and no general restrictions specified in the BOUND and RESTRICT statements, then PROC VARMAX applies the maximum likelihood analysis proposed by Johansen and Juselius (1990); Johansen (1995a, b). Otherwise, the likelihood is maximized using an optimizer whose options can be specified in the NLOPTIONS statement.

The following statements fit a VECM(2):

```plaintext
proc varmax data=one;
  model y1-y3 / p=2;
  cointeg rank=1;
run;
```

To test restrictions on \(\alpha\) and \(\beta,\) you specify the \(J=\) option and the \(H=\) option, respectively. You specify the EXOGENEITY option in the COINTEG statement for tests of weak exogeneity in the long-run parameters.

The following example of the COINTEG statement specifies tests of restrictions on \(\alpha\) and \(\beta,\) along with tests of weak exogeneity:

```plaintext
proc varmax data=one;
  model y1-y3 / p=2;
  cointeg rank=1 h=(1 0, -1 0, 0 1)
                j=(1 0, 0 0, 0 1) exogeneity;
run;
```

You must specify the following option:

```plaintext
RANK=number
```

specifies the cointegration rank of the cointegrated system. The rank of cointegration should be greater than 0 and less than the number of dependent (endogenous) variables. If \(number\) is different from
the value of the RANK= option specified in the ECM= option in the MODEL statement, the *number* specified here is used for the rank.

You can also specify the following *options* in the COINTEG statement:

**ECTREND**

specifies the restriction on the drift in the VECM. This option is used in the following cases:

- There is no separate drift in the VECM, but a constant enters only through the error correction term. For example, for VECM($p$),

\[
\Delta y_t = \alpha(\beta', \beta_0)(y_{t-1}', 1)' + \sum_{i=1}^{p-1} \Phi_i^* \Delta y_{t-i} + \epsilon_t
\]

An example of the ECTREND option follows:

```plaintext
model y1 y2 / p=2;
cointeg rank=1 ectrend;
```

- There is a separate drift and no separate linear trend in the VECM, but a linear trend enters only through the error correction term. For example, for VECM($p$),

\[
\Delta y_t = \alpha(\beta', \beta_1)(y_{t-1}', 1)' + \sum_{i=1}^{p-1} \Phi_i^* \Delta y_{t-i} + \delta_0 + \epsilon_t
\]

An example of the ECTREND option with the TREND= option follows:

```plaintext
model y1 y2 / p=2 trend=linear;
cointeg rank=1 ectrend;
```

If you specify both this option and the NSEASON option in the MODEL statement, then the NSEASON option is ignored. If you specify the NOINT option in the MODEL statement, then this option is ignored.

**EXOGENITY**

formulates the likelihood ratio tests for testing weak exogeneity in the long-run parameters. The null hypothesis is that one variable is weakly exogenous for the others.

**H=(matrix)**

specifies the restrictions $H$ on the $k \times r$ or $(k + 1) \times r$ cointegrated coefficient matrix $\tilde{\beta}$ such that $\tilde{\beta} = H\phi$, where $H$ is known and $\phi$ is unknown. If you do not specify the ECTREND option, then the cointegrated coefficient matrix $\tilde{\beta}$ is the cointegrating matrix $\beta$ and the $H$ matrix has dimension $k \times m$. If you specify the ECTREND option, then the cointegrated coefficient matrix $\tilde{\beta}$ is the cointegrating matrix $\beta$ stacked with the coefficient row vector $\beta_0$ or $\beta_1$ for the constant or linear trend in the error correction term, and the $H$ matrix has dimension $(k + 1) \times m$. Here $k$ is the number of dependent variables and $m$ is $r \leq m < k$, where $r$ is defined in the RANK=r option.

For example, consider a VECM(2) with rank equal to 1 on four dependent variables. Then, $\beta = (\beta_{11}, \beta_{21}, \beta_{31}, \beta_{41})'$. To test the null hypothesis $\beta_{11} + \beta_{21} = 0$ (that is, $H_{11} H_{11} = 0$, where $H_{11} = (1 - 1 0 0)'$), you can use the following statements to specify the restriction matrix $H$: 
model y1-y4 / p=2;
cointeg rank=1 h=(1 0 0, -1 0 0, 0 1 0, 0 0 1);

Here the dimension of matrix $H$ is $4 \times 3$ because $k = 4$ and $m = 3$, and each row of the matrix $H$ is separated by commas. Note that $H'_\perp H = 0$; that is, the $H$ and $H'_\perp$ matrices are orthogonal.

When the series has no separate deterministic trend, and therefore you specify the ECTREND option, the constant term should be restricted by $\alpha'_\perp \delta = 0$. The matrix $\alpha'_\perp$ is a $k \times (k - r)$ full-rank matrix orthogonal to $\alpha$, such that $\text{rank}(\alpha'_\perp) = k - r$ and $\alpha'_\perp \alpha = 0$. The $\tilde{\beta}$ becomes $(\beta', \beta_0)'$ or $\tilde{\beta} = (\beta_{11}, \beta_{21}, \beta_{31}, \beta_{41}, \beta_{11}^{(0)})'$. As for the previous test of $\beta_{11} + \beta_{21} = 0$ (that is, $H'_\perp \tilde{\beta} = 0$, where $H'_\perp = (1 - 1 0 0 0)'$), you can specify the restriction matrix $H$ as follows:

model y1-y4 / p=2;
cointeg rank=1 ectrend
  h=(1 0 0 0, -1 0 0 0, 0 1 0 0, 0 0 1 0, 0 0 0 1);

Because the dimension is changed in the $H'_\perp$ matrix, the dimension of $H$ matrix has to be adjusted accordingly.

When the cointegrated system contains three dependent variables and the RANK=2 option is specified, the test of $\beta_{1j} = -\beta_{2j}$ for $j = 1, 2$ can be run with the following restriction matrix $H$, where $H'_\perp = (1 1 0)'$ and $H'_\perp \tilde{\beta} = 0$:

```plaintext
cointeg rank=2 h=(1 0, -1 0, 0 1);
```

There are many ways to achieve a matrix that is orthogonal to a particular matrix. The following statements illustrate how to obtain the orthogonal matrix through QR decomposition:

```plaintext
proc iml;
/* For a given matrix H_dot, */
H_dot = {1 1 0}
/* get its QR decomposition, i.e., H_dot = QR. */
call qr(Q, R, piv, lindep, H_dot);
/* Then, the matrix orthogonal to H_dot can be extracted from Q. */
H = Q[,ncol(H_dot)+1:nrow(H_dot)];
/* Finally, normalize each column of H if necessary. */
do i = 1 to ncol(H);
  k = 0;
  do j = nrow(H) to 1 by -1;
    if (H[j,i]^=0) then k=j;
  end;
  if (k=0) then
    print "Error: H is not full rank!";
  else
    do j = nrow(H) to 1 by -1;
      H[j,i] = H[j,i] / H[k,i];
    end;
end;
end;
```
print "The given matrix is:";
print H_dot;
print "The matrix orthogonal to it is:";
print H;
quid;

J=(matrix)
specifies the restrictions \( J \) on the \( k \times r \) adjustment matrix \( \alpha \) such that \( \alpha = J \psi \), where \( J \) is known and \( \psi \) is unknown. The \( k \times m \) matrix \( J \) is specified by using this option, where \( k \) is the number of dependent variables, \( m \) is \( r \leq m < k \), and \( r \) is defined in the RANK=\( r \) option.

For example, suppose the system contains four variables, the RANK=1 option is specified, and you want to test \( \alpha_j = 0 \) for \( j = 2, 3, 4 \) — that is, \( J_\perp \alpha = 0 \), where

\[
J_\perp = \begin{pmatrix}
0 & 0 & 0 \\
1 & 0 & 0 \\
0 & 1 & 0 \\
0 & 0 & 1 \\
\end{pmatrix}
\]

Then you can specify the restriction matrix \( J \) as follows:

```
cointeg rank=1 j=(1, 0, 0, 0);
```

Suppose the system contains three variables, the RANK=2 option is specified, and you want to test \( \alpha_{2j} = 0 \) for \( j = 1, 2 \) — that is, \( J'_{\perp} \alpha = 0 \), where \( J_{\perp} = (0 1 0)' \). Then you can specify the restriction matrix \( J \) as follows:

```
cointeg rank=2 j=(1 0, 0 0, 0 1);
```

NLC

specifies the nonlinear constraints that \( \alpha \) and \( \beta \) are full column rank. Although the constraints are required for a well-defined VECM, only the TECH=QUANEW and TECH=NMSIMP optimization methods in the NLOPTIONS statement support nonlinear constraints. The full-rank constraints are not imposed by default so that other optimization methods, such as TECH=CONGRA or TECH=TRUREG, can be tried. The NLC option works only when numerical optimization is used for estimating VECM (for example, when the BOUND, INITIAL, or RESTRICT statement is specified, or the VEC-ARMA or VEC-ARMA-GARCH model is estimated). That is, the NLC option is ignored if the closed-form solution of parameter estimates and maximum likelihood analysis, which is provided in Johansen and Juselius (1990) and Johansen (1995a, b), can be applied.

NORMALIZE=variable

specifies a single dependent (endogenous) variable whose cointegrating vectors are normalized. If the variable is different from the variable specified in the COINTTEST=(JOHANSEN=) or ECM= option in the MODEL statement, the variable in this option is used. If this option is not specified, cointegrating vectors are not normalized.

If the EXOGENEITY, H=, J=, or NORMALIZE= option is specified, the BOUND, GARCH, INITIAL, and RESTRICT statements are all ignored, and the Q= option in the MODEL statement is also ignored.
The GARCH statement specifies a GARCH-type multivariate conditional heteroscedasticity model.

You can specify the following options:

**CORRCONSTANT=ESTIMATE | EXPECT**

specifies how to calculate the constant or unconditional correlation matrix in the CCC or DCC GARCH model, respectively. If you specify CORRCONSTANT=EXPECT, the constant conditional correlation matrix in the CCC GARCH model or the unconditional correlation matrix in the DCC GARCH model is calculated through the standardized residuals, given the other parameters. After parameter estimates are output, the constant or unconditional correlation matrix for the CCC or DCC GARCH model is output in the CCCCorrConstant or DCCCorrConstant ODS table, respectively. If you specify CORRCONSTANT=ESTIMATE, the correlation matrix is estimated like all other parameters in the model. By default, CORRCONSTANT=ESTIMATE.

**FORM=value**

specifies the representation for a GARCH model. Valid values are as follows:

- **BEKK** specifies a BEKK representation. This is the default.
- **CCC** specifies a constant conditional correlation representation.
- **DCC** specifies a dynamic conditional correlation representation.

**OUTHT=SAS-data-set**

writes the conditional covariance matrix to an output data set.

**P=number**

**P=(number-list)**

specifies the order of the process or the subset of GARCH terms to be fitted. For example, you can specify the P=(1,3) option. The P=3 option is equivalent to the P=(1,2,3) option. By default, P=0.

**Q=number**

**Q=(number-list)**

specifies the order of the process or the subset of ARCH terms to be fitted. This option is required in the GARCH statement. For example, you can specify the Q=(2) option. The Q=2 option is equivalent to the Q=(1,2) option.

**SUBFORM=value**

specifies the type of the univariate GARCH model for each innovation in the CCC or DCC GARCH model. If you specify the FORM=BEKK option, the SUBFORM= option is ignored. The values of the SUBFORM= option are as follows:

- **EGARCH** specifies the exponential GARCH, or EGARCH, model.
- **GARCH** specifies the GARCH model with no constraints.
- **GJR | TGARCH** specifies the GJR GARCH (also called threshold GARCH, or TGARCH) model.
ID Statement

ID variable INTERVAL=value < ALIGN=value> ;

The ID statement specifies a variable that identifies observations in the input data set. The datetime variable specified in the ID statement is included in the OUT= data set if the OUTPUT statement is specified. The ID variable is usually a SAS datetime variable. The values of the ID variable are extrapolated for the forecast observations based on the value of the INTERVAL= option.

ALIGN= value
controls the alignment of SAS dates used to identify output observations. The ALIGN= option allows the following values: BEGINNING | BEG | B, MIDDLE | MID | M, and ENDING | END | E. The default is BEGINNING. The ALIGN= option is used to align the ID variable to the beginning, middle, or end of the time ID interval specified by the INTERVAL= option.

INTERVAL=value
specifies the time interval between observations. This option is required in the ID statement. The INTERVAL= option is used in conjunction with the ID variable to check that the input data are in order and have no missing periods. The INTERVAL= option is also used to extrapolate the ID values past the end of the input data when the OUTPUT statement is specified.

The following is an example of the ID statement:
INITIAL Statement

INITIAL equation, . . . , equation;

The INITIAL statement sets up the initial parameter values for nonlinear optimization when the maximum likelihood method is applied to the estimation of VARMAX, VECM, VARMAX-GARCH, or VEC-ARMAX-GARCH models. Only one INITIAL statement is allowed. If you specify more than one equation, separate them with commas. The equations are specified in the same manner as the restrictions in the RESTRICT statement. For information about how to define equations by using matrix expressions, operators, and functions, see the section “RESTRICT Statement” on page 3050.

To use the INITIAL statement, you need to know the form of the model. If you do not specify the GARCH statement, the COINTEG statement, or the ECM=, P=, Q=, or XLAG= option in the MODEL statement, then the INITIAL statement is not applicable. If you specify the ECM=(NORMALIZE=), METHOD=LS, or PRIOR= option in the MODEL statement, or if you specify the EXOGENEITY, H=, J=, or NORMALIZE= option in the COINTEG statement, the INITIAL statement is ignored. Nonlinear restrictions on parameters are not supported.

The initial parameter values are the solution of the specified linear equations. If you do not specify initial values for all parameters, the default initial value for any parameter that is not specified in the INITIAL statement is 0, except for the following:

- The diagonal elements of the COV parameter matrix are set to ones if the COV parameter matrix is to be estimated.
- The diagonal elements of the GCHC parameter matrix are set to ones if the GCHC parameter matrix is to be estimated and the SUBFORM=EGARCH option is not specified.
- The diagonal elements of the PACH parameter matrix are set to ones if the SUBFORM=PGARCH option is specified.

The following is an example of the INITIAL statement for a bivariate (k=2) zero-mean VARMA(1,1) model, which is estimated by the maximum likelihood method by default because a moving average (MA) term is present:

```
proc varmax data=one;
   model y1 y2 / noint p=1 q=1;
   initial AR = 0, MA = 0,
       COV={1 0.5, 0.5 4};
run;
```

Like the RESTRICT statement, the preceding INITIAL statement can be abbreviated as follows:
initial AR = MA = 0,
    COV={1 0.5, 0.5 4};

or

initial AR, MA, COV={1 0.5, 0.5 4};

Furthermore, you can omit AR and MA in the INITIAL statement as follows, because by default the AR and MA matrices are set to zeros if they are not specified in the INITIAL statement:

    initial COV={1 0.5, 0.5 4};

If you use the INITIAL statement for a vector error correction model (VECM), you must specify initial values for both the ALPHA and BETA matrices and make sure they are both full rank; otherwise, the INITIAL statement is ignored.

In the following example, the INITIAL statement is ignored because initial values for ALPHA and BETA are not specified:

    proc varmax data=one;
    model y1 y2 / noint p=1;
    cointeg rank=1;
    initial cov=I(2)*4;
    run;

In the following example, the INITIAL statement is ignored because initial values for ALPHA are not specified:

    proc varmax data=one;
    model y1 y2 / noint p=1;
    cointeg rank=1;
    initial beta=1;
    run;

In the following example, the INITIAL statement is ignored because the initial BETA matrix is not full rank:

    proc varmax data=one;
    model y1 y2 y3 / noint p=1;
    cointeg rank=2;
    initial alpha={1 0, 0 1, 0 0},
        beta ={1 2, 2 4, 3 6};
    run;

In the following example, the INITIAL statement works fine because the specified initial ALPHA and BETA matrices are both full rank:

    proc varmax data=one;
    model y1 y2 y3 / noint p=1;
    cointeg rank=2;
    initial alpha={1 0, 0 1, 0 0},
        beta ={1 2, 2 4, 3 5};
    run;
MODEL Statement

MODEL  dependents < = regressors >
   < , dependents < = regressors > ... >
   < / options > ;

The MODEL statement specifies dependent (endogenous) variables and independent (exogenous) variables for the VARMAX model. The multivariate model can have the same or different independent variables corresponding to the dependent variables. As a special case, the VARMAX procedure allows you to analyze one dependent variable. Only one MODEL statement is allowed.

For example, the following statements are equivalent ways of specifying the multivariate model for the vector $(y_1, y_2, y_3)$:

```plaintext
model y1 y2 y3 </options>;
model y1-y3 </options>;
```

The following statements are equivalent ways of specifying the multivariate model with independent variables, where $y_1, y_2, y_3,$ and $y_4$ are the dependent variables and $x_1, x_2, x_3, x_4,$ and $x_5$ are the independent variables:

```plaintext
model y1 y2 y3 y4 = x1 x2 x3 x4 x5 </options>;
model y1 y2 y3 y4 = x1-x5 </options>;
model y1 = x1-x5, y2 = x1-x5, y3 y4 = x1-x5 </options>;
model y1-y4 = x1-x5 </options>;
```

When the multivariate model has different independent variables that correspond to each of the dependent variables, equations are separated by commas (,) and the model can be specified as illustrated by the following MODEL statement:

```plaintext
model y1 = x1-x3, y2 = x3-x5, y3 y4 = x1-x5 </options>;
```

The following options can be used in the MODEL statement after a forward slash (/):

**CENTER**

centers the dependent (endogenous) variables by subtracting their means. Note that centering is done after differencing when the DIF= or DIFY= option is specified. If there are exogenous (independent) variables, this option is not applicable.

```plaintext
model y1 y2 / p=1 center;
```

**DIF(variable (number-list) < ... variable (number-list)>)**

**DIF=(variable (number-list) < ... variable (number-list)>)**

specifies the degrees of differencing to be applied to the specified dependent or independent variables. The number-list must contain one or more numbers, each of which should be greater than zero. The differencing can be the same for all variables, or it can vary among variables. For example, the DIF=$(y_1(1,4)\ y_3(1)\ x_2(2))$ option specifies that the series $y_1$ is differenced at lag 1 and at lag 4, which is

$$(1 - B^4)(1 - B)y_{1t} = (y_{1t} - y_{1,t-1}) - (y_{1,t-4} - y_{1,t-5})$$
the series $y_3$ is differenced at lag 1, which is $(y_{3t} - y_{3,t-1})$; and the series $x_2$ is differenced at lag 2, which is $(x_{2t} - x_{2,t-2})$.

The following uses the data $dy_1, y_2, x_1$, and $dx_2$, where $dy_1 = (1 - B)y_{1t}$ and $dx_2 = (1 - B)^2 x_{2t}$:

```
model y1 y2 = x1 x2 / p=1 dif=(y1(1) x2(2));
```

**DIFX(number-list)**

specifies the degrees of differencing to be applied to all independent variables. The *number-list* must contain one or more numbers, each of which should be greater than zero. For example, the DIFX=(1) option specifies that all of the independent series are differenced once at lag 1. The DIFX=(1,4) option specifies that all of the independent series are differenced at lag 1 and at lag 4. If independent variables are specified in the DIF= option, then the DIFX= option is ignored.

The following statement uses the data $y_1, y_2, dx_1$, and $dx_2$, where $dx_1 = (1 - B)x_{1t}$ and $dx_2 = (1 - B)^2 x_{2t}$:

```
model y1 y2 = x1 x2 / p=1 difx(1);
```

**DIFY(number-list)**

specifies the degrees of differencing to be applied to all dependent (endogenous) variables. The *number-list* must contain one or more numbers, each of which should be greater than zero. For more information, see the DIFX= option. If dependent variables are specified in the DIF= option, then the DIFY= option is ignored.

```
model y1 y2 / p=1 dify(1);
```

**FI**

specifies that the vector autoregressive fractionally integrated moving average model with exogenous variables will be used.

```
model y1 y2 / fi method = ML;
```

**METHOD=value**

requests the type of estimates to be computed. The possible values of the METHOD= option are as follows:

- **LS**: specifies least squares estimates.
- **ML**: specifies maximum likelihood estimates.
- **CML**: specifies conditional maximum likelihood estimates.

For VARX models, you can apply least squares method, maximum likelihood method, or conditional maximum likelihood method; for VARMAX models, you can apply either maximum likelihood method or conditional maximum likelihood method; for other type of models, namely, vector error correction models, GARCH models, and Bayesian models, the default maximum likelihood method is applied.
model y1 y2 / p=1 method=ml;

**NOCURRENTX**
suppresses the current values \(x_t\) of the independent variables. In general, the \(\text{VARX}(p,s)\) model is

\[
y_t = \delta + \sum_{i=1}^{p} \Phi_i y_{t-i} + \sum_{i=0}^{s} \Theta_i x_{t-i} + \epsilon_t
\]

where \(p\) is the number of lags of the dependent variables included in the model, and \(s\) is the number of lags of the independent variables included in the model, including the contemporaneous values of \(x_t\).

A \(\text{VARX}(1,2)\) model can be specified as:

model y1 y2 = x1 x2 / p=1 xlag=2;

If the NOCURRENTX option is specified, it suppresses the current values \(x_t\) and starts with \(x_{t-1}\). The \(\text{VARX}(p,s)\) model is redefined as:

\[
y_t = \delta + \sum_{i=1}^{p} \Phi_i y_{t-i} + \sum_{i=1}^{s} \Theta_i x_{t-i} + \epsilon_t
\]

This model with \(p = 1\) and \(s = 2\) can be specified as:

model y1 y2 = x1 x2 / p=1 xlag=2 nocurrentx;

**NOINT**
suppresses the intercept parameter \(\delta\).

model y1 y2 / p=1 noint;

**NSEASON=number**
specifies the number of seasonal periods. When the NSEASON=number option is specified, \((\text{number} - 1)\) seasonal dummies are added to the regressors. If the NOINT option is specified, the NSEASON=option is not applicable. For more information, see the section “Seasonal Dummies and Time Trends” on page 3098.

model y1 y2 / p=1 nseason=4;

**SCENTER**
centers seasonal dummies specified by the NSEASON= option. The centered seasonal dummies are generated by \(c - (1/s)\), where \(c\) is a seasonal dummy generated by the NSEASON=s option.
MODEL Statement  ➤  3037

model y1 y2 / p=1 nseason=4 scenter;

TREND=value
specifies the degree of deterministic time trend included in the model. Valid values are as follows:

LINEAR includes a linear time trend as a regressor.
QUAD includes linear and quadratic time trends as regressors.

The TREND=QUAD option is not applicable for a cointegration analysis. For more information, see the section “Seasonal Dummies and Time Trends” on page 3098.

model y1 y2 / p=1 trend=linear;

VARDEF=value
corrects for the degrees of freedom of the denominator for computing an error covariance matrix for the METHOD=LS option. If the METHOD=ML option is specified, the VARDEF=N option is always used. Valid values are as follows:

DF specifies that the number of nonmissing observation minus the number of regressors be used.
N specifies that the number of nonmissing observation be used.

model y1 y2 / p=1 vardef=n;

Printing Control Options

LAGMAX=number
specifies the maximum number of lags for which results are computed and displayed by the PRINT=(CORRX CORRY COVX COVY IARR IMPULSE= IMPULSX= PARCOEF PCANCORR PCORR) options. This option is also used to limit the printed results for the cross covariances and cross-correlations of residuals. The default is LAGMAX=min(12, T-2), where T is the number of nonmissing observations.

model y1 y2 / p=1 lagmax=6;

NOPRINT suppresses all printed output.

model y1 y2 / p=1 noprint;
Chapter 43: The VARMAX Procedure

**PRINTALL**
requests all printing control options. The options set by the option PRINTALL are DFTEST=, MINIC=, PRINTFORM=BOTH, and PRINT=(CORRB CORRX CORRY COVB COVPE COVX COVY DECOMPOSE DYNAMIC IARR IMPULSE=(ALL) IMPULSX=(ALL) PARCOEF PCAN-CORR PCORR ROOTS YW).

You can also specify this option as the option ALL.

```plaintext
model y1 y2 / p=1 printall;
```

**PRINTFORM=** *value*
requests the printing format of the output generated by the PRINT= option and cross covariances and cross-correlations of residuals. Valid values are as follows:

- **BOTH** prints output in both MATRIX and UNIVARIATE forms.
- **MATRIX** prints output in matrix form. This is the default.
- **UNIVARIATE** prints output by variables.

```plaintext
model y1 y2 / p=1 print=(impulse) printform=univariate;
```

**Printing Options**

**PRINT=(options)**
The following options can be used in the PRINT=( ) option. The options are listed within parentheses. If a number in parentheses follows an option listed below, then the option prints the number of lags specified by *number* in parentheses. The default is the number of lags specified by the LAGMAX= *number* option.

**CORRB**
prints the estimated correlations of the parameter estimates.

**CORRX**
CORRX(*number*)
prints the cross-correlation matrices of exogenous (independent) variables. The *number* should be greater than zero.

**CORRY**
CORRY(*number*)
prints the cross-correlation matrices of dependent (endogenous) variables. The *number* should be greater than zero.

**COVB**
prints the estimated covariances of the parameter estimates.
COVPE

**COVPE**(number)

prints the covariance matrices of number-ahead prediction errors for the VARMAX(p,q,s) model. The number should be greater than zero. If the DIF= or DIFY= option is specified, the covariance matrices of multistep prediction errors are computed based on the differenced data. This option is not applicable when the PRIOR= option is specified. For more information, see the section “Forecasting” on page 3082.

COVX

**COVX**(number)

prints the cross-covariance matrices of exogenous (independent) variables. The number should be greater than zero.

COVY

**COVY**(number)

prints the cross-covariance matrices of dependent (endogenous) variables. The number should be greater than zero.

DECOMPOSE

**DECOMPOSE**(number)

prints the decomposition of the prediction error covariances using up to the number of lags specified by number in parentheses for the VARMA(p,q) model. The number should be greater than zero. It can be interpreted as the contribution of innovations in one variable to the mean squared error of the multistep forecast of another variable. The DECOMPOSE option also prints proportions of the forecast error variance.

If the DIF= or DIFY= option is specified, the covariance matrices of multistep prediction errors are computed based on the differenced data. This option is not applicable when the PRIOR= option is specified. For more information, see the section “Forecasting” on page 3082.

DIAGNOSE

prints the residual diagnostics and model diagnostics.

DYNAMIC

prints the contemporaneous relationships among the components of the vector time series.

ESTIMATES

prints the coefficient estimates and a schematic representation of the significance and sign of the parameter estimates.

IARR

**IARR**(number)

prints the infinite order AR representation of a VARMA process. The number should be greater than zero. If the ECM= option or the COINTEG statement is specified, then the reparameterized AR coefficient matrices are printed.
IMPULSE
IMPULSE(\textit{number})
IMPULSE=(\textit{SIMPLE ACCUM ORTH STDERR ALL})
IMPULSE(\textit{number})=(\textit{SIMPLE ACCUM ORTH STDERR ALL})

prints the impulse response function. The \textit{number} should be greater than zero. It investigates the response of one variable to an impulse in another variable in a system that involves a number of other variables as well. It is an infinite order MA representation of a VARMA process. For more information, see the section “Impulse Response Function” on page 3071.

You can specify the following options within parentheses:

- \textbf{ACCUM} prints the accumulated impulse response function.
- \textbf{ALL} is equivalent to specifying \textit{SIMPLE}, \textit{ACCUM}, \textit{ORTH}, and \textit{STDERR}.
- \textbf{ORTH} prints the orthogonalized impulse response function.
- \textbf{SIMPLE} prints the impulse response function. This is the default.
- \textbf{STDERR} prints the standard errors of the impulse response function, the accumulated impulse response function, or the orthogonalized impulse response function.

IMPULSX
IMPULSX(\textit{number})
IMPULSX=(\textit{SIMPLE ACCUM STDERR ALL})
IMPULSX(\textit{number})=(\textit{SIMPLE ACCUM STDERR ALL})

prints the impulse response function related to exogenous (independent) variables. The \textit{number} should be greater than zero. For more information, see the section “Impulse Response Function” on page 3071.

You can specify the following options within parentheses:

- \textbf{ACCUM} prints the accumulated impulse response matrices for the transfer function.
- \textbf{ALL} is equivalent to specifying \textit{SIMPLE}, \textit{ACCUM}, and \textit{STDERR}.
- \textbf{SIMPLE} prints the impulse response matrices for the transfer function.
- \textbf{STDERR} prints the standard errors of the simple impulse response function or the accumulated impulse response function.

By default, IMPULSX(\textit{number})=(\textit{SIMPLE}).

PARCOEF
PARCOEF(\textit{number})

prints the partial autoregression coefficient matrices, $\Phi_{mm}$ up to the lag \textit{number}. The \textit{number} should be greater than zero. With a VAR process, this option is useful for the identification of the order since the $\Phi_{mm}$ have the property that they equal zero for $m > p$ under the hypothetical assumption of a VAR($p$) model. For more information, see the section “Tentative Order Selection” on page 3087.

PCANCORR
PCANCORR(\textit{number})

prints the partial canonical correlations of the process at lag \textit{m} and the test for testing $\Phi_{m}=0$ for $m > p$ up to the lag \textit{number}. The \textit{number} should be greater than zero. The lag \textit{m} partial canonical correlations are the canonical correlations between $y_{t}$ and $y_{t-m}$, after adjustment for the dependence of these variables on the intervening values $y_{t-1}, \ldots, y_{t-m+1}$. For more information, see the section “Tentative Order Selection” on page 3087.
PCORR

PCORR(number)

prints the partial correlation matrices. The number should be greater than zero. With a VAR process, this option is useful for a tentative order selection by the same property as the partial autoregression coefficient matrices, as described in the PRINT=(PARCOEF) option. For more information, see the section “Tentative Order Selection” on page 3087.

ROOTS

prints the eigenvalues of the $k p \times k p$ companion matrix associated with the AR characteristic function $\Phi(B)$, where $k$ is the number of dependent (endogenous) variables, and $\Phi(B)$ is the finite order matrix polynomial in the backshift operator $B$, such that $B^i y_t = y_{t-i}$. These eigenvalues indicate the stationary condition of the process since the stationary condition on the roots of $|\Phi(B)| = 0$ in the VAR($p$) model is equivalent to the condition in the corresponding VAR(1) representation that all eigenvalues of the companion matrix be less than one in absolute value. Similarly, you can use this option to check the invertibility of the MA process. In addition, when the GARCH statement is specified, this option prints the roots of the GARCH characteristic polynomials to check covariance stationarity for the GARCH process.

YW

prints Yule-Walker estimates of the preliminary autoregressive model for the dependent (endogenous) variables. The coefficient matrices are printed using the maximum order of the autoregressive process.

Some examples of the PRINT= option are as follows:

   model y1 y2 / p=1 print=(covy(10) corry(10));
   model y1 y2 / p=1 print=(parcoef pcancorr pcorr);
   model y1 y2 / p=1 print=(impulse(8) decompose(6) covpe(6));
   model y1 y2 / p=1 print=(dynamic roots yw);

Lag Specification Options

P=number

P=(number-list)

specifies the order of the vector autoregressive process. Subset models of vector autoregressive orders can be specified by listing the desired set of lags. For example, you can specify the P=(1,3,4) option. The P=3 option is equivalent to the P=(1,2,3) option. The default is P=0.

If P=0 and there are no exogenous (independent) variables, then the AR polynomial order is automatically determined by minimizing an information criterion. If P=0 and the PRIOR= or ECM= option or COINTEG statement are specified, then the AR polynomial order is determined automatically.

If the ECM= option or the COINTEG statement is specified, then subset models of vector autoregressive orders are not allowed and the AR maximum order specified is used.

Examples illustrating the P= option follow:

   model y1 y2 / p=3;
   model y1 y2 / p=(1,3);
   model y1 y2 / p=(1,3) prior;
**Q=number**

**Q=(number-list)**

specifies the order of the moving-average error process. Subset models of moving-average orders can be specified by listing the desired set of lags. For example, you can specify the Q=(1,5) option. The default is Q=0.

```plaintext
model y1 y2 / p=1 q=1;
model y1 y2 / q=(2);
```

**XLAG=number**

**XLAG=(number-list)**

specifies the lags of exogenous (independent) variables. Subset models of distributed lags can be specified by listing the desired set of lags. For example, XLAG=(2) selects only a lag 2 of the exogenous variables. The default is XLAG=0. To exclude the present values of exogenous variables from the model, the NOCURRENTX option must be used.

```plaintext
model y1 y2 = x1-x3 / xlag=2 nocurrentx;
model y1 y2 = x1-x3 / p=1 xlag=(2);
```

**Tentative Order Selection Options**

**MINIC**

**MINIC=( P=number PERROR=number Q=number TYPE=value )**

prints the information criterion for the appropriate AR and MA tentative order selection.

You can specify the following options within parentheses in the MINIC= option:

**P=number**

**P=(p_{min} : p_{max})**

specifies the range of AR orders to be considered in the tentative order selection. The default is P=(0:5). P=3 is equivalent to P=(0:3).

**PERROR=number**

**PERROR=(p_{e, min} : p_{e, max})**

specifies the range of AR orders for obtaining the error series. The default is PERROR=(p_{max} : p_{max} + q_{max}).

**Q=number**

**Q=(q_{min} : q_{max})**

specifies the range of MA orders to be considered in the tentative order selection. The default is Q=(0:5).

**TYPE= AIC | AICC | FPE | HQC | SBC**

specifies the criterion for the model order selection. Valid criteria are as follows:
\textbf{AIC} specifies Akaike’s information criterion.

\textbf{AICC} specifies the corrected Akaike’s information criterion.

\textbf{FPE} specifies the final prediction error criterion.

\textbf{HQC} specifies the Hanna-Quinn criterion.

\textbf{SBC} specifies the Schwarz Bayesian criterion. You can also specify this value as TYPE=BIC.

By default, TYPE=AICC.

The following examples show how to use the MINIC or MINIC= option:

```
model y1 y2 / minic;
```

```
model y1 y2 / minic=(type=aic p=13);
```

In the selection of AR and MA orders, the model that has the smallest criterion value is chosen. For the definitions of the information criteria used in the MINIC option, see the section “The Minimum Information Criterion (MINIC) Method” on page 3091.

\section*{Cointegration Related Options}

Two options are related to integrated time series; one is the DFTEST option to test for a unit root and the other is the COINTTEST option to test for cointegration.

\section*{DFTEST
\textbf{DFTEST}=(\textbf{DLAG}=number)
\textbf{DFTEST}=(\textbf{DLAG}=(number) . . . (number))
prints the Dickey-Fuller unit root tests. The DLAG=(number) . . . (number) option specifies the regular or seasonal unit root test. Supported values of number are in 1, 2, 4, 12. If the number is greater than one, a seasonal Dickey-Fuller test is performed. If the TREND= option is specified, the seasonal unit root test is not available. The default is DLAG=1.

For example, the DFTEST=(DLAG=(1)(12)) option produces two tables: the Dickey-Fuller regular unit root test and the seasonal unit root test.

Some examples of the DFTEST= option follow:

```
model y1 y2 / p=2 dftest;
model y1 y2 / p=2 dftest=(dlag=4);
model y1 y2 / p=2 dftest=(dlag=(1)(12));
model y1 y2 / p=2 dftest cointtest;
```

\textbf{COINTTEST
\textbf{COINTTEST}=(\textbf{JOHANSEN} < (=options) > \textbf{SW} < (=options) > \textbf{SIGLEVEL}=number )
specifies the cointegration tests.

You can specify the following options within parentheses in the COINTTEST= option:
JOHANSEN

JOHANSEN=(TYPE=value IORDER=number NORMALIZE=variable)

prints the cointegration rank test for multivariate time series based on Johansen’s method. This
test is provided when the number of dependent (endogenous) variables is less than or equal to 64.
For more information, see the section “Vector Error Correction Modeling” on page 3115.

The VARX(p,s) model can be written as the error correction model

\[ \Delta y_t = \Pi y_{t-1} + \sum_{i=1}^{p-1} \Phi^*_i \Delta y_{t-i} + AD_t + \sum_{i=0}^{s} \Theta^*_i x_{t-i} + \epsilon_t \]

where \( \Pi \), \( \Phi^*_i \), \( A \), and \( \Theta^*_i \) are coefficient parameters and \( D_t \) is a deterministic term such as a
constant, a linear trend, or seasonal dummies.

The \( I.1/ \) model is defined by one reduced-rank condition. If the cointegration rank is
\( r < k \), then there exist \( k \times r \) matrices \( \alpha \) and \( \beta \) of rank \( r \) such that
\( \Pi = \alpha \beta' \).

The \( I.1/ \) model is rewritten as the \( I.2/ \) model

\[ \Delta^2 y_t = \Psi \Delta y_{t-1} + \sum_{i=1}^{p-2} \Psi_i \Delta^2 y_{t-i} + AD_t + \sum_{i=0}^{s} \Theta^*_i x_{t-i} + \epsilon_t \]

where \( \Psi = I_k - \sum_{i=1}^{p-1} \Phi^*_i \) and \( \Psi_i = -\sum_{j=i+1}^{p-1} \Phi^*_i \).

The \( I.2/ \) model is defined by two reduced-rank conditions. One is that \( \Pi = \alpha \beta' \), where \( \alpha \) and \( \beta \) are \( k \times r \) matrices of full-rank \( r \). The other is that \( \alpha \perp \Psi \beta \perp = \xi \eta' \), where \( \xi \) and \( \eta \) are \( (k-r) \times s \) matrices with \( s \leq k-r \), and \( \alpha \perp \) and \( \beta \perp \) are \( k \times (k-r) \) matrices of full-rank \( k-r \) such that
\( \alpha' \alpha \perp = 0 \) and \( \beta' \beta \perp = 0 \).

You can specify the following options within parentheses in the JOHANSEN= option:

**IORDER=1 | 2**

specifies the integrated order. You can specify the following values:

1

prints the cointegration rank test for an integrated order 1 and prints the
long-run parameter, \( \beta \), and the adjustment coefficient, \( \alpha \). If you specify
IORDER=1, then the AR order should be greater than or equal to 1. If
you specify P=0 in the MODEL statement, the value of P is set to 1 for
the Johansen test.

2

prints the cointegration rank test for integrated orders 1 and 2. If you
specify IORDER=2, then the AR order should be greater than or equal to 2. If you specify P=1 and IORDER=2, then the value of IORDER is set
to 1; if you specify P=0 and IORDER=2, then the value of P is set to 2.

By default, IORDER=1.

**NORMALIZE=variable**

specifies the dependent (endogenous) variable whose cointegration vectors are to be normal-
ized. If the variable is different from the variable specified in the COINTEG statement or
in the ECM= option in the MODEL statement, then the value specified in the COINTEG
statement is used. If you specify this option and you want to estimate an error correction
model, then the BOUND, GARCH, INITIAL, and RESTRICT statements are all ignored and the Q= option in the MODEL statement is also ignored.

**TYPE=MAX | TRACE**

specifies the type of cointegration rank test to be printed. You can specify the following values:

- **MAX**
  - prints the cointegration maximum eigenvalue test.
- **TRACE**
  - prints the cointegration trace test.

By default, TYPE=TRACE. If the NOINT option is not specified, PROC VARMAX prints two different cointegration rank tests in the presence of the unrestricted and restricted deterministic terms (constant or linear trend) models. If you specify IORDER=2, the procedure automatically sets the TYPE=TRACE option.

The following examples illustrate the JOHANSEN= option:

```model y1 y2 / p=2 cointest=(johansen=(type=max normalize=y1));

model y1 y2 / p=2 cointest=(johansen=(iorder=2 normalize=y1));
```

**SIGLEVEL=value**

sets the size (the significance level) of the common trends tests. The SIGLEVEL=value can be set to 0.1, 0.05, or 0.01. By default, SIGLEVEL=0.05.

```model y1 y2 / p=2 cointest=(sw siglevel=0.1);

model y1 y2 / p=2 cointest=(sw siglevel=0.01);
```

**SW**

**SW=(TYPE=value LAG=number)**

prints common trends tests for a multivariate time series based on the Stock-Watson method. This test is provided when the number of dependent (endogenous) variables is less than or equal to 6. For more information, see the section “Common Trends” on page 3113.

You can specify the following options within parentheses in the SW= option:

**LAG=number**

specifies the number of lags. The default is LAG=max(1,p) for the TYPE=FILTDIF or TYPE=FILTRES option, where p is the AR maximum order specified by the P= option. The default is LAG=T^{1/4} for the TYPE=KERNEL option, where T is the number of nonmissing observations. If the specified LAG=number exceeds the default, then it is replaced by the default.

**TYPE=FILTDIF | FILTRES | KERNEL**

specifies the type of common trends test to be printed. You can specify the following values:
FILTDIF prints the common trends test based on the filtering method applied to the differenced series.

FILTRES prints the common trends test based on the filtering method applied to the residual series.

KERNEL prints the common trends test based on the kernel method.

By default, TYPE=FILTDIF.

The following examples illustrate the SW option:

```plaintext
model y1 y2 / p=2 cointtest=(sw);
model y1 y2 / p=2 cointtest=(sw=(type=kernel));
model y1 y2 / p=2 cointtest=(sw=(type=kernel lag=3));
```

Bayesian VARX Estimation Options

**PRIOR**

PRIOR=(prior-options)

specifies the prior value of parameters for the BVARX(p, s) model. The BVARX model allows for a subset model specification. If the ECM= option or the COINTEG statement is specified with the PRIOR option, the BVECMX(p, s) form is fitted. When the PRIOR option is specified, the Q= option in the MODEL statement is ignored, and the BOUND, GARCH, INITIAL, RESTRICT, and TEST statements are all ignored. For more information, see the section “Bayesian VAR and VARX Modeling” on page 3099.

The following options can be used with the PRIOR=(prior-options) option. The prior-options are listed within parentheses.

**IVAR**

IVAR=(variables)

specifies an integrated BVAR(p) model. The variables should be specified in the MODEL statement as dependent variables. If you use the IVAR option without variables, then it sets the overall prior mean of the first lag of each variable equal to one in its own equation and sets all other coefficients to zero. If variables are specified, it sets the prior mean of the first lag of the specified variables equal to one in its own equation and sets all other coefficients to zero. When the series $y_t = (y_1, y_2)'$ follows a bivariate BVAR(2) process, the IVAR or IVAR=(y1, y2) option is equivalent to specifying MEAN=(1 0 0 0 0 1 0 0).

If the PRIOR=(MEAN=) or ECM= option or the COINTEG statement is specified, the IVAR= option is ignored.

**LAMBDA=value**

specifies the prior standard deviation of the AR coefficient parameter matrices. It should be a positive number. The default is LAMBDA=1. As the value of the LAMBDA= option is increased, the BVAR(p) model becomes closer to a VAR(p) model.
**MEAN=(vector)**

specifies the mean vector in the prior distribution for the AR coefficients. If the vector is not specified, the prior value is assumed to be a zero vector. For more information, see the section “Bayesian VAR and VARX Modeling” on page 3099.

You can specify the mean vector by order of the equation. Let \( \delta, \Phi_1, \ldots, \Phi_P \) be the parameter sets to be estimated and \( \Phi = (\Phi_1, \ldots, \Phi_P) \) be the AR parameter sets. The mean vector is specified by row-wise from \( \Phi \); that is, the MEAN=(vec(\( \Phi' \))) option.

For the PRIOR=(mean) option in the BVAR(2),

\[
\Phi = \begin{pmatrix}
\phi_{1,11} & \phi_{1,12} & \phi_{2,11} & \phi_{2,12} \\
\phi_{1,21} & \phi_{1,22} & \phi_{2,21} & \phi_{2,22}
\end{pmatrix} = \begin{pmatrix}
2 & 0.1 & 1 & 0 \\
0.5 & 3 & 0 & -1
\end{pmatrix}
\]

where \( \phi_{l,ij} \) is an element of \( \Phi \), \( l \) is a lag, \( i \) is associated with the first dependent variable, and \( j \) is associated with the second dependent variable.

```plaintext
model y1 y2 / p=2 prior=(mean=(2 0.1 1 0 0.5 3 0 -1));
```

The deterministic terms and exogenous variables are considered to shrink toward zero; you must omit prior means of exogenous variables and deterministic terms such as a constant, seasonal dummies, or trends.

For a Bayesian error correction model estimated when both the ECM= option (or the COINTEG statement) and the PRIOR= option are used, a mean vector for only lagged AR coefficients, \( \Phi_t^* \), in terms of regressors \( \Delta y_{t-i} \), for \( i = 1, \ldots, (p-1) \) is used in the VECM(\( p \)) representation. The diffused prior variance of \( \alpha \) is used, since \( \beta \) is replaced by \( \hat{\beta} \) estimated in a nonconstrained VECM(\( p \)) form.

\[
\Delta y_t = \alpha z_{t-1} + \sum_{i=1}^{p-1} \Phi_t^* \Delta y_{t-i} + AD_t + \sum_{i=0}^{s} \Theta_t^* x_{t-i} + \epsilon_t
\]

where \( z_t = \beta' y_t \).

For example, in the case of a bivariate \(( k = 2 )\) BVECMTM(2) form, the option

\[
\text{MEAN} = (\phi_{1,11}^* \phi_{1,12}^* \phi_{1,21}^* \phi_{1,22}^*)
\]

where \( \phi_{1,ij}^* \) is the \(( i, j) \) element of the matrix \( \Phi_t^* \).

**NREP=number**

determines the number of repetitions that are used to compute the measure of forecast accuracy. For more information, see the equation in the section “Forecasting of BVAR Modeling” on page 3101. The default is NREP=0.5\( T \), where \( T \) is the number of observations. If NREP is above 0.5\( T \), it is decreased to 0.5\( T \); if NREP is below the value of the LEAD= option, it is increased to the value of the LEAD= option.

**THETA=value**

specifies the prior standard deviation of the AR coefficient parameter matrices. The \textit{value} is in the interval \((0,1)\). The default is THETA=0.1. As the value of the THETA= option approaches 1, the specified BVAR(\( p \)) model approaches a VAR(\( p \)) model.

Some examples of the PRIOR= option follow:
model y1 y2 / p=2 prior;
model y1 y2 / p=2 prior=(theta=0.2 lambda=5);
model y1 y2 = x1 / p=2 prior=(theta=0.2 lambda=5);
model y1 y2 = x1 / p=2
  prior=(theta=0.2 lambda=5 mean=(2 0.1 0 0.5 3 0 -1));

For more information, see the section “Bayesian VAR and VARX Modeling” on page 3099.

Vector Error Correction Model Options

*ECM=(RANK=number <ECTREND> <NORMALIZE=variable>)*

specifies a vector error correction model.

The ECM= option is obsolete. Use the COINTEG statement instead.

You must specify the following option within parentheses in the ECM= option:

**RANK=number**

specifies the cointegration rank of the cointegrated system. The rank of cointegration should be greater than 0 and less than the number of dependent (endogenous) variables. If *number* is different from the RANK= option specified in the COINTEG statement, the value specified in the COINTEG statement is used for the rank.

You can also specify the following options within parentheses in the ECM= option:

**ECTREND**

specifies the restriction on the drift in the VECM. This option is used in the following cases:

- There is no separate drift in the VECM, but a constant enters only through the error correction term. For example, for VECM(*p*),

\[
\Delta y_t = \alpha (\beta', \beta_0) (y'_{t-1}, 1)' + \sum_{i=1}^{p-1} \Phi_i^* \Delta y_{t-i} + \epsilon_t
\]

An example of the ECTREND option follows:

```plaintext
model y1 y2 / p=2 ecm=(rank=1 ectrend);
```

- There is a separate drift and no separate linear trend in the VECM, but a linear trend enters only through the error correction term. For example, for VECM(*p*),

\[
\Delta y_t = \alpha (\beta', \beta_1) (y'_{t-1}, t)' + \sum_{i=1}^{p-1} \Phi_i^* \Delta y_{t-i} + \delta_0 + \epsilon_t
\]

An example of the ECTREND option with the TREND= option follows:

```plaintext
model y1 y2 / p=2 ecm=(rank=1 ectrend) trend=linear;
```
If you specify both this option and the NSEASON option in the MODEL statement, then the NSEASON option is ignored. If you specify the NOINT option in the MODEL statement, then this option is ignored.

NORMALIZE=variable
specifies a single dependent (endogenous) variable whose cointegrating vectors are normalized. If the variable is different from the variable specified in the NORMALIZE= option in the COINTEG statement, the variable specified in the NORMALIZE= option in the COINTEG statement is used. If this option is not specified, cointegrating vectors are not normalized. If you specify this option, then the BOUND, GARCH, INITIAL, and RESTRICT statements are all ignored and the Q= option in the MODEL statement is also ignored.

The following examples illustrate the ECM= option:

```
model y1 y2 / p=2 ecm=(rank=1 normalize=y1);

model y1 y2 / p=2 ecm=(rank=1 ectrend) trend=linear;
```

For more information, see the section “Vector Error Correction Modeling” on page 3115.

---

**NLOPTONS Statement**

**NLOPTONS options ;**

The VARMAX procedure uses the nonlinear optimization (NLO) subsystem to perform nonlinear optimization tasks. For a list of all the options in the NLOPTONS statement, see Chapter 6, “Nonlinear Optimization Methods.”

An example of the NLOPTONS statement is as follows:

```
proc varmax data=one;
   nloptions tech=qn maxit=1000 pall;
   model y1 y2 / p=2;
run;
```

By default, the VARMAX procedure uses the dual quasi-Newton optimization method.

---

**OUTPUT Statement**

**OUTPUT < options > ;**

The OUTPUT statement generates and prints forecasts based on the model estimated in the previous MODEL statement and, optionally, creates an output SAS data set that contains these forecasts.

When the GARCH model is estimated, the upper and lower confidence limits of forecasts are calculated according to the conditional covariance of errors.
**ALPHA=** *number*
sets the forecast confidence limit size, where *number* is between 0 and 1. When you specify the ALPHA= *number* option, the upper and lower confidence limits define the 100(1 − α)% confidence interval. The default is ALPHA=0.05, which produces 95% confidence intervals.

**BACK=** *number*
specifies the number of observations before the end of the data at which the multistep forecasts begin. The BACK= option value must be less than or equal to the number of observations minus the number of lagged regressors in the model. The default is BACK=0, which means that the forecasts start at the end of the available data.

**LEAD=** *number*
specifies the number of multistep forecast values to compute. The default is LEAD=12.

**NOPRINT**
suppresses the printed forecast values of each dependent (endogenous) variable.

**OUT=** *SAS-data-set*
writes the forecast values to an output data set. If the OUT= option is not included in the OUTPUT statement, then the output data set is named using the DATA*_n* naming convention.

Some examples of the OUTPUT statements follow:

```
proc varmax data=one;
   model y1 y2 / p=2;
   output lead=6 back=2;
run;
```

```
proc varmax data=one;
   model y1 y2 / p=2;
   output out=for noprint;
run;
```

**RESTRICT Statement**

```
RESTRICT restriction, . . . , restriction ;
```

The RESTRICT statement places linear restrictions on the parameters and provides constrained estimation. Only one RESTRICT statement is allowed. If you specify more than one *restriction* in a RESTRICT statement, separate them with commas. Both equality and inequality constraints are allowed in the RESTRICT statement, although usually equality constraints are specified in the RESTRICT statement and inequality constraints are specified in the BOUND statement. If the least squares method is used, the inequality constraints are not applicable.

To use the RESTRICT statement, you need to know the form of the model. If you do not specify the GARCH statement, the COINTEG statement, or the ECM=, P=, Q=, or XLAG= option in the MODEL statement then the RESTRICT statement is not applicable. If you specify the ECM=(NORMALIZE=) option or PRIOR= option in the MODEL statement or if you specify the EXOGENEITY, H=, J=, or NORMALIZE= option in
the COINTEG statement, then the RESTRICT statement is ignored. Nonlinear restrictions on parameters are not supported.

Restricted parameter estimates are computed by introducing a Lagrangian parameter for each restriction (Pringle and Rayner 1971). The Lagrangian parameter measures the sensitivity of the sum of squared errors to the restriction. The estimates of these Lagrangian parameters and their significance are printed in the Restrict ODS table.

**Matrix Expression**

The RESTRICT statement operates on matrices. That is, you can specify the parameter matrices or constant matrices through the RESTRICT statement’s built-in operators and functions. You can add elements of the matrices \( A \) and \( B \) with the expression \( A + B \), and you can perform matrix multiplication with the expression \( A \times B \) and elementwise multiplication with the expression \( A \# B \). You can get the diagonal elements of the matrix \( A \) through the function \( \text{DIAG}(A) \), and you can get the \( n \times n \) identity matrix through the function \( I(n) \).

Each restriction is written as a matrix expression composed of constants, operators, and functions.

**Constants**

Constants are either scalar constants (such as \(-1.2, 0.3, \) and so on) or matrix constants enclosed in braces (such as the \(2 \times 2\) matrix, \(\{1 2, 3 4\}\), or the \(1 \times 3\) row vector, \(\{-0.2 5.3 12\}\)). Constants also include the dependent variable names and exogenous variable names that represent their index values and are mostly used in the subscripts or function arguments. For example, in the following PROC VARMAX statements, the dependent and exogenous variables have the following index values (based on their orders in the MODEL statement): GDP is equal to 1, CPI to 2, M2 to 3, FFR to 1, and CP to 2. Hence, the function call \( \text{AR}(2, \text{GDP}, \{\text{CPI M2}\}) \) is equivalent to \( \text{AR}(2,1,{2 3}) \), and \( \text{XL}(0, \text{CPI}, \{\text{FFR CP}\}) \) is equivalent to \( \text{XL}(0,2,{1 2}) \). For more information about the use of \( \text{AR} \) and \( \text{XL} \) functions to access parameters, see the section “Functions” on page 3053.

```plaintext
proc varmax data=macrodata;
  model GDP CPI M2 = FFR CP / p=12 xlag=12;
  restrict AR(2, GDP, {CPI M2}) = 0,
       XL(0, CPI, {FFR CP}) = 0;
run;
```

The matrix constant cannot be the first item in the RESTRICT statement. For example, you cannot specify the following statement:

```plaintext
restrict {-0.1 -0.2, -0.3 -0.4} <= AR <= {0.1 0.2, 0.3 0.4};
```

However, you can put the first matrix constant in parentheses and specify the preceding example in the following way:

```plaintext
restrict ((-0.1 -0.2, -0.3 -0.4)) <= AR <= (0.1 0.2, 0.3 0.4);
```

**Operators**

Operators define the operations on operands. Table 43.2 lists all built-in operators supported by the RESTRICT statement.
Table 43.2 Operators

<table>
<thead>
<tr>
<th>Operator Name</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>+</td>
<td>Addition</td>
</tr>
<tr>
<td>=</td>
<td>Comparison, equal</td>
</tr>
<tr>
<td>&lt;</td>
<td>Comparison, less than</td>
</tr>
<tr>
<td>&lt;=</td>
<td>Comparison, not greater than</td>
</tr>
<tr>
<td>&gt;</td>
<td>Comparison, greater than</td>
</tr>
<tr>
<td>&gt;=</td>
<td>Comparison, not less than</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>//</td>
<td>Concatenation, vertical</td>
</tr>
<tr>
<td>@</td>
<td>Direct product</td>
</tr>
<tr>
<td>:</td>
<td>Index creation</td>
</tr>
<tr>
<td>#</td>
<td>Multiplication, elementwise</td>
</tr>
<tr>
<td>*</td>
<td>Multiplication, matrix</td>
</tr>
<tr>
<td>–</td>
<td>Sign reverse</td>
</tr>
<tr>
<td>[ ]</td>
<td>Subscripts</td>
</tr>
<tr>
<td>–</td>
<td>Subtraction</td>
</tr>
<tr>
<td>`</td>
<td>Transpose</td>
</tr>
</tbody>
</table>

For more information about each operator, see the section “Details of Operators” on page 3057.

Table 43.3 shows the precedence of matrix operators in the RESTRICT statement.

Table 43.3 Operator Precedence

<table>
<thead>
<tr>
<th>Priority Group</th>
<th>Operators</th>
</tr>
</thead>
<tbody>
<tr>
<td>I (highest)</td>
<td>[ ] (subscripts) ` (transpose)</td>
</tr>
<tr>
<td>II</td>
<td>~ (sign reverse) #</td>
</tr>
<tr>
<td>III</td>
<td>* @</td>
</tr>
<tr>
<td>IV</td>
<td>– (subtraction) +</td>
</tr>
<tr>
<td>V</td>
<td></td>
</tr>
<tr>
<td>VI (lowest)</td>
<td>= &lt; &lt;= &gt; &gt;=</td>
</tr>
</tbody>
</table>

Each restriction can be a compound expression that involves several matrix operators and operands. The rules for evaluating compound expressions are as follows:

- Evaluation follows the order of operator precedence, as described in Table 43.3. Group I has the highest priority; that is, Group I operators are evaluated first. Group II operators are evaluated after Group I operators, and so on. For example, $1 + 2 \times 3$ returns 7.

- If neighboring operators in an expression have equal precedence, the expression is evaluated from left to right, except for the Group I operators. For example, $1 - 2 - 3$ returns $-4$.

- All expressions in parentheses are evaluated first, following the two preceding rules. For example, $3 \times (2 + 1)$ returns 9.
Functions

Functions are mainly divided into two categories: one type of function refers to parameters to be estimated, such as $\text{AR}(L, I, J)$ and $\text{CCC}(I, J)$; the other type does not, such as $I(n)$ and $\text{DIAG}(A)$.

Functions that refer to the parameters are listed in Table 43.4. The arguments for functions can be matrices. The simplest case, scalar arguments, is discussed first. For convenience, the scalar indices $i$ and $j$ refer to the position of the element in the coefficient matrix, and scalar $l$ refers to the lag value.

<table>
<thead>
<tr>
<th>Function</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\text{ACH}(l, i, j)$</td>
<td>ARCH parameter of the lag $l$ value of $\epsilon_t \epsilon'_t$ in a GARCH model</td>
</tr>
<tr>
<td>$\text{ALPHA}(i, j)$</td>
<td>The $(i, j)$ element in the adjustment coefficient matrix $\alpha$ for the vector error correction model</td>
</tr>
<tr>
<td>$\text{AR}(l, i, j)$</td>
<td>Autoregressive parameter of the lag $l$ value of the $j$th dependent (endogenous) variable, $y_{j,t-l}$, to the $i$th dependent variable at time $t$, $y_{it}$ for models other than error-correction models. For error correction models, $\text{AR}(1, i, j)$ is the $(i, j)$ element in $\Pi(= \alpha \beta')$ for $y_{t-1}$, and $\text{AR}(1, i, j)$, $l &gt; 1$, is the autoregressive parameter of the lag $(l - 1)$ value of the $j$th differenced dependent (endogenous) variable, $\Delta y_{j,t-(l-1)}$, to the $i$th differenced dependent variable at time $t$, $\Delta y_{it}$.</td>
</tr>
<tr>
<td>$\text{BETA}(i, j)$</td>
<td>The $(i, j)$ element in the cointegrating matrix $\beta$ for the vector error correction model</td>
</tr>
<tr>
<td>$\text{CCC}(i, j)$</td>
<td>Constant conditional correlation parameter between the $i$th and $j$th standardized error processes for the CCC GARCH model</td>
</tr>
<tr>
<td>$\text{CONST}(i)$</td>
<td>Intercept parameter of the $i$th time series, $y_{it}$</td>
</tr>
<tr>
<td>$\text{COV}(i, j)$</td>
<td>Covariance of innovations parameter between the $i$th and $j$th error processes when the maximum likelihood method is used for the fitted non-GARCH model</td>
</tr>
<tr>
<td>$D(i)$</td>
<td>Long-range dependent parameter of the $i$th time series, $y_{it}$, when the FI option is specified. By default, the LRD parameters are restricted between $-1/2$ and $1/2$.</td>
</tr>
<tr>
<td>$\text{DCCA}()$</td>
<td>Parameter $\alpha$ in the correlation equation for the DCC GARCH model</td>
</tr>
<tr>
<td>$\text{DCCB}()$</td>
<td>Parameter $\beta$ in the correlation equation for the DCC GARCH model</td>
</tr>
<tr>
<td>$\text{DCCS}(i, j)$</td>
<td>Unconditional correlation parameter between the $i$th and $j$th standardized error processes for the DCC GARCH model</td>
</tr>
<tr>
<td>$\text{EACH}(l, i, j)$</td>
<td>Exponential ARCH parameter of the lag $l$ value of $\epsilon_{it}/\sigma_{it}$ in the CCC or DCC GARCH model when SUBFORM=EGARCH is specified and $i = j$. If $i \neq j$, the value is set to 0.</td>
</tr>
<tr>
<td>$\text{ECONST}(i)$</td>
<td>The $i$th element for the constant in the error correction term for the vector error correction model when the ECTREND option in the COINTEG statement is specified</td>
</tr>
<tr>
<td>$\text{ECLTREND}(i)$</td>
<td>The $i$th element for the linear trend in the error correction term for vector error correction model when the ECTREND option in the COINTEG statement is specified</td>
</tr>
<tr>
<td>$\text{GCH}(l, i, j)$</td>
<td>GARCH parameter of the lag $l$ value of the covariance matrix, $H_t$, in a GARCH model</td>
</tr>
<tr>
<td>$\text{GCHC}(i, j)$</td>
<td>Constant parameter of the covariance matrix, $H_t$, in a GARCH model</td>
</tr>
</tbody>
</table>
Table 43.4  continued

<table>
<thead>
<tr>
<th>Function</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>LAMBDA(i)</td>
<td>Power parameter for the ith error process in the CCC or DCC GARCH model when SUBFORM=PGARCH is specified</td>
</tr>
<tr>
<td>LTREND(i)</td>
<td>Linear trend parameter of the ith time series, $y_{it}$, when the TREND= option is specified</td>
</tr>
<tr>
<td>MA(l,i,j)</td>
<td>Moving average parameter of the lag l value of the jth error process, $\epsilon_{j,t-l}$, to the ith dependent variable at time $t$, $y_{it}$</td>
</tr>
<tr>
<td>PARCH(l,i,j)</td>
<td>Power ARCH parameter of the lag l value of $\epsilon_{it}$ in the CCC or DCC GARCH model when SUBFORM=PGARCH is specified and $i = j$. If $i \neq j$, the value is set to 0.</td>
</tr>
<tr>
<td>QARCH(l,i,j)</td>
<td>Quadratic ARCH center parameter of the lag l value of $\epsilon_{it}$ in the CCC or DCC GARCH model when SUBFORM=QGARCH is specified and $i = j$. If $i \neq j$, the value is set to 0.</td>
</tr>
<tr>
<td>QTREND(i)</td>
<td>Quadratic trend parameter of the ith time series, $y_{it}$, when TREND=QUAD is specified</td>
</tr>
<tr>
<td>SD(i,j)</td>
<td>Same as SDUMMY(i,j)</td>
</tr>
<tr>
<td>SDUMMY(i,j)</td>
<td>The jth seasonal dummy of the ith time series at time $t$, $y_{it}$, where $j = 1, \ldots, (nseason-1)$, where nseason is the value of the NSEASON= option in the MODEL statement</td>
</tr>
<tr>
<td>TACH(l,i,j)</td>
<td>Threshold ARCH parameter of the lag l value of $1_{\epsilon_{it}&lt;0}\epsilon_{it}^2$ in the CCC or DCC GARCH model when SUBFORM=GJR is specified and $i = j$. If $i \neq j$, the value is set to 0.</td>
</tr>
<tr>
<td>XL(l,i,j)</td>
<td>Exogenous parameter of the lag l value of the jth exogenous (independent) variable, $x_{j,t-l}$, to the ith dependent variable at time $t$, $y_{it}$</td>
</tr>
</tbody>
</table>

The functions that refer to parameters, as shown in Table 43.4, accept vector arguments and return the matrix that is constructed by the corresponding parameters. According to the number of arguments, the following list shows what matrix a function returns when the arguments are vectors:

- A function, FUNC0, that has zero arguments, always returns the corresponding scalar parameter. DCCA and DCCB are types of FUNC0.

- A function, FUNC1, that has one vector argument I, where $I = (i_1 i_2 \ldots i_{n_I})'$, returns a vector $R = (r_1 r_2 \ldots r_{n_I})'$, where $r_k = FUNC1(i_k), k = 1, \ldots, n_I$. CONST, ECCONST, ECLTREND, LAMBDA, LTREND, and QTREND are types of FUNC1.

- A function, FUNC2, that has two vector arguments I and J, where $I = (i_1 i_2 \ldots i_{n_I})'$ and $J = (j_1 j_2 \ldots j_{n_J})'$, returns a matrix

$$ R = \begin{pmatrix} r_{1,1} & r_{1,2} & \cdots & r_{1,n_J} \\ r_{2,1} & r_{2,2} & \cdots & r_{2,n_J} \\ \vdots & \vdots & \ddots & \vdots \\ r_{n_I,1} & r_{n_I,2} & \cdots & r_{n_I,n_J} \end{pmatrix} $$

where $r_{k,m} = FUNC2(i_k, j_m), k = 1, \ldots, n_I, m = 1, \ldots, n_J$. ALPHA, BETA, CCC, COV, DCCS, GCHC, SD, and SDUMMY and types of FUNC2.
A function, \texttt{FUNC3}, that has three vector arguments \(L, I, \) and \(J\), where \(L = (l_1 \ l_2 \ \ldots \ l_n)\)', \(I = (i_1 \ i_2 \ \ldots \ i_m)\)', and \(J = (j_1 \ j_2 \ \ldots \ j_p)\)', returns a matrix

\[
R = \begin{pmatrix}
  r_{1,1} & r_{1,2} & \cdots & r_{1,n_L \ n_J} \\
  r_{2,1} & r_{2,2} & \cdots & r_{2,n_L \ n_J} \\
  \vdots & \vdots & \ddots & \vdots \\
  r_{n_L,1} & r_{n_L,2} & \cdots & r_{n_L,n_J} 
\end{pmatrix}
\]

where \(r_{k,m} = \text{FUNC3}(l_m, i_k, j_m), \ k = 1, \ldots, n_I, \ m = 1, \ldots, n_L \ n_J\), and \(l_m \) and \(j_m\) are the quotient and remainder of \(m\) divided by \(n_J\), respectively. \texttt{ACH, AR, EACH, GCH, MA, PACH, QACH, TACH, and XL} are types of \texttt{FUNC3}.

The functions that refer to parameters can accept empty arguments or omit any number of last arguments. The empty or omitted arguments are replaced with all possible values for those arguments. For example, PROC VARMAX is used to fit a bivariate \((k=2)\) VARX(1,1) model with three exogenous variables as follows:

\[
\text{model y1 y2 = x1 x2 x3 / p=1 xlag=3;}
\]

In order to restrict the third exogenous variable from having an effect on the first dependent variable, and to restrict the first exogenous variable from having an effect on the second dependent variable, you can use the following statement:

\[
\text{restrict XL(0 1 2 3), 1, 3) = 0,}
\]
\[
\text{XL(0 1 2 3), 2, 1) = 0;}
\]

Taking advantage of empty arguments, you can specify the preceding example as follows:

\[
\text{restrict XL(, 1, 3) = 0,}
\]
\[
\text{XL(, 2, 1) = 0;}
\]

To get all coefficients of the first lag exogenous variables on dependent variables, you can use \texttt{XL(1, \ (1 2), (1 2 3))} or \texttt{XL(1, \ , )} or \texttt{XL(1)}. To get all coefficients of exogenous variables on dependent variables, you can use \texttt{XL((0 1 2 3), \ (1 2), \ (1 2 3)),} or \texttt{XL(, \ , )} or \texttt{XL(}) or even just \texttt{XL}.

Another type of function does not refer to parameters but generates useful matrices. Table 43.5 lists all built-in functions supported by the RESTRICT statement.

<table>
<thead>
<tr>
<th>Function</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>\texttt{DIAG(A)}</td>
<td>Creates a diagonal matrix from a vector or extracts the diagonal elements of a matrix</td>
</tr>
<tr>
<td>\texttt{I(n)}</td>
<td>Creates an (n \times n) identity matrix</td>
</tr>
<tr>
<td>\texttt{J(m, n, elem)}</td>
<td>Creates an (m \times n) matrix with all elements equal to \texttt{elem}</td>
</tr>
<tr>
<td>\texttt{SHAPE(A, m, n)}</td>
<td>Creates a (m \times n) matrix with elements of matrix \texttt{A}</td>
</tr>
</tbody>
</table>

For more information about each function in Table 43.5, see the section “Details of Functions” on page 3061.
Examples

The following examples show how to use the RESTRICT statement.

This example shows a bivariate (k=2) VAR(2) model:

```sas
proc varmax data=one;
  model y1 y2 / p=2;
  restrict AR(1,1,2)=0, AR(2,1,2)=0.3;
run;
```

The AR(1,1,2) and AR(2,1,2) parameters are fixed as AR(1,1,2)=0 and AR(2,1,2)=0.3, respectively, and other parameters are to be estimated.

The following example shows a bivariate (k=2) VAR(1) model, estimated using the ML method:

```sas
proc varmax data=two;
  model y1 y2 = / p=1 method=ml;
  restrict cov(1,1)=cov(2,2), cov(1,2)=0;
run;
```

The COV(1,1) and COV(2,2) parameters are equal, and the correlation between the two series is fixed at 0. You can also express the preceding restrictions in matrix expressions as follows. This approach is very convenient when the number of dependent variables is large:

```sas
proc varmax data=two;
  model y1 y2 = / p=1 method=ml;
  restrict cov = cov(1,1)*I(2);
run;
```

When restricting a linear combination of parameters to be 0, you can omit the equal sign. For example, the following two RESTRICT statements are equivalent:

```sas
restrict AR(1)[1,1]-AR(1)[2,2], 2*MA(1)[1,2]-MA(1)[2,1];
restrict AR(1)[1,1]-AR(1)[2,2] = 0, 2*MA(1)[1,2]-MA(1)[2,1] = 0;
```

The following RESTRICT statement constrains four parameter estimates to be equal:

```sas
restrict AR(1)[1,1] = AR(1)[1,2],
      AR(1)[1,2] = AR(1)[2,1],
      AR(1)[2,1] = AR(1)[2,2];
```

This restriction can be abbreviated as follows:

```sas
restrict AR(1)[1,1] = AR(1)[1,2] = AR(1)[2,1] = AR(1)[2,2];
```

Or, in matrix expressions,

```sas
restrict AR(1,1:2,1:2) = J(2,2,AR(1,1,1));
```

The VARMA representation \( A(L) y_t = \Theta(L) \varepsilon_t \), where \( A(L) = I_k - A_1 L - \cdots - A_p L^p \) and \( \Theta(L) = I_k - \Theta_1 L - \cdots - \Theta_q L^q \), is said to be in final equation form if \( A(L) = a(L) I_k \), where \( a(L) = 1-a_1 L - \cdots - a_p L^p \) is a scalar operator with \( a_p \neq 0 \). If \( p \) and \( k \) are large, it would be difficult and inconvenient to restrict AR parameters element by element in standard form to estimate the VARMA model in final equation form. However, when you use matrix expressions, the restrictions become very simple, as shown in the following statement for a trivariate (\( k = 3 \)) VARMA\((p,q)\) model, where \( p \) might be any positive integer:
restrict AR = AR(1,1,1) @ I(3);

For the vector error correction models, the AR(1,.,.) parameters (that is, \( \Pi \)) are not supported in the RESTRICT statement, because AR(1) is in fact the product of the estimated parameters \( \alpha \) and the transpose of \( \beta \). Any linear constraints on AR(1) should be regarded as nonlinear constraints on the estimated parameters.

For the same reason, the CONST(.) or LTREND(.) functions are not supported in the RESTRICT statement if the ECTREND option in the COINTEG statement is specified. For example, the following statements are supported:

```plaintext
model y1-y4 / p=2;
cointeg rank=1 ectrend;
restrict ALPHA + BETA = 1.0,
     ECCONST;
```

However, neither of the following sets of statements is supported:

```plaintext
model y1-y4 / p=2;
cointeg rank=1 ectrend;
restrict AR(1,1,1) = 0;

model y1-y4 / p=2;
cointeg rank=1 ectrend;
restrict CONST(2) = 0.2;
```

### Details of Operators

This section describes all operators that are available in the RESTRICT statement. Each subsection shows how the operator is used, followed by a description of the operator.

**Addition Operator: +**

- `matrix1 + matrix2`
- `matrix + scalar`
- `matrix + vector`

The addition operator (+) computes a new matrix whose elements are the sums of the corresponding elements of `matrix1` and `matrix2`. If `matrix1` and `matrix2` are both \( n \times p \) matrices, then the addition operator adds the element in the \( i \)th row and \( j \)th column of the first matrix to the element in the \( i \)th row and \( j \)th column of the second matrix, for \( i = 1, \ldots, n \), \( j = 1, \ldots, p \). For example, \( \{1 \ 2 \ 3, \ 4 \ 5 \ 6\} + \{7 \ 8 \ 9, \ 10 \ 11 \ 12\} \) results in \( \{8 \ 10 \ 12, \ 14 \ 16 \ 18\} \).

You can also use the addition operator as follows to conveniently add a value to each element of a matrix, to each column of a matrix, or to each row of a matrix:

- When you use the `matrix + scalar` form, the scalar value is added to each element of the matrix.
- When you use the `matrix + vector` form, the vector is added to each row or column of the \( n \times p \) matrix.
  - If you add an \( n \times 1 \) column vector, each row of the vector is added to each row of the matrix.
  - If you add a \( 1 \times p \) row vector, each column of the vector is added to each column of the matrix.
For example, you can obtain \((2, 3, 4, 5, 6, 7)\) from \((1, 2, 3, 4, 5, 6) + 1\) or \((1, 2, 3, 4, 5, 6) + (1, 1, 1)\) or \((1, 2, 3, 4, 5, 6) + (1, 1)\).

**Comparison Operators:** \(=, <, <=, >, >\)

\[
\begin{align*}
\text{matrix1} & = \text{matrix2} \\
\text{matrix1} & < \text{matrix2} \\
\text{matrix1} & <= \text{matrix2} \\
\text{matrix1} & > \text{matrix2} \\
\text{matrix1} & >= \text{matrix2}
\end{align*}
\]

The comparison operators \(=, <, <=, >, >=\) compare two matrices element by element and return a list of equivalent restrictions on only scalar constants and parameters.

For example, the RESTRICT statement with matrix expressions

\[
\text{restrict AR}(1, \{1, 2\}, \{1, 2\}) = \text{MA}(2, \{3, 4\}, \{3, 4\});
\]

is transformed into the following equivalent RESTRICT statement with scalar parameters:

\[
\begin{align*}
\text{restrict AR}(1, 1, 1) & = \text{MA}(2, 3, 3), \\
\text{AR}(1, 1, 2) & = \text{MA}(2, 3, 4), \\
\text{AR}(1, 2, 1) & = \text{MA}(2, 4, 3), \\
\text{AR}(1, 2, 2) & = \text{MA}(2, 4, 4);
\end{align*}
\]

You can also use the comparison operators to conveniently compare all elements of a matrix with a scalar:

- If either argument is a scalar, then the VARMAX procedure performs an elementwise comparison between each element of the matrix and the scalar.

You can also compare an \(n \times p\) matrix with a row or column vector:

- If the comparison is with an \(n \times 1\) column vector, the VARMAX procedure compares each row of the vector to each row of the matrix.
- If the comparison is with a \(1 \times p\) row vector, the VARMAX procedure compares each column of the vector to each column of the matrix.

For example, the following statements are equivalent:

\[
\begin{align*}
\text{restrict AR}(1, 1:2, 1:3) & >= 0.2; \\
\text{restrict AR}(1, 1:2, 1:3) & >= \{0.2, 0.2\}; \\
\text{restrict AR}(1, 1:2, 1:3) & >= \{0.2 0.2 0.2\};
\end{align*}
\]

**Concatenation Operator, Horizontal:** \(\|\)

\[
\begin{align*}
\text{matrix1} & \| \text{matrix2}
\end{align*}
\]

The horizontal concatenation operator \(\|\) produces a new matrix by horizontally joining \text{matrix1} and \text{matrix2}. The matrices must have the same number of rows, which is also the number of rows in the new
matrix. The number of columns in the new matrix is the number of columns in `matrix1` plus the number of columns in `matrix2`.

For example, `{1 1 1, 7 7 7} || {0 0 0, 8 8 8}` returns `{1 1 0 0 0, 7 7 8 8 8}`.

**Concatenation Operator, Vertical:** `//`

`matrix1 // matrix2`

The vertical concatenation operator (`//`) produces a new matrix by vertically joining `matrix1` and `matrix2`. The matrices must have the same number of columns, which is also the number of columns in the new matrix. The number of rows in the new matrix is the number of rows in `matrix1` plus the number of rows in `matrix2`.

For example, `{1 1 1} // {0 0 0, 8 8 8}` returns `{1 1 1, 0 0 0, 8 8 8}`.

**Direct Product Operator:** `@`

`matrix1 @ matrix2`

The direct product operator (`@`) computes a new matrix that is the direct product (also called the *Kronecker product*) of `matrix1` and `matrix2`. For matrices `A` and `B`, the direct product is denoted by `A ⊗ B`. The number of rows in the new matrix equals the product of the number of rows in `matrix1` and the number of rows in `matrix2`; the number of columns in the new matrix equals the product of the number of columns in `matrix1` and the number of columns in `matrix2`.

Specifically, if `A` is an `n × p` matrix and `B` is a `m × q` matrix, then the Kronecker product `A ⊗ B` is the following `nm × pq` block matrix:

```
A ⊗ B =

\[
\begin{bmatrix}
A_{11}B & \cdots & A_{1p}B \\
\vdots & \ddots & \vdots \\
A_{n1}B & \cdots & A_{np}B
\end{bmatrix}
\]
```

For example, `{1 2, 3 4} @ {0 2}` returns `{0 2 0 4, 0 6 0 8}`, and `{0 2} @ {1 2, 3 4}` returns `{0 0 2 4, 0 0 6 8}`. Note that the direct product of two matrices is not commutative.

**Index Creation Operator:** `:`

`value1 : value2`

The index creation operator `:` creates a column vector whose first element is `value1`, whose second element is `value1 + 1`, and so on, until the last element, which is less than or equal to `value2`.

For example, `3 : 6` returns `{3 4 5 6}`.

If `value1` is greater than `value2`, a reverse-order index is created. For example, `6 : 3` returns `{6 5 4 3}`.

Neither `value1` nor `value2` is required to be an integer.

**Multiplication Operator, Elementwise:** `#`

`matrix1 # matrix2`

`matrix # scalar`

`matrix # vector`
The elementwise multiplication operator (#) computes a new matrix whose elements are the products of the corresponding elements of $\text{matrix1}$ and $\text{matrix2}$.

For example, $(1 \ 2, \ 3 \ 4) \# (4 \ 8, \ 0 \ 5)$ returns $(4 \ 16, \ 0 \ 20)$.

In addition to multiplying matrices that have the same dimensions, you can use the elementwise multiplication operator to multiply a matrix and a scalar:

- When either argument is a scalar, each element in $\text{matrix}$ is multiplied by the scalar value.

When you use the $\text{matrix} \# \text{vector}$ form, each row or column of the $n \times p$ matrix is multiplied by a corresponding element of the vector:

- If you multiply by an $n \times 1$ column vector, each row of the matrix is multiplied by the corresponding row of the vector.
- If you multiply by a $1 \times p$ row vector, each column of the matrix is multiplied by the corresponding column of the vector.

For example, a $2 \times 3$ matrix can be multiplied on either side by a $2 \times 3$, $1 \times 3$, $2 \times 1$, or $1 \times 1$ scalar.

The product of elementwise multiplication is also known as the Schur or Hadamard product. Elementwise multiplication (which uses the # operator) should not be confused with matrix multiplication (which uses the * operator).

**Multiplication Operator, Matrix:** $\star$

$\text{matrix1} \star \text{matrix2}$

The matrix multiplication operator ($\star$) computes a new matrix by performing matrix multiplication. The first matrix must have the same number of columns as the second matrix has rows. The new matrix has the same number of rows as the first matrix and the same number of columns as the second matrix. That is, if $A$ is an $n \times p$ matrix and $B$ is a $p \times m$ matrix, then the product $A \star B$ is an $n \times m$ matrix. The $(i, j)$ element of the product is the sum $\sum_{k=1}^{p} A_{ik} B_{kj}$.

For example, $(1 \ 2, \ 3 \ 4) \star (1, \ 2)$ returns $(5, \ 11)$.

**Sign Reversal Operator:** $-$

$- \text{matrix}$

The sign reversal operator (--) computes a new matrix whose elements are formed by reversing the sign of each element in $\text{matrix}$. The sign reversal operator is also called the *unary minus* operator.

For example, $-(-1 \ 7 \ 6, \ 2 \ 0 \ -8)$ returns $(1 -7 -6, -2 \ 0 \ 8)$.

**Subscripts:** [ ]

$\text{matrix[rows, columns]}$

$\text{matrix[elements]}$

Subscripts are used with matrices to select submatrices, where rows, columns, and elements are expressions that evaluate to scalars or vectors. If these expressions are numeric, they must contain valid subscript values of rows and columns, or the indices, in the argument matrix.
For example, \(\{1 \ 2 \ 3, \ 4 \ 5 \ 6, \ 7 \ 8 \ 9\}\)[2,3] returns 6, \(\{1 \ 2 \ 3, \ 4 \ 5 \ 6, \ 7 \ 8 \ 9\}\)[2,1:3] returns \(\{4 \ 5 \ 6\}\), and \(\{1 \ 2 \ 3, \ 4 \ 5 \ 6, \ 7 \ 8 \ 9\}\)[,3] returns \(\{3, \ 6, \ 9\}\). Because the VARMAX procedure stores matrices in row-major order, \(\{11 \ 22 \ 33, \ 44 \ 55 \ 66, \ 77 \ 88 \ 99\}\)[3 5 9] returns \(\{33, \ 55, \ 99\}\).

**Subtraction Operator:** –
- \(\text{matrix1} - \text{matrix2}\)
- \(\text{matrix} - \text{scalar}\)
- \(\text{matrix} - \text{vector}\)

The subtraction operator (–) computes a new matrix whose elements are formed by subtracting the corresponding elements of \(\text{matrix2}\) from those of \(\text{matrix1}\).

In addition to subtracting conformable matrices, you can also use the subtraction operator to subtract a scalar from a matrix or subtract a vector from a matrix:

- When either argument is a scalar, the VARMAX procedure performs the subtraction between the scalar and each element of the matrix argument. For example, when you use the \(\text{matrix} - \text{scalar}\) form, the scalar value is subtracted from each element of the matrix.
- When you use the \(\text{matrix} - \text{vector}\) form, the vector is subtracted from each row or column of the \(n \times p\) matrix.
  - If you subtract an \(n \times 1\) column vector, each row of the vector is subtracted from each row of the matrix.
  - If you subtract a \(1 \times p\) row vector, each column of the vector is subtracted from each column of the matrix.

For example, \(\{1 \ 2 \ 3, \ 4 \ 5 \ 6\} - \{1 \ 1 \ 1, \ 1 \ 1 \ 1\}\) returns \(\{0 \ 1 \ 2, \ 3 \ 4 \ 5\}\). The same results can be obtained by \(\{1 \ 2 \ 3, \ 4 \ 5 \ 6\} - \{1\}\) or \(\{1 \ 2 \ 3, \ 4 \ 5 \ 6\} - \{1 \ 1\}\) or \(\{1 \ 2 \ 3, \ 4 \ 5 \ 6\} - \{1, \ 1\}\).

**Transpose Operator:** `
- \(\text{matrix}'\)

The transpose operator, denoted by the backquote character ('), exchanges the rows and columns of \(\text{matrix}\), producing the transpose of \(\text{matrix}\). If \(v\) is the value in the \(i\)th row and \(j\)th column of \(\text{matrix}\), then the transpose of \(\text{matrix}\) contains \(v\) in the \(j\)th row and \(i\)th column. If \(\text{matrix}\) contains \(n\) rows and \(p\) columns, the transpose has \(p\) rows and \(n\) columns.

For example, \(\{1 \ 2, \ 3 \ 4, \ 5 \ 6\}'\) returns \(\{1 \ 3, \ 5, \ 2 \ 4 \ 6\}\).

**Details of Functions**

**DIAG Function**
- \(\text{DIAG(matrix)}\)

The DIAG function creates a diagonal matrix from a vector or extracts the diagonal elements of a matrix. The matrix argument can be either a square matrix or a vector.

If \(\text{matrix}\) is a vector, the DIAG function creates a matrix whose diagonal elements are the values in the vector. All off-diagonal elements are zeros.
If \texttt{matrix} is a square matrix, the \texttt{DIAG} function creates a vector from the diagonal elements of the matrix.

For example, \texttt{DIAG(\{1 2 3, 4 5 6, 7 8 9\})} returns \{1, 5, 9\}. Also, \texttt{DIAG(\{1 5 9\})} or \texttt{DIAG(DIAG(\{1 2 3, 4 5 6, 7 8 9\}))} returns \{1 0 0, 0 5 0, 0 0 9\}.

\textit{I Function}

\texttt{I(dim)}

The \texttt{I} function creates an identity matrix that contains \texttt{dim} rows and columns. The diagonal elements of an identity matrix are ones; all other elements are zeros. The value of \texttt{dim} must be an integer greater than or equal to 1. Noninteger operands are rounded to the nearest integer.

For example, \texttt{I(3)} returns \{1 0 0, 0 1 0, 0 0 1\}.

\textit{J Function}

\texttt{J(nrow, ncol, value)}

The \texttt{J} function creates a matrix that contains \texttt{nrow} rows and \texttt{ncol} columns, in which all elements are equal to \texttt{value}.

The arguments \texttt{nrow} and \texttt{ncol} are both integers; \texttt{value} can be any expression that returns a linear combination of scalar constants and parameters.

For example, \texttt{J(2, 3, 1)} returns \{1 1 1, 1 1 1\}. \texttt{J(2, 3, 5+2*AR(1,1,1))} returns the same result as \texttt{J(2, 3, 1) * (5+2*AR(1,1,1))}.

\textit{SHAPE Function}

\texttt{SHAPE(matrix, nrow, ncol)}

The \texttt{SHAPE} function creates a new matrix from data in \texttt{matrix}. The values \texttt{nrow} and \texttt{ncol} specify the number of rows and columns, respectively, in the new matrix. The \texttt{SHAPE} function produces the result matrix by traversing the argument matrix in row-major order until it reaches the specified number of elements. If necessary, the \texttt{SHAPE} function reuses elements.

For example, \texttt{SHAPE(\{1 2 3, 4 5 6\}, 3, 2)} returns \{1 2 3, 4 5 6\}; \texttt{SHAPE(\{1 2 3, 4 5 6\}, 5, 2)} returns \{1 2, 3 4, 5 6, 1 2, 3 4\}; and \texttt{SHAPE(\{1 2 3, 4 5 6\}, 1, 4)} returns \{1 2 3 4\}.

\textbf{TEST Statement}

\texttt{TEST restriction, \ldots, restriction ;}

The TEST statement performs the Wald test for the joint linear hypothesis that is specified in the statement. Each restriction specifies a linear hypothesis to be tested. If you specify more than one \texttt{restriction}, separate them with commas. Specify the \texttt{restrictions} in the same manner as in the RESTRICT statement. For information about how to define restriction by using matrix expressions, operators, and functions, see the section “RESTRICT Statement” on page 3050. You can specify any number of TEST statements.
To use the TEST statement, you need to know the form of the model. If you do not specify the GARCH statement, the COINTEG statement, or the ECM=, P=, Q=, or XLAG= option in the MODEL statement, then the TEST statement is not applicable. Nonlinear restrictions on parameters are not supported.

For information about the Wald test, see the section “Granger Causality Test” on page 3095.

The following is an example of the TEST statement for a bivariate \((k=2)\) VAR(2) model:

```plaintext
proc varmax data=one;
  model y1 y2 / p=2;
  test AR(1,1,2) = 0, AR(2,1,2) = 0;
run;
```

After estimating the parameters, the TEST statement tests the null hypothesis that \(AR(1,1,2)=0\) and \(AR(2,1,2)=0\). Like the RESTRICT statement, the preceding TEST statement can be abbreviated as follows:

```plaintext
test AR(1,1,2) = AR(2,1,2) = 0;
```

or

```plaintext
test AR(1,1,2), AR(2,1,2);
```

Note that the following statements are different from the preceding statement:

```plaintext
  test AR(1,1,2);
  test AR(2,1,2);
```

These two TEST statements are to test two null hypotheses separately: one is \(AR(1,1,2)=0\), and the other is \(AR(2,1,2)=0\).

For the vector error correction models, you can test the hypothesis on the \(AR(1,\ldots)\) parameters (that is, \(\Pi\)) by using the TEST statement, because asymptotically these parameters follow a normal distribution and the Wald test can be applied. For the same reason, you can use the `CONST(.)` or `LTREND(.)` function in the TEST statement if the ECTREND option in the COINTEG statement is specified. However, the `BETA(.,.), ECCONST(.),` and `ECLTREND(.)` functions are not supported in the TEST statement. For example, the following statements are supported:

```plaintext
model y1-y4 / p=2;
  cointeg rank=1 ectrend;
  test AR(1,1,1);
  test CONST(2);
```

However, the following statements are not supported:

```plaintext
model y1-y4 / p=2;
  cointeg rank=1 ectrend;
  test BETA(1,1) = BETA(2,1) = 0;
```

or

```plaintext
model y1-y4 / p=2;
  cointeg rank=1 ectrend;
  test ECCONST(1) = 0.2;
```
Details: VARMAX Procedure

Missing Values

The VARMAX procedure currently does not support missing values. PROC VARMAX uses the first contiguous group of observations that have no missing values for any of the MODEL statement variables. Observations at the beginning of the data set that have missing values for any MODEL statement variables are not used or included in the output data set. At the end of the data set, observations can have dependent (endogenous) variables with missing values and independent (exogenous) variables with nonmissing values.

VARMAX Model

The vector autoregressive moving-average model with exogenous variables is called the VARMAX($p,q,s$) model. The form of the model can be written as

$$y_t = \sum_{i=1}^{p} \Phi_i y_{t-i} + \sum_{i=0}^{s} \Theta^*_i x_{t-i} + \epsilon_t - \sum_{i=1}^{q} \Theta_i \epsilon_{t-i}$$

where the output variables of interest, $y_t = (y_{1t}, \ldots, y_{kt})'$, can be influenced by other input variables, $x_t = (x_{1t}, \ldots, x_{rt})'$, which are determined outside of the system of interest. The variables $y_t$ are referred to as dependent, response, or endogenous variables, and the variables $x_t$ are referred to as independent, input, predictor, regressor, or exogenous variables. The unobserved noise variables, $\epsilon_t = (\epsilon_{1t}, \ldots, \epsilon_{kt})'$, are a vector white noise process.

The VARMAX($p,q,s$) model can be written

$$\Phi(B)y_t = \Theta^*(B)x_t + \Theta(B)\epsilon_t$$

where

$$\Phi(B) = I_k - \Phi_1 B - \cdots - \Phi_p B^p$$
$$\Theta^*(B) = \Theta^*_0 + \Theta^*_1 B + \cdots + \Theta^*_s B^s$$
$$\Theta(B) = I_k - \Theta_1 B - \cdots - \Theta_q B^q$$

are matrix polynomials in $B$ in the backshift operator, such that $B^i y_t = y_{t-i}$, the $\Phi_i$ and $\Theta_i$ are $k \times k$ matrices, and the $\Theta^*_i$ are $k \times r$ matrices.

The following assumptions are made:

- $E(\epsilon_t) = 0$, $E(\epsilon_t \epsilon'_s) = \Sigma$, which is positive-definite, and $E(\epsilon_t \epsilon'_t) = 0$ for $t \neq s$.
- For stationarity and invertibility of the VARMAX process, the roots of $|\Phi(z)| = 0$ and $|\Theta(z)| = 0$ are outside the unit circle.
The exogenous (independent) variables $x_t$ are not correlated with residuals $\epsilon_t$, $E(x_t\epsilon'_t) = 0$. The exogenous variables can be stochastic or nonstochastic. When the exogenous variables are stochastic and their future values are unknown, forecasts of these future values are needed to forecast the future values of the endogenous (dependent) variables. On occasion, future values of the exogenous variables can be assumed to be known because they are deterministic variables. The VARMAX procedure assumes that the exogenous variables are nonstochastic if future values are available in the input data set. Otherwise, the exogenous variables are assumed to be stochastic and their future values are forecasted by assuming that they follow the VARMA($p,q$) model, prior to forecasting the endogenous variables, where $p$ and $q$ are the same as in the VARMAX($p,q,s$) model.

State Space Representation

Another representation of the VARMAX($p,q,s$) model is in the form of a state variable or a state space model, which consists of a state equation

$$z_t = Fz_{t-1} + Kx_t + G\epsilon_t$$

and an observation equation

$$y_t = Hz_t$$

where

$$z_t = \begin{bmatrix} y_t \\ \vdots \\ y_{t-p+1} \\ x_t \\ \vdots \\ x_{t-s+1} \\ \epsilon_t \\ \vdots \\ \epsilon_{t-q+1} \end{bmatrix}, \quad K = \begin{bmatrix} \Theta^*_0 \\ \vdots \\ \Theta^*_r \end{bmatrix}, \quad G = \begin{bmatrix} I_k \\ \vdots \\ 0_{r \times k} \end{bmatrix}$$

$$F = \begin{bmatrix} \Phi_1 & \cdots & \Phi_{p-1} & \Phi_p & \Theta^*_1 & \cdots & \Theta^*_{s-1} & \Theta^*_s & -\Theta_1 & \cdots & -\Theta_{q-1} & -\Theta_q \\ I_k & \cdots & 0 & 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 & 0 \\ 0 & \cdots & I_k & 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 & 0 \\ 0 & \cdots & 0 & 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 & 0 \\ 0 & \cdots & 0 & 0 & I_r & \cdots & 0 & 0 & 0 & \cdots & 0 & 0 \\ 0 & \cdots & 0 & 0 & 0 & \cdots & I_r & 0 & 0 & \cdots & 0 & 0 \\ 0 & \cdots & 0 & 0 & 0 & \cdots & 0 & 0 & I_k & \cdots & 0 & 0 \\ 0 & \cdots & 0 & 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & I_k & 0 \end{bmatrix}$$

and

$$H = [I_k, 0_{k \times k}, \ldots, 0_{k \times k}, 0_{k \times r}, \ldots, 0_{k \times r}, 0_{k \times k}, \ldots, 0_{k \times k}]$$
On the other hand, it is assumed that $x_t$ follows a VARMA($p,q$) model

$$x_t = \sum_{i=1}^{p} A_i x_{t-i} + \alpha_t - \sum_{i=1}^{q} C_i \alpha_{t-i}$$

The model can also be expressed as

$$A(B)x_t = C(B)\alpha_t$$

where $A(B) = I_r - A_1 B - \cdots - A_p B^p$ and $C(B) = I_r - C_1 B - \cdots - C_q B^q$ are matrix polynomials in $B$, and the $A_i$ and $C_i$ are $r \times r$ matrices. Without loss of generality, the AR and MA orders can be taken to be the same as the VARMAX($p,q,s$) model, and $\alpha_t$ and $\epsilon_t$ are independent white noise processes.

Under suitable conditions such as stationarity, $x_t$ is represented by an infinite order moving-average process

$$x_t = A(B)^{-1} C(B)\alpha_t = \sum_{j=0}^{\infty} \Psi^x_j \alpha_{t-j}$$

where $\Psi^x(B) = A(B)^{-1} C(B) = \sum_{j=0}^{\infty} \Psi^x_j B^j$.

The optimal minimum mean squared error (minimum MSE) $i$-step-ahead forecast of $x_{t+i}$ is

$$x_{t+i|t} = \sum_{j=i}^{\infty} \Psi^x_j \alpha_{t+i-j}$$

$$x_{t+i|t+1} = x_{t+i|t} + \Psi^x_{i-1} \alpha_{t+1}$$

For $i > q$,

$$x_{t+i|t} = \sum_{j=1}^{p} A_j x_{t+i-j|t}$$

The VARMAX($p,q,s$) model has an absolutely convergent representation as

$$y_t = \Phi(B)^{-1} \Theta^*(B)x_t + \Phi(B)^{-1} \Theta(B)\epsilon_t$$

$$= \Psi^*(B) \Psi^x(B) \alpha_t + \Phi(B)^{-1} \Theta(B)\epsilon_t$$

$$= V(B) \alpha_t + \Psi(B)\epsilon_t$$

or

$$y_t = \sum_{j=0}^{\infty} V_j \alpha_{t-j} + \sum_{j=0}^{\infty} \Psi_j \epsilon_{t-j}$$

where $\Psi(B) = \Phi(B)^{-1} \Theta(B) = \sum_{j=0}^{\infty} \Psi_j B^j$, $\Psi^*(B) = \Phi(B)^{-1} \Theta^*(B)$, and $V(B) = \Psi^*(B) \Psi^x(B) = \sum_{j=0}^{\infty} V_j B^j$.

The optimal (minimum MSE) $i$-step-ahead forecast of $y_{t+i}$ is

$$y_{t+i|t} = \sum_{j=i}^{\infty} V_j a_{t+i-j} + \sum_{j=i}^{\infty} \Psi_j \epsilon_{t+i-j}$$

$$y_{t+i|t+1} = y_{t+i|t} + \sum_{j=1}^{i-1} V_{i-j} a_{t+1} + \sum_{j=1}^{i-1} \Psi_{i-j} \epsilon_{t+1}$$
for $i = 1, \ldots, v$ with $v = \max(p, q + 1)$. For $i > q$,

$$
\begin{align*}
Y_{t+i|t} &= \sum_{j=1}^{p} \Phi_j Y_{t+i-j|t} + \sum_{j=0}^{s} \Theta^*_j x_{t+i-j|t} \\
&= \sum_{j=1}^{p} \Phi_j Y_{t+i-j|t} + \Theta^*_0 x_{t+i|t} + \sum_{j=1}^{s} \Theta^*_j x_{t+i-j|t} \\
&= \sum_{j=1}^{p} \Phi_j Y_{t+i-j|t} + \Theta^*_0 \sum_{j=1}^{p} A_j x_{t+i-j|t} + \sum_{j=1}^{s} \Theta^*_j x_{t+i-j|t} \\
&= \sum_{j=1}^{p} \Phi_j Y_{t+i-j|t} + \sum_{j=1}^{u} (\Theta^*_0 A_j + \Theta^*_j) x_{t+i-j|t}
\end{align*}
$$

where $u = \max(p, s)$.

Define $\Pi_j = \Theta^*_0 A_j + \Theta^*_j$. For $i = v > q$ with $v = \max(p, q + 1)$, you obtain

$$
\begin{align*}
Y_{t+v|t} &= \sum_{j=1}^{p} \Phi_j Y_{t+v-j|t} + \sum_{j=1}^{u} \Pi_j x_{t+v-j|t} \text{ for } u \leq v \\
Y_{t+v|t} &= \sum_{j=1}^{p} \Phi_j Y_{t+v-j|t} + \sum_{j=1}^{r} \Pi_j x_{t+v-j|t} \text{ for } u > v
\end{align*}
$$

From the preceding relations, a state equation is

$$
Z_{t+1} = F z_t + K x^*_t + G e_{t+1}
$$

and an observation equation is

$$
Y_t = H z_t
$$

where

$$
\begin{bmatrix}
Y_t \\
Y_{t+1|t} \\
\vdots \\
Y_{t+v-1|t} \\
x_t \\
X_{t+1|t} \\
\vdots \\
x_{t+v-1|t}
\end{bmatrix}
= 
\begin{bmatrix}
x_{t+v-u} \\
x_{t+v-u+1} \\
\vdots \\
x_{t-1}
\end{bmatrix},
\begin{bmatrix}
ea_t \\
e_{t+1}
\end{bmatrix}
Chapter 43: The VARMAX Procedure

Dynamic Simultaneous Equations Modeling

In the econometrics literature, the VARMAX\((p,q,s)\) model is sometimes written in a form that is slightly different than the one shown in the previous section. This alternative form is referred to as a dynamic simultaneous equations model or a dynamic structural equations model.

Because \(E(\epsilon_t \epsilon'_t) = \Sigma\) is assumed to be positive-definite, there exists a lower triangular matrix \(A_0\) that has ones on the diagonals such that \(A_0 \Sigma A'_0 = \Sigma^d\), where \(\Sigma^d\) is a diagonal matrix that has positive diagonal elements.

\[
A_0 y_t = \sum_{i=1}^{p} A_i y_{t-i} + \sum_{i=0}^{s} C^*_i x_{t-i} + A_0 \epsilon_t - \sum_{i=1}^{q} C_i A_0 \epsilon_{t-i}
\]

where \(A_i = A_0 \Phi_i, C^*_i = A_0 \Theta^*_i,\) and \(C_i = A_0 \Theta_i A_0^{-1}\).
As an alternative form,

\[ A_0 y_t = \sum_{i=1}^{p} A_i y_{t-i} + \sum_{i=0}^{s} C_i^* x_{t-i} + a_t - \sum_{i=1}^{q} C_i a_{t-i} \]

where \( A_i = A_0 \Phi_i, C_i^* = A_0 \Theta_i^*, C_i = A_0 \Theta_i A_0^{-1}, \) and \( a_t = A_0 \epsilon_t \). The covariance matrix of \( a_t \) is the diagonal matrix \( \Sigma^d \). The PRINT=(DYNAMIC) option returns the parameter estimates that result from estimating the model in this form.

A dynamic simultaneous equations model involves a leading (lower triangular) coefficient matrix for \( y_t \) at lag 0 or a leading coefficient matrix for \( \epsilon_t \) at lag 0. Such a representation of the VARMAX\((p,q,s)\) model can be more useful in certain circumstances than the standard representation. From the linear combination of the dependent variables obtained by \( A_0 y_t \), you can easily see the relationship between the dependent variables in the current time.

The following statements provide the dynamic simultaneous equations of the VAR(1) model:

```plaintext
proc iml;
    sig = {1.0 0.5, 0.5 1.25};
    phi = {1.2 -0.5, 0.6 0.3};
    /* simulate the vector time series */
    call varmasim(y,phi) sigma = sig n = 100 seed = 34657;
    cn = {'y1' 'y2'};
    create simull from y[colname=cn];
    append from y;
quit;

data simull;
    set simull;
    date = intnx('year', '01jan1900'd, _n_-1 );
    format date year4.;
run;

proc varmax data=simull;
    model y1 y2 / p=1 noint print=(dynamic);
run;
```

This is the same data set and model used in the section “Getting Started: V ARMAX Procedure” on page 2982. You can compare the results of the VARMA model form and the dynamic simultaneous equations model form.
In Figure 43.4 in the section “Getting Started: VARMAX Procedure” on page 2982, the covariance of $\epsilon_t$ estimated from the VARMAX model form is

$$
\Sigma_{\epsilon} = \begin{pmatrix}
1.28875 & 0.39751 \\
0.39751 & 1.41839
\end{pmatrix}
$$

Figure 43.44 shows the results from estimating the model as a dynamic simultaneous equations model. By the decomposition of $\Sigma_{\epsilon}$, you get a diagonal matrix ($\Sigma_{\alpha}$) and a lower triangular matrix ($A_0$) such as $\Sigma_{\alpha} = A_0 \Sigma_{\epsilon} A_0'$ where

$$
\Sigma_{\alpha} = \begin{pmatrix}
1.28875 & 0 \\
0 & 1.29578
\end{pmatrix} \quad \text{and} \quad A_0 = \begin{pmatrix}
1 & 0 \\
-0.30845 & 1
\end{pmatrix}
$$

The lower triangular matrix ($A_0$) is shown in the left side of the simultaneous equations model. The parameter estimates in equations system are shown in the right side of the two-equations system.
The simultaneous equations model is written as

\[
\begin{pmatrix}
1 & 0 \\
-0.30845 & 1
\end{pmatrix} y_t = \begin{pmatrix}
1.15977 & -0.51058 \\
0.18861 & 0.54247
\end{pmatrix} y_{t-1} + a_t
\]

The resulting two-equation system can be written as

\[
y_{1t} = 1.15977y_{1,t-1} - 0.51058y_{2,t-1} + a_{1t}
\]

\[
y_{2t} = 0.30845y_{1t} + 0.18861y_{1,t-1} + 0.54247y_{2,t-1} + a_{2t}
\]

---

**Impulse Response Function**

**Simple Impulse Response Function (IMPULSE=SIMPLE Option)**

The VARMAX\((p, q, s)\) model has a convergent representation

\[
y_t = \Psi^*(B)x_t + \Psi(B)\epsilon_t
\]

where \(\Psi^*(B) = \Phi(B)^{-1}\Theta^*(B) = \sum_{j=0}^{\infty} \Psi_j B^j\) and \(\Psi(B) = \Phi(B)^{-1}\Theta(B) = \sum_{j=0}^{\infty} \Psi_j B^j\).

The elements of the matrices \(\Psi_j\) from the operator \(\Psi(B)\), called the impulse response, can be interpreted as the response of a variable to a shock in another variable. Let \(\psi_{j,i,n}\) be the \((i, n)\) element of \(\Psi_j\) at lag \(j\), where \(n\) is the index for the impulse variable, and \(i\) is the index for the response variable (impulse \(\rightarrow\) response); that is to say, \(\psi_{j,i,n}\) shows the reaction of the \(i\)-th variable to a unit shock in variable \(n\), \(j\) periods ago, assuming that the effect is not contaminated by other shocks (Lütkepohl 1993). For instance, \(\psi_{j,11}\) is an impulse response to \(y_{1t} \rightarrow y_{1t}\), and \(\psi_{j,12}\) is an impulse response to \(y_{2t} \rightarrow y_{1t}\).

**Accumulated Impulse Response Function (IMPULSE=ACCUM Option)**

The accumulated impulse response function is the cumulative sum of the impulse response function, \(\Psi^a_t = \sum_{j=0}^t \Psi_j\).

**Orthogonalized Impulse Response Function (IMPULSE=ORTH Option)**

The MA representation of a VARMAX\((p, q)\) model with a standardized white noise innovation process offers another way to interpret a VARMAX\((p, q)\) model. Since \(\Sigma\) is positive-definite, there is a lower triangular matrix \(P\) such that \(\Sigma = PP'\). The alternate MA representation of a VARMAX\((p, q)\) model is written as

\[
y_t = \Psi^o(B)u_t
\]

where \(\Psi^o(B) = \sum_{j=0}^{\infty} \Psi_j^o B^j\), \(\Psi_j^o = \Psi_j P\), and \(u_t = P^{-1}\epsilon_t\).

The elements of the matrices \(\Psi_j^o\), called the orthogonal impulse response, can be interpreted as the effects of the components of the standardized shock process \(u_t\) on the process \(y_t\) at lag \(j\).
Impulse Response of Transfer Function (IMPULSX=SIMPLE Option)

The coefficient matrix $\Psi_j^*$ from the transfer function operator $\Psi^*(B)$ can be interpreted as the effects that changes in the exogenous variables $x_t$ have on the output variable $y_t$ at lag $j$; it is called an impulse response matrix in the transfer function.

Accumulated Impulse Response of Transfer Function (IMPULSX=ACCUM Option)

The accumulated impulse response in the transfer function is the cumulative sum of the impulse response in the transfer function, $\Psi_{l,a}^* = \sum_{j=0}^{l} \Psi_j^*$.

The asymptotic distributions of the impulse functions can be seen in the section “VAR and VARX Modeling” on page 3092.

The following statements provide the impulse response and the accumulated impulse response in the transfer function for a VARX(1,0) model:

```plaintext
proc varmax data=grunfeld plot=impulse;
  model y1-y3 = x1 x2 / p=1 lagmax=5
       printform=univariate
       print=(impulsx=(all) estimates);
run;
```

In Figure 43.45, the variables $x1$ and $x2$ are impulses, and the variables $y1$, $y2$, and $y3$ are responses. The keyword STD stands for the standard errors of the elements. You can read the table that matches the impulse $\rightarrow$ response pairs, such as $x1 \rightarrow y1$, $x1 \rightarrow y2$, $x1 \rightarrow y3$, $x2 \rightarrow y1$, $x2 \rightarrow y2$, and $x2 \rightarrow y3$. In the pair $x1 \rightarrow y1$, you can see the long-run responses of $y1$ to an impulse in $x1$ (the values are 1.69281, 0.35399, 0.09090, and so on for lag 0, lag 1, lag 2, and so on, respectively).
**The VARMAX Procedure**

**Simple Impulse Response of Transfer Function by Variable**

<table>
<thead>
<tr>
<th>Variable Response</th>
<th>Impulse</th>
<th>Lag</th>
<th>( x_1 )</th>
<th>( x_2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( y_1 )</td>
<td>0</td>
<td>1.69281</td>
<td>-0.00859</td>
<td></td>
</tr>
<tr>
<td></td>
<td>STD 0.54395</td>
<td>0.05361</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1 0.35399</td>
<td>0.01727</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>STD 0.36482</td>
<td>0.03762</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>2 0.09090</td>
<td>0.00714</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>STD 0.17419</td>
<td>0.01592</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>3 0.05136</td>
<td>0.00214</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>STD 0.08203</td>
<td>0.00524</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>4 0.04717</td>
<td>0.00072</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>STD 0.07969</td>
<td>0.00229</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>5 0.04620</td>
<td>0.00040</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>STD 0.08216</td>
<td>0.00170</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( y_2 )</td>
<td>0</td>
<td>-6.09850</td>
<td>2.57980</td>
<td></td>
</tr>
<tr>
<td></td>
<td>STD 5.07849</td>
<td>0.50056</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1 -5.15484</td>
<td>0.45445</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>STD 3.89665</td>
<td>0.40534</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>2 -3.04168</td>
<td>0.04391</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>STD 1.56519</td>
<td>0.13268</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>3 -2.23797</td>
<td>-0.01376</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>STD 1.15163</td>
<td>0.08723</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>4 -1.98183</td>
<td>-0.01647</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>STD 1.08738</td>
<td>0.07844</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>5 -1.87415</td>
<td>-0.01453</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>STD 0.99384</td>
<td>0.07250</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( y_3 )</td>
<td>0</td>
<td>-0.02317</td>
<td>-0.01274</td>
<td></td>
</tr>
<tr>
<td></td>
<td>STD 0.20418</td>
<td>0.02012</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1 1.57476</td>
<td>-0.01435</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>STD 0.56132</td>
<td>0.05515</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>2 1.80231</td>
<td>0.00398</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>STD 0.61049</td>
<td>0.05896</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>3 1.77024</td>
<td>0.01062</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>STD 0.64476</td>
<td>0.06380</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>4 1.70435</td>
<td>0.01197</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>STD 0.62648</td>
<td>0.06353</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>5 1.63913</td>
<td>0.01187</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>STD 0.59511</td>
<td>0.06142</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Figure 43.46 shows the responses of $y_1$, $y_2$, and $y_3$ to a forecast error impulse in $x_1$.

**Figure 43.46** Plot of Impulse Response in Transfer Function
Figure 43.47 shows the accumulated impulse response in transfer function.

Figure 43.47  Accumulated Impulse Response in Transfer Function (IMPLS= Option)

<table>
<thead>
<tr>
<th>Variable</th>
<th>Lag</th>
<th>x1</th>
<th>x2</th>
</tr>
</thead>
<tbody>
<tr>
<td>y1</td>
<td>0</td>
<td>1.69281</td>
<td>-0.00859</td>
</tr>
<tr>
<td></td>
<td>STD</td>
<td>0.54395</td>
<td>0.05361</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>2.04680</td>
<td>0.00868</td>
</tr>
<tr>
<td></td>
<td>STD</td>
<td>0.36482</td>
<td>0.03762</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>2.13770</td>
<td>0.01582</td>
</tr>
<tr>
<td></td>
<td>STD</td>
<td>0.17419</td>
<td>0.01592</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>2.18906</td>
<td>0.01796</td>
</tr>
<tr>
<td></td>
<td>STD</td>
<td>0.08203</td>
<td>0.00524</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>2.23623</td>
<td>0.01867</td>
</tr>
<tr>
<td></td>
<td>STD</td>
<td>0.07969</td>
<td>0.00229</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>2.28243</td>
<td>0.01907</td>
</tr>
<tr>
<td></td>
<td>STD</td>
<td>0.08216</td>
<td>0.00170</td>
</tr>
<tr>
<td>y2</td>
<td>0</td>
<td>-6.09850</td>
<td>2.57980</td>
</tr>
<tr>
<td></td>
<td>STD</td>
<td>5.07849</td>
<td>0.50056</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>-11.25334</td>
<td>3.03425</td>
</tr>
<tr>
<td></td>
<td>STD</td>
<td>3.89665</td>
<td>0.40534</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>-14.29502</td>
<td>3.07816</td>
</tr>
<tr>
<td></td>
<td>STD</td>
<td>1.56519</td>
<td>0.13268</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>-16.53299</td>
<td>3.06440</td>
</tr>
<tr>
<td></td>
<td>STD</td>
<td>1.15163</td>
<td>0.08723</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>-18.51482</td>
<td>3.04793</td>
</tr>
<tr>
<td></td>
<td>STD</td>
<td>1.08738</td>
<td>0.07844</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>-20.38897</td>
<td>3.03340</td>
</tr>
<tr>
<td></td>
<td>STD</td>
<td>0.99384</td>
<td>0.07250</td>
</tr>
<tr>
<td>y3</td>
<td>0</td>
<td>0.02317</td>
<td>-0.01274</td>
</tr>
<tr>
<td></td>
<td>STD</td>
<td>0.20418</td>
<td>0.02012</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>1.55159</td>
<td>-0.02709</td>
</tr>
<tr>
<td></td>
<td>STD</td>
<td>0.56132</td>
<td>0.05515</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>3.35390</td>
<td>-0.02311</td>
</tr>
<tr>
<td></td>
<td>STD</td>
<td>0.61049</td>
<td>0.05896</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>5.12414</td>
<td>-0.01249</td>
</tr>
<tr>
<td></td>
<td>STD</td>
<td>0.64476</td>
<td>0.06380</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>6.82848</td>
<td>-0.00052</td>
</tr>
<tr>
<td></td>
<td>STD</td>
<td>0.62648</td>
<td>0.06353</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>8.46762</td>
<td>0.01135</td>
</tr>
<tr>
<td></td>
<td>STD</td>
<td>0.59511</td>
<td>0.06142</td>
</tr>
</tbody>
</table>
Figure 43.48 shows the accumulated responses of $y_1$, $y_2$, and $y_3$ to a forecast error impulse in $x_1$.

**Figure 43.48 Plot of Accumulated Impulse Response in Transfer Function**

The following statements provide the impulse response function, the accumulated impulse response function, and the orthogonalized impulse response function with their standard errors for a VAR(1) model. Parts of the VARMAX procedure output are shown in Figure 43.49, Figure 43.51, and Figure 43.53.

```latex
\begin{verbatim}
proc varmax data=simul1 plot=impulse;
  model y1 y2 / p=1 noint lagmax=5
       print=(impulse=(all))
       printform=univariate;
run;
\end{verbatim}
```
Figure 43.49 is the output in a univariate format associated with the PRINT=(IMPULSE=) option for the impulse response function. The keyword STD stands for the standard errors of the elements. The matrix in terms of the lag 0 does not print since it is the identity. In Figure 43.49, the variables $y_1$ and $y_2$ of the first row are impulses, and the variables $y_1$ and $y_2$ of the first column are responses. You can read the table matching the impulse $\rightarrow$ response pairs, such as $y_1 \rightarrow y_1$, $y_1 \rightarrow y_2$, $y_2 \rightarrow y_1$, and $y_2 \rightarrow y_2$. For example, in the pair of $y_1 \rightarrow y_1$ at lag 3, the response is 0.8055. This represents the impact on $y_1$ of one-unit change in $y_1$ after 3 periods. As the lag gets higher, you can see the long-run responses of $y_1$ to an impulse in itself.

**Figure 43.49** Impulse Response Function (IMPULSE= Option)

<table>
<thead>
<tr>
<th>Variable</th>
<th>Simple Impulse Response by Variable</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Lag</td>
</tr>
<tr>
<td>$y_1$</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>STD</td>
</tr>
<tr>
<td></td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>STD</td>
</tr>
<tr>
<td></td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>STD</td>
</tr>
<tr>
<td></td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>STD</td>
</tr>
<tr>
<td></td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>STD</td>
</tr>
<tr>
<td>$y_2$</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>STD</td>
</tr>
<tr>
<td></td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>STD</td>
</tr>
<tr>
<td></td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>STD</td>
</tr>
<tr>
<td></td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>STD</td>
</tr>
<tr>
<td></td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>STD</td>
</tr>
</tbody>
</table>
Figure 43.50 shows the responses of $y_1$ and $y_2$ to a forecast error impulse in $y_1$ with two standard errors.

**Figure 43.50** Plot of Impulse Response

**Response to Impulse in $y_1$**
With Two Standard Errors

<table>
<thead>
<tr>
<th></th>
<th>$y_1$</th>
<th>$y_2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Response</td>
<td>1.0</td>
<td>1.0</td>
</tr>
<tr>
<td></td>
<td>0.5</td>
<td>0.5</td>
</tr>
<tr>
<td></td>
<td>0.0</td>
<td>0.0</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Lag</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>$y_1$</td>
<td>1.0</td>
<td>0.6</td>
<td>0.4</td>
<td>0.2</td>
<td>0.0</td>
</tr>
<tr>
<td>$y_2$</td>
<td>0.0</td>
<td>0.4</td>
<td>0.6</td>
<td>1.0</td>
<td>0.6</td>
</tr>
</tbody>
</table>
**Figure 43.51** is the output in a univariate format associated with the PRINT=(IMPULSE=) option for the accumulated impulse response function. The matrix in terms of the lag 0 does not print since it is the identity.

**Figure 43.51** Accumulated Impulse Response Function (IMPULSE= Option)

<table>
<thead>
<tr>
<th>Variable Response\Impulse</th>
<th>Lag</th>
<th>y1</th>
<th>y2</th>
</tr>
</thead>
<tbody>
<tr>
<td>y1</td>
<td>1</td>
<td>2.15977</td>
<td>-0.51058</td>
</tr>
<tr>
<td></td>
<td>STD</td>
<td>0.05508</td>
<td>0.05898</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>3.22589</td>
<td>-1.29929</td>
</tr>
<tr>
<td></td>
<td>STD</td>
<td>0.21684</td>
<td>0.22776</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>4.03144</td>
<td>-2.14728</td>
</tr>
<tr>
<td></td>
<td>STD</td>
<td>0.52217</td>
<td>0.53649</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>4.50241</td>
<td>-2.88504</td>
</tr>
<tr>
<td></td>
<td>STD</td>
<td>0.96922</td>
<td>0.97088</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>4.64556</td>
<td>-3.40953</td>
</tr>
<tr>
<td></td>
<td>STD</td>
<td>1.51137</td>
<td>1.47122</td>
</tr>
<tr>
<td>y2</td>
<td>1</td>
<td>0.54634</td>
<td>1.38499</td>
</tr>
<tr>
<td></td>
<td>STD</td>
<td>0.05779</td>
<td>0.06188</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>1.39030</td>
<td>1.25426</td>
</tr>
<tr>
<td></td>
<td>STD</td>
<td>0.17614</td>
<td>0.18392</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>2.29768</td>
<td>0.77302</td>
</tr>
<tr>
<td></td>
<td>STD</td>
<td>0.36166</td>
<td>0.36874</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>3.08711</td>
<td>0.12447</td>
</tr>
<tr>
<td></td>
<td>STD</td>
<td>0.65129</td>
<td>0.65333</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>3.64834</td>
<td>-0.52829</td>
</tr>
<tr>
<td></td>
<td>STD</td>
<td>1.07510</td>
<td>1.06309</td>
</tr>
</tbody>
</table>
Figure 43.52 shows the accumulated responses of $y_1$ and $y_2$ to a forecast error impulse in $y_1$ with two standard errors.

![Figure 43.52 Plot of Accumulated Impulse Response](image)

Figure 43.53 is the output in a univariate format associated with the PRINT=(IMPULSE=) option for the orthogonalized impulse response function. The two right-hand side columns, $y_1$ and $y_2$, represent the $y_1\_innovation$ and $y_2\_innovation$ variables. These are the impulses variables. The left-hand side column contains responses variables, $y_1$ and $y_2$. You can read the table by matching the impulse $\rightarrow$ response pairs such as $y_1\_innovation \rightarrow y_1$, $y_1\_innovation \rightarrow y_2$, $y_2\_innovation \rightarrow y_1$, and $y_2\_innovation \rightarrow y_2$. 
In Figure 43.4, there is a positive correlation between $\varepsilon_{1t}$ and $\varepsilon_{2t}$. Therefore, shock in $y_1$ can be accompanied by a shock in $y_2$ in the same period. For example, in the pair of $y_1\_\text{innovation} \rightarrow y_2$, you can see the long-run responses of $y_2$ to an impulse in $y_1\_\text{innovation}$. 

<table>
<thead>
<tr>
<th>Lag</th>
<th>$y_1$</th>
<th>$y_2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1.13523</td>
<td>0.00000</td>
</tr>
<tr>
<td>1</td>
<td>1.13783</td>
<td>-0.58120</td>
</tr>
<tr>
<td>2</td>
<td>0.93412</td>
<td>-0.89782</td>
</tr>
<tr>
<td>3</td>
<td>0.61756</td>
<td>-0.96528</td>
</tr>
<tr>
<td>4</td>
<td>0.15348</td>
<td>0.18595</td>
</tr>
<tr>
<td>5</td>
<td>-0.02115</td>
<td>-0.59705</td>
</tr>
<tr>
<td>0</td>
<td>0.35016</td>
<td>1.13832</td>
</tr>
<tr>
<td>1</td>
<td>0.75503</td>
<td>0.43824</td>
</tr>
<tr>
<td>2</td>
<td>0.91231</td>
<td>-0.14881</td>
</tr>
<tr>
<td>3</td>
<td>0.10553</td>
<td>0.13565</td>
</tr>
<tr>
<td>4</td>
<td>0.86158</td>
<td>-0.54780</td>
</tr>
<tr>
<td>5</td>
<td>0.12266</td>
<td>0.14825</td>
</tr>
<tr>
<td>4</td>
<td>0.66909</td>
<td>-0.73827</td>
</tr>
<tr>
<td>5</td>
<td>0.13305</td>
<td>0.15846</td>
</tr>
<tr>
<td>5</td>
<td>0.40856</td>
<td>-0.74304</td>
</tr>
<tr>
<td>STD</td>
<td>0.14189</td>
<td>0.16765</td>
</tr>
</tbody>
</table>
Figure 43.54 shows the orthogonalized responses of \( y_1 \) and \( y_2 \) to a forecast error impulse in \( y_1 \) with two standard errors.

**Figure 43.54** Plot of Orthogonalized Impulse Response

<table>
<thead>
<tr>
<th>y1</th>
<th>y2</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image" alt="Response to Orthogonalized Impulse in y1" /></td>
<td><img src="image" alt="Response to Orthogonalized Impulse in y2" /></td>
</tr>
</tbody>
</table>

### Forecasting

The optimal (minimum MSE) \( l \)-step-ahead forecast of \( y_{t+l} \) is

\[
y_{t+l|t} = \sum_{j=1}^{p} \Phi_j y_{t+l-j|t} + \sum_{j=0}^{s} \Theta_j^* x_{t+l-j|t} - \sum_{j=l}^{q} \Theta_j \epsilon_{t+l-j}, \quad l \leq q
\]

\[
y_{t+l|t} = \sum_{j=1}^{p} \Phi_j y_{t+l-j|t} + \sum_{j=0}^{s} \Theta_j^* x_{t+l-j|t}, \quad l > q
\]

where \( y_{t+l-j|t} = y_{t+l-j} \) and \( x_{t+l-j|t} = x_{t+l-j} \) for \( l \leq j \). For information about the forecasts \( x_{t+l-j|t} \), see the section “State Space Representation” on page 3065.
Covariance Matrices of Prediction Errors without Exogenous (Independent) Variables

Under the stationarity assumption, the optimal (minimum MSE) $l$-step-ahead forecast of $y_{t+l}$ has an infinite moving-average form, $y_{t+l} = \sum_{j=0}^{\infty} \Psi_j \epsilon_{t+l-j}$. The prediction error of the optimal $l$-step-ahead forecast is $e_{t+l} = y_{t+l} - y_{t+l} = \sum_{j=0}^{l-1} \Psi_j \epsilon_{t+l-j}$, with zero mean and covariance matrix,

$$\Sigma(l) = \text{Cov}(e_{t+l}) = \sum_{j=0}^{l-1} \Psi_j \Sigma_j' = \sum_{j=0}^{l-1} \Psi_j^{\prime} \Psi_j^{\prime}$$

where $\Psi_j^{\prime} = \Psi_j P$ with a lower triangular matrix $P$ such that $\Sigma = PP'$. Under the assumption of normality of the $\epsilon_t$, the $l$-step-ahead prediction error $e_{t+l}$ is also normally distributed as multivariate $N(0, \Sigma(l))$. Hence, it follows that the diagonal elements $\sigma^2_{ii}(l)$ of $\Sigma(l)$ can be used, together with the point forecasts $y_{i,t+l}$, to construct $l$-step-ahead prediction intervals of the future values of the component series, $y_{i,t+l}$.

The following statements use the COVPE option to compute the covariance matrices of the prediction errors for a VAR(1) model. The parts of the VARMAX procedure output are shown in Figure 43.55 and Figure 43.56.

```latex
proc varmax data=simul1;
  model y1 y2 / p=1 noint lagmax=5
    printform=both
    print=(decompose(5) impulse=(all) covpe(5));
run;
```

Figure 43.55 is the output in a matrix format associated with the COVPE option for the prediction error covariance matrices.

### Figure 43.55 Covariances of Prediction Errors (COVPE Option)

<table>
<thead>
<tr>
<th>Lead Variable</th>
<th>y1</th>
<th>y2</th>
</tr>
</thead>
<tbody>
<tr>
<td>y1</td>
<td>1.28875</td>
<td>0.39751</td>
</tr>
<tr>
<td>y2</td>
<td>0.39751</td>
<td>1.41839</td>
</tr>
<tr>
<td>2 y1</td>
<td>2.92119</td>
<td>1.00189</td>
</tr>
<tr>
<td>y2</td>
<td>1.00189</td>
<td>2.18051</td>
</tr>
<tr>
<td>3 y1</td>
<td>4.59984</td>
<td>1.98771</td>
</tr>
<tr>
<td>y2</td>
<td>1.98771</td>
<td>3.03498</td>
</tr>
<tr>
<td>4 y1</td>
<td>5.91299</td>
<td>3.04856</td>
</tr>
<tr>
<td>y2</td>
<td>3.04856</td>
<td>4.07738</td>
</tr>
<tr>
<td>5 y1</td>
<td>6.69463</td>
<td>3.85346</td>
</tr>
<tr>
<td>y2</td>
<td>3.85346</td>
<td>5.07010</td>
</tr>
</tbody>
</table>

The VARMAX Procedure
Figure 43.56 is the output in a univariate format associated with the COVPE option for the prediction error covariances. This printing format more easily explains the prediction error covariances of each variable.

Figure 43.56 Covariances of Prediction Errors

<table>
<thead>
<tr>
<th>Variable</th>
<th>Lead</th>
<th>y1</th>
<th>y2</th>
</tr>
</thead>
<tbody>
<tr>
<td>y1</td>
<td></td>
<td>1.28875</td>
<td>0.39751</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>2.92119</td>
<td>1.00189</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>4.59984</td>
<td>1.98771</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>5.91299</td>
<td>3.04856</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>6.69463</td>
<td>3.85346</td>
</tr>
<tr>
<td>y2</td>
<td>1</td>
<td>0.39751</td>
<td>1.41839</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>1.00189</td>
<td>2.18051</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>1.98771</td>
<td>3.03498</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>3.04856</td>
<td>4.07738</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>3.85346</td>
<td>5.07010</td>
</tr>
</tbody>
</table>

Covariance Matrices of Prediction Errors in the Presence of Exogenous (Independent) Variables

Exogenous variables can be both stochastic and nonstochastic (deterministic) variables. Considering the forecasts in the V ARMAX($p,q,s$) model, there are two cases.

When exogenous (independent) variables are stochastic (future values not specified):

As defined in the section “State Space Representation” on page 3065, $y_{t+l|t}$ has the representation

$$y_{t+l|t} = \sum_{j=l}^{\infty} V_j a_{t+l-j} + \sum_{j=l}^{\infty} \psi_j \epsilon_{t+l-j}$$

and hence

$$e_{t+l|t} = \sum_{j=0}^{l-1} V_j a_{t+l-j} + \sum_{j=0}^{l-1} \psi_j \epsilon_{t+l-j}$$

Therefore, the covariance matrix of the $l$-step-ahead prediction error is given as

$$\Sigma(l) = \text{Cov}(e_{t+l|t}) = \sum_{j=0}^{l-1} V_j \Sigma_a V_j' + \sum_{j=0}^{l-1} \psi_j \Sigma_x \psi_j'$$

where $\Sigma_a$ is the covariance of the white noise series $a_t$, and $a_t$ is the white noise series for the VARMA($p,q$) model of exogenous (independent) variables, which is assumed not to be correlated with $\epsilon_t$ or its lags.
When future exogenous (independent) variables are specified:

The optimal forecast $y_{t+l|t}$ of $y_t$ conditioned on the past information and also on known future values $x_{t+1}, \ldots, x_{t+l}$ can be represented as

$$y_{t+l|t} = \sum_{j=0}^{\infty} \Psi_j^* x_{t+l-j} + \sum_{j=1}^{\infty} \Psi_j \epsilon_{t+l-j}$$

and the forecast error is

$$e_{t+l|t} = \sum_{j=0}^{l-1} \Psi_j \epsilon_{t+l-j}$$

Thus, the covariance matrix of the $l$-step-ahead prediction error is given as

$$\Sigma(l) = \text{Cov}(e_{t+l|t}) = \sum_{j=0}^{l-1} \Psi_j \sum_{\epsilon} \Psi'_j$$

**Decomposition of Prediction Error Covariances**

In the relation $\Sigma(l) = \sum_{j=0}^{l-1} \Psi_j^o \Psi'_j$, the diagonal elements can be interpreted as providing a decomposition of the $l$-step-ahead prediction error covariance $\sigma^2_{i|l}(l)$ for each component series $y_{i|l}$ into contributions from the components of the standardized innovations $\epsilon_t$.

If you denote the ($i$, $n$) element of $\Psi_j^o$ by $\psi_{j,in}$, the MSE of $y_{i,t+h|t}$ is

$$\text{MSE}(y_{i,t+h|t}) = \text{E}(y_{i,t+h} - y_{i,t+h|t})^2 = \sum_{j=0}^{l-1} \sum_{k=1}^{K} \psi_{j,in}^2$$

Note that $\sum_{j=0}^{l-1} \psi_{j,in}^2$ is interpreted as the contribution of innovations in variable $n$ to the prediction error covariance of the $l$-step-ahead forecast of variable $i$.

The proportion, $\omega_{l,in}$, of the $l$-step-ahead forecast error covariance of variable $i$ accounting for the innovations in variable $n$ is

$$\omega_{l,in} = \sum_{j=0}^{l-1} \psi_{j,in}^2 / \text{MSE}(y_{i,t+h|t})$$

The following statements use the DECOMPOSE option to compute the decomposition of prediction error covariances and their proportions for a VAR(1) model:

```fortran
proc varmax data=simul1;
   model y1 y2 / p=1 noint print=(decompose(15)) printform=univariate;
run;
```

The proportions of decomposition of prediction error covariances of two variables are given in Figure 43.57. The output explains that about 91.356% of the one-step-ahead prediction error covariances of the variable $y_{2t}$ is accounted for by its own innovations and about 8.644% is accounted for by $y_{1t}$ innovations.
Forecasting of the Centered Series

If the CENTER option is specified, the sample mean mean vector is added to the forecast.

Forecasting of the Differenced Series

If dependent (endogenous) variables are differenced, the final forecasts and their prediction error covariances are produced by integrating those of the differenced series. However, if the PRIOR option is specified, the forecasts and their prediction error variances of the differenced series are produced.

Let \( z_t \) be the original series with some appended zero values that correspond to the unobserved past observations. Let \( \Delta(B) \) be the \( k \times k \) matrix polynomial in the backshift operator that corresponds to the differencing specified by the MODEL statement. The off-diagonal elements of \( \Delta_i \) are zero, and the diagonal elements can be different. Then \( y_t = \Delta(B)z_t \).

This gives the relationship

\[
z_t = \Delta^{-1}(B)y_t = \sum_{j=0}^{\infty} \Lambda_j y_{t-j}
\]

where \( \Delta^{-1}(B) = \sum_{j=0}^{\infty} \Lambda_j B^j \) and \( \Lambda_0 = I_k \).

The \( l \)-step-ahead prediction of \( z_{t+l} \) is

\[
z_{t+l|t} = \sum_{j=0}^{l-1} \Lambda_j y_{t+l-j} + \sum_{j=l}^{\infty} \Lambda_j y_{t+l-j}
\]

The \( l \)-step-ahead prediction error of \( z_{t+l} \) is

\[
\sum_{j=0}^{l-1} \Lambda_j (y_{t+l-j} - y_{t+l-j}) = \sum_{j=0}^{l-1} \left( \sum_{u=0}^{j} \Lambda_u \Psi_{j-u} \right) \epsilon_{t+l-j}
\]
Letting $\Sigma_z(0) = 0$, the covariance matrix of the $l$-step-ahead prediction error of $z_{t+l}$, $\Sigma_z(l)$, is

$$
\Sigma_z(l) = \sum_{j=0}^{l-1} \left( \sum_{u=0}^{\lfloor j \rfloor} \Lambda_u \Psi_{j-u} \right) \Sigma_\varepsilon \left( \sum_{u=0}^{\lfloor j \rfloor} \Lambda_u \Psi_{j-u} \right)'
$$

$$
= \Sigma_z(l-1) + \left( \sum_{j=0}^{l-1} \Lambda_j \Psi_{l-1-j} \right) \Sigma_\varepsilon \left( \sum_{j=0}^{l-1} \Lambda_j \Psi_{l-1-j} \right)'
$$

If there are stochastic exogenous (independent) variables, the covariance matrix of the $l$-step-ahead prediction error of $z_{t+l}$, $\Sigma_z(l)$, is

$$
\Sigma_z(l) = \Sigma_z(l-1) + \left( \sum_{j=0}^{l-1} \Lambda_j \Psi_{l-1-j} \right) \Sigma_\varepsilon \left( \sum_{j=0}^{l-1} \Lambda_j \Psi_{l-1-j} \right)'
$$

$$
+ \left( \sum_{j=0}^{l-1} \Lambda_j V_{l-1-j} \right) \Sigma_a \left( \sum_{j=0}^{l-1} \Lambda_j V_{l-1-j} \right)'
$$

**Tentative Order Selection**

**Sample Cross-Covariance and Cross-Correlation Matrices**

Given a stationary multivariate time series $y_t$, cross-covariance matrices are

$$
\Gamma(l) = \mathbb{E}[(y_t - \mu)(y_{t+l} - \mu)']
$$

where $\mu = \mathbb{E}(y_t)$, and cross-correlation matrices are

$$
\rho(l) = D^{-1} \Gamma(l) D^{-1}
$$

where $D$ is a diagonal matrix with the standard deviations of the components of $y_t$ on the diagonal.

The sample cross-covariance matrix at lag $l$, denoted as $C(l)$, is computed as

$$
\hat{C}(l) = C(l) = \frac{1}{T} \sum_{t=1}^{T-l} \tilde{y}_t \tilde{y}_{t+l}'
$$

where $\tilde{y}_t$ is the centered data and $T$ is the number of nonmissing observations. Thus, the $(i, j)$ element of $\hat{C}(l)$ is $\hat{c}_{ij}(l) = c_{ij}(l)$. The sample cross-correlation matrix at lag $l$ is computed as

$$
\hat{\rho}_{ij}(l) = c_{ij}(l)/[c_{ii}(0)c_{jj}(0)]^{1/2}, \quad i, j = 1, ..., k
$$

The following statements use the CORRY option to compute the sample cross-correlation matrices and their summary indicator plots in terms of $+$, $-$, and $\cdot$, where $+$ indicates significant positive cross-correlations, $-$ indicates significant negative cross-correlations, and $\cdot$ indicates insignificant cross-correlations:
proc varmax data=simul1;
    model y1 y2 / p=1 noint lagmax=3 print=(corry)
        printform=univariate;
run;

Figure 43.58 shows the sample cross-correlation matrices of $y_{1t}$ and $y_{2t}$. As shown, the sample autocorrelation functions for each variable decay quickly, but are significant with respect to two standard errors.

### Figure 43.58 Cross-Correlations (CORRY Option)

<table>
<thead>
<tr>
<th>Variable</th>
<th>Lag</th>
<th>y1</th>
<th>y2</th>
</tr>
</thead>
<tbody>
<tr>
<td>y1</td>
<td>0</td>
<td>1.00000</td>
<td>0.67041</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>0.83143</td>
<td>0.84330</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>0.56094</td>
<td>0.81972</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>0.26629</td>
<td>0.66154</td>
</tr>
<tr>
<td>y2</td>
<td>0</td>
<td>0.67041</td>
<td>1.00000</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>0.29707</td>
<td>0.77132</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>-0.00936</td>
<td>0.48658</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>-0.22058</td>
<td>0.22014</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Schematic Representation of Cross Correlations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Variable/Lag</td>
</tr>
<tr>
<td>y1</td>
</tr>
<tr>
<td>y2</td>
</tr>
</tbody>
</table>

+ is > 2*std error, - is < -2*std error, . is between

### Partial Autoregressive Matrices

For each $m = 1, 2, \ldots, p$, you can define a sequence of matrices $\Phi_{mm}$, which is called the partial autoregression matrices of lag $m$, as the solution for $\Phi_{mm}$ to the Yule-Walker equations of order $m$,

$$\Gamma(l) = \sum_{i=1}^{m} \Gamma(l-i) \Phi_{im}'$$

The sequence of the partial autoregression matrices $\Phi_{mm}$ of order $m$ has the characteristic property that if the process follows the AR($p$), then $\Phi_{pp} = \Phi_p$ and $\Phi_{mm} = 0$ for $m > p$. Hence, the matrices $\Phi_{mm}$ have the cutoff property for a VAR($p$) model, and so they can be useful in the identification of the order of a pure VAR model.

The following statements use the PARCOEF option to compute the partial autoregression matrices:

```bash
proc varmax data=simul1;
    model y1 y2 / p=1 noint lagmax=3
        printform=univariate
        print=(corry parcoef pcorr)
```
Figure 43.59 shows that the model can be obtained by an AR order \( m = 1 \) since partial autoregression matrices are insignificant after lag 1 with respect to two standard errors. The matrix for lag 1 is the same as the Yule-Walker autoregressive matrix.

**Figure 43.59** Partial Autoregression Matrices (PARCOEF Option)

<table>
<thead>
<tr>
<th>The VARMAX Procedure</th>
</tr>
</thead>
<tbody>
<tr>
<td>Partial Autoregression</td>
</tr>
<tr>
<td>Lag</td>
</tr>
<tr>
<td>-----</td>
</tr>
<tr>
<td>1</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>2</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>3</td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>

Schematic Representation of Partial Autoregression

<table>
<thead>
<tr>
<th>Variable/Lag</th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>y1</td>
<td>+-</td>
<td>..</td>
<td>..</td>
</tr>
<tr>
<td>y2</td>
<td>++</td>
<td>..</td>
<td>..</td>
</tr>
</tbody>
</table>

+ is > 2*std error, - is < -2*std error, . is between

**Partial Correlation Matrices**

Define the forward autoregression

\[
y_t = \sum_{i=1}^{m-1} \Phi_{i,m-1} y_{t-i} + u_{m,t}
\]

and the backward autoregression

\[
y_{t-m} = \sum_{i=1}^{m-1} \Phi_{i,m-1} y_{t-m+i} + u_{m,t-m}^*
\]

The matrices \( P(m) \) defined by Ansley and Newbold (1979) are given by

\[
P(m) = \Sigma_{m-1}^{1/2} \Phi_{m}^* \Sigma_{m-1}^{-1/2}
\]

where

\[
\Sigma_{m-1} = \text{Cov}(u_{m,t}) = \Gamma(0) - \sum_{i=1}^{m-1} \Gamma(-i) \Phi_{i,m-1}^*
\]
and

\[ \Sigma_{m-1}^* = \text{Cov}(u_{m,t-m}^*) = \Gamma(0) - \sum_{i=1}^{m-1} \Gamma(m-i) \Phi_{m-i,m-1}^{*'} \]

\( P(m) \) are the partial cross-correlation matrices at lag \( m \) between the elements of \( y_t \) and \( y_{t-m} \), given \( y_{t-1}, \ldots, y_{t-m+1} \). The matrices \( P(m) \) have the cutoff property for a \( \text{VAR}(p) \) model, and so they can be useful in the identification of the order of a pure \( \text{VAR} \) structure.

The following statements use the PCORR option to compute the partial cross-correlation matrices:

```latex
proc varmax data=simul1;
  model y1 y2 / p=1 noint lagmax=3
    print=(pcorr)
    printform=univariate;
run;
```

The partial cross-correlation matrices in Figure 43.60 are insignificant after lag 1 with respect to two standard errors. This indicates that an AR order of \( m = 1 \) can be an appropriate choice.

**Figure 43.60** Partial Correlations (PCORR Option)

### The VARMAX Procedure

<table>
<thead>
<tr>
<th>Partial Cross Correlations by Variable</th>
</tr>
</thead>
<tbody>
<tr>
<td>Variable</td>
</tr>
<tr>
<td>----------</td>
</tr>
<tr>
<td>y1</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>y2</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>

**Schematic Representation of Partial Cross Correlations**

<table>
<thead>
<tr>
<th>Variable/Lag</th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>y1</td>
<td>++</td>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td>y2</td>
<td>-+</td>
<td>.</td>
<td>.</td>
</tr>
</tbody>
</table>

+ is > 2*std error, - is < -2*std error, . is between

### Partial Canonical Correlation Matrices

The partial canonical correlations at lag \( m \) between the vectors \( y_t \) and \( y_{t-m} \), given \( y_{t-1}, \ldots, y_{t-m+1} \), are \( 1 \geq \rho_1(m) \geq \rho_2(m) \cdots \geq \rho_k(m) \). The partial canonical correlations are the canonical correlations between the residual series \( u_{m,t} \) and \( u_{m,t-m}^* \), where \( u_{m,t} \) and \( u_{m,t-m}^* \) are defined in the previous section. Thus, the squared partial canonical correlations \( \rho_i^2(m) \) are the eigenvalues of the matrix

\[
\{\text{Cov}(u_{m,t})\}^{-1}E(u_{m,t}u_{m,t-m}^{'})\{\text{Cov}(u_{m,t-m}^*)\}^{-1}E(u_{m,t-m}^*u_{m,t}^{'}) = \Phi_{m,m}^{*'}\Phi_{m,m}^{*'}
\]
It follows that the test statistic to test for $\Phi_m = 0$ in the VAR model of order $m > p$ is approximately

$$(T - m) \text{tr} \{\Phi_m^\prime \Phi_m\} \approx (T - m) \sum_{i=1}^{k} \rho_i^2(m)$$

and has an asymptotic chi-square distribution with $k^2$ degrees of freedom for $m > p$.

The following statements use the PCANCORR option to compute the partial canonical correlations:

```sas
proc varmax data=simul1;
  model y1 y2 / p=1 noint lagmax=3 print=(pcancorr);
run;
```

Figure 43.61 shows that the partial canonical correlations $\rho_i(m)$ between $y_t$ and $y_{t-m}$ are $\{0.918, 0.773\}$, $\{0.092, 0.018\}$, and $\{0.109, 0.011\}$ for lags $m = 1$ to 3. After lag $m = 1$, the partial canonical correlations are insignificant with respect to the 0.05 significance level, indicating that an AR order of $m = 1$ can be an appropriate choice.

The Minimum Information Criterion (MINIC) Method

The minimum information criterion (MINIC) method can tentatively identify the orders of a VARMA($p,q$) process (Spliid 1983; Koreisha and Pukkila 1989; Quinn 1980). The first step of this method is to obtain estimates of the innovations series, $\epsilon_t$, from the VAR($p_e$), where $p_e$ is chosen sufficiently large. The choice of the autoregressive order, $p_e$, is determined by use of a selection criterion. From the selected VAR($p_e$) model, you obtain estimates of residual series

$$\tilde{\epsilon}_t = y_t - \sum_{i=1}^{p_e} \tilde{\Phi}_i^e y_{t-i} - \tilde{\delta}^e \epsilon_t, \ t = p_e + 1, \ldots, T$$

In the second step, you select the order $(p,q)$ of the VARMA model for $p$ in $(p_{min} : p_{max})$ and $q$ in $(q_{min} : q_{max})$

$$y_t = \delta + \sum_{i=1}^{p} \Phi_i y_{t-i} - \sum_{i=1}^{q} \Theta_i \tilde{\epsilon}_{t-i} + \epsilon_t$$

which minimizes a selection criterion like SBC or HQ.

According to Lütkepohl (1993), the information criteria, namely Akaike’s information criterion (AIC), the corrected Akaike’s information criterion (AICC), the final prediction error criterion (FPE), the Hannan-Quinn criterion (HQC), and the Schwarz Bayesian criterion (SBC), are defined as
AIC = \log(|\hat{\Sigma}|) + 2r_bk/T
AICC = \log(|\hat{\Sigma}|) + 2r_bk/(T - r_b)
FPE = \left(\frac{T + r_b}{T - r_b}\right)^{k} |\hat{\Sigma}|
HQC = \log(|\hat{\Sigma}|) + 2r_bk \log(\log(T))/T
SBC = \log(|\hat{\Sigma}|) + r_bk \log(T)/T

where \hat{\Sigma} is the maximum likelihood estimate of the innovation covariance matrix \Sigma, r_b is the number of parameters in each mean equation, k is the number of dependent variables, and T is the number of observations used to estimate the model. Compared to the definitions of AIC, AICC, HQC, and SBC discussed in the section “Multivariate Model Diagnostic Checks” on page 3110, the preceding definitions omit some constant terms and are normalized by T. More specifically, only the parameters in each of the mean equations are counted; the parameters in the innovation covariance matrix \Sigma are not counted.

The following statements use the MINIC= option to compute a table that contains the information criterion associated with various AR and MA orders:

```plaintext
proc varmax data=simul1;
   model y1 y2 / p=1 noint minic=(p=3 q=3);
run;
```

Figure 43.62 shows the output associated with the MINIC= option. The criterion takes the smallest value at AR order 1.

### Figure 43.62 MINIC= Option

The **VAR and VARX Modeling**

The pth-order VAR process is written as

\[ y_t - \mu = \sum_{i=1}^{p} \Phi_i (y_{t-i} - \mu) + \epsilon_t \quad \text{or} \quad \Phi(B) (y_t - \mu) = \epsilon_t \]

with \( \Phi(B) = I_k - \sum_{i=1}^{p} \Phi_i B^i \).
Equivalently, it can be written as

\[ y_t = \delta + \sum_{i=1}^{p} \Phi_i y_{t-i} + \epsilon_t \quad \text{or} \quad \Phi(B) y_t = \delta + \epsilon_t \]

with \( \delta = (I_k - \sum_{i=1}^{p} \Phi_i) \mu \).

**Stationarity**

For stationarity, the VAR process must be expressible in the convergent causal infinite MA form as

\[ y_t = \mu + \sum_{j=0}^{\infty} \Psi_j \epsilon_{t-j} \]

where \( \Psi(B) = \Phi(B)^{-1} = \sum_{j=0}^{\infty} \Psi_j B^j \) with \( \sum_{j=0}^{\infty} ||\Psi_j|| < \infty \), where \( ||A|| \) denotes a norm for the matrix \( A \) such as \( ||A||^2 = \text{tr}(A'A) \). The matrix \( \Psi_j \) can be recursively obtained from the relation \( \Phi(B)\Psi(B) = I \); it is

\[ \Psi_j = \Phi_1 \Psi_{j-1} + \Phi_2 \Psi_{j-2} + \cdots + \Phi_p \Psi_{j-p} \]

where \( \Psi_0 = I_k \) and \( \Psi_j = 0 \) for \( j < 0 \).

The stationarity condition is satisfied if all roots of \( |\Phi(z)| = 0 \) are outside of the unit circle. The stationarity condition is equivalent to the condition in the corresponding VAR(1) representation, \( Y_t = \Phi Y_{t-1} + \epsilon_t \), that all eigenvalues of the \( kp \times kp \) companion matrix \( \Phi \) be less than one in absolute value, where \( Y_t = (y'_t, \ldots, y'_{t-p+1})' \), \( \epsilon_t = (\epsilon'_t, 0', \ldots, 0')' \), and

\[
\Phi = \begin{bmatrix}
\Phi_1 & \Phi_2 & \cdots & \Phi_{p-1} & \Phi_p \\
I_k & 0 & \cdots & 0 & 0 \\
0 & I_k & \cdots & 0 & 0 \\
\vdots & \vdots & \ddots & \vdots & \vdots \\
0 & 0 & \cdots & I_k & 0
\end{bmatrix}
\]

If the stationarity condition is not satisfied, a nonstationary model (a differenced model or an error correction model) might be more appropriate.

The following statements estimate a VAR(1) model and use the ROOTS option to compute the characteristic polynomial roots:

```
proc varmax data=simul1;
    model y1 y2 / p=1 noint print=(roots);
run;
```

Figure 43.63 shows the output associated with the ROOTS option, which indicates that the series is stationary since the modulus of the eigenvalue is less than one.
Parameter Estimation

Consider the stationary $\text{VAR}(p)$ model

$$y_t = \delta + \sum_{i=1}^{p} \Phi_i y_{t-i} + \epsilon_t$$

where $y_{-p+1}, \ldots, y_0$ are assumed to be available (for convenience of notation). This can be represented by the general form of the multivariate linear model,

$$Y = XB + E \quad \text{or} \quad y = (X \otimes I_k)\beta + e$$

where

$$Y = (y_1, \ldots, y_T)'$$
$$B = (\delta, \Phi_1, \ldots, \Phi_p)'$$
$$X = (X_0, \ldots, X_{T-1})'$$
$$X_t = (1, y_t', \ldots, y_{t-p+1}')'$$
$$E = (\epsilon_1, \ldots, \epsilon_T)'$$
$$y = \text{vec}(Y')$$
$$\beta = \text{vec}(B')$$
$$e = \text{vec}(E')$$

with vec denoting the column stacking operator.

The conditional least squares estimator of $\beta$ is

$$\hat{\beta} = ((X'X)^{-1} X' \otimes I_k)y$$

and the estimate of $\Sigma$ is

$$\hat{\Sigma} = (T - (kp + 1))^{-1} \sum_{i=1}^{T} \epsilon_i \epsilon_i'$$

where $\epsilon_i$ is the residual vectors. Consistency and asymptotic normality of the LS estimator are that

$$\sqrt{T}(\hat{\beta} - \beta) \xrightarrow{d} N(0, \Gamma_p^{-1} \otimes \Sigma)$$

where $X'X/T$ converges in probability to $\Gamma_p$ and $\xrightarrow{d}$ denotes convergence in distribution.

The (conditional) maximum likelihood estimator in the $\text{VAR}(p)$ model is equal to the (conditional) least squares estimator on the assumption of normality of the error vectors.
Asymptotic Distributions of Impulse Response Functions

As before, \( \text{vec} \) denotes the column stacking operator and \( \text{vech} \) is the corresponding operator that stacks the elements on and below the diagonal. For any \( k \times k \) matrix \( A \), the commutation matrix \( K_k \) is defined as \( K_k \text{vec}(A) = \text{vec}(A') \); the duplication matrix \( D_k \) is defined as \( D_k \text{vec}(A) = \text{vec}(A) \); the elimination matrix \( L_k \) is defined as \( L_k \text{vec}(A) = \text{vech}(A) \).

The asymptotic distribution of the impulse response function (Lütkepohl 1993) is

\[
\sqrt{T} \text{vec}(\hat{\psi}_j - \Psi_j) \xrightarrow{d} N(0, G_j \Sigma_\beta G_j') \quad j = 1, 2, \ldots
\]

where \( \Sigma_\beta = \Gamma_p^{-1} \otimes \Sigma \) and

\[
G_j = \frac{\partial \text{vec}(\Psi_j)}{\partial \beta'} = \sum_{i=0}^{j-1} J(\Phi')^{j-1-i} \otimes \Psi_i
\]

where \( J = [I_k, 0, \ldots, 0] \) is a \( k \times kp \) matrix and \( \Phi \) is a \( kp \times kp \) companion matrix.

The asymptotic distribution of the accumulated impulse response function is

\[
\sqrt{T} \text{vec}(\hat{\Psi}_l^a - \Psi_l^a) \xrightarrow{d} N(0, F_l \Sigma_\beta F_l') \quad l = 1, 2, \ldots
\]

where \( F_l = \sum_{j=1}^l G_j \).

The asymptotic distribution of the orthogonalized impulse response function is

\[
\sqrt{T} \text{vec}(\hat{\psi}_j^o - \psi_j^o) \xrightarrow{d} N(0, C_j \Sigma_\beta C_j' + \tilde{C}_j \Sigma_\sigma \tilde{C}_j') \quad j = 0, 1, 2, \ldots
\]

where \( C_0 = 0, C_j = (\Psi_0' \otimes I_k)G_j, \tilde{C}_j = (I_k \otimes \Psi_j)H \),

\[
H = \frac{\partial \text{vec}(\Psi_0^o)}{\partial \sigma'} = L_k'k(2L_k(I_k^2 + K_k)(\Psi_0^o \otimes I_k)L_k')^{-1}
\]

and \( \Sigma_\sigma = 2D_k^+(\Sigma \otimes \Sigma)D_k^+ \) with \( D_k^+ = (D_k' D_k)^{-1}D_k' \) and \( \sigma = \text{vech}(E) \).

Granger Causality Test

Let \( y_t \) be arranged and partitioned in subgroups \( y_{1t} \) and \( y_{2t} \) with dimensions \( k_1 \) and \( k_2 \), respectively \( (k = k_1 + k_2) \); that is, \( y_t = (y_{1t}', y_{2t}')' \) with the corresponding white noise process \( \epsilon_t = (\epsilon_{1t}', \epsilon_{2t}')' \). Consider the VAR(\( p \)) model with partitioned coefficients \( \Phi_{ij}(B) \) for \( i, j = 1, 2 \) as follows:

\[
\begin{bmatrix}
\Phi_{11}(B) & \Phi_{12}(B) \\
\Phi_{21}(B) & \Phi_{22}(B)
\end{bmatrix}
\begin{bmatrix}
y_{1t} \\
y_{2t}
\end{bmatrix}
= 
\begin{bmatrix}
\delta_1 \\
\delta_2
\end{bmatrix}
+ 
\begin{bmatrix}
\epsilon_{1t} \\
\epsilon_{2t}
\end{bmatrix}
\]

The variables \( y_{1t} \) are said to cause \( y_{2t} \), but \( y_{2t} \) do not cause \( y_{1t} \) if \( \Phi_{12}(B) = 0 \). The implication of this model structure is that future values of the process \( y_{1t} \) are influenced only by its own past and not by the past of \( y_{2t} \), where future values of \( y_{2t} \) are influenced by the past of both \( y_{1t} \) and \( y_{2t} \). If the future \( y_{1t} \) are not influenced by the past values of \( y_{2t} \), then it can be better to model \( y_{1t} \) separately from \( y_{2t} \).
Consider testing \( H_0: C \beta = c \), where \( C \) is a \( s \times (k^2 p + k) \) matrix of rank \( s \) and \( c \) is an \( s \)-dimensional vector where \( s = k_1 k_2 p \). Assuming that

\[
\sqrt{T}(\hat{\beta} - \beta) \overset{d}{\rightarrow} N(0, \Gamma_p^{-1} \otimes \Sigma)
\]

you get the Wald statistic

\[
T(C \hat{\beta} - c)' [C(\Gamma_p^{-1} \otimes \hat{\Sigma})C']^{-1} (C \hat{\beta} - c) \overset{d}{\rightarrow} \chi^2(s)
\]

For the Granger causality test, the matrix \( C \) consists of zeros or ones and \( c \) is the zero vector. For more information about the Granger causality test, see Lütkepohl (1993).

**VARX Modeling**

The vector autoregressive model with exogenous variables is called the \( \text{VARX}(p,s) \) model. The form of the \( \text{VARX}(p,s) \) model can be written as

\[
y_t = \delta + \sum_{i=1}^{p} \Phi_i y_{t-i} + \sum_{i=0}^{s} \Theta^*_i x_{t-i} + \epsilon_t
\]

The parameter estimates can be obtained by representing the general form of the multivariate linear model,

\[
Y = XB + E \quad \text{or} \quad y = (X \otimes I_k) \beta + e
\]

where

\[
\begin{align*}
Y & = (y_1, \ldots, y_T)' \\
B & = (\delta, \Phi_1, \ldots, \Phi_p, \Theta^*_0, \ldots, \Theta^*_s)' \\
X & = (X_0, \ldots, X_{T-1})' \\
X_t & = (1, y'_t, \ldots, y'_{t-p+1}, x'_{t+1}, \ldots, x'_{t-s+1})' \\
E & = (\epsilon_1, \ldots, \epsilon_T)' \\
y & = \text{vec}(Y') \\
\beta & = \text{vec}(B') \\
e & = \text{vec}(E')
\end{align*}
\]

The conditional least squares estimator of \( \beta \) can be obtained by using the same method in a \( \text{VAR}(p) \) modeling. If the multivariate linear model has different independent variables that correspond to dependent variables, the \( \text{SUR} \) (seemingly unrelated regression) method is used to improve the regression estimates.

The following example fits the ordinary regression model:

```sas
proc varmax data=one;
    model y1-y3 = x1-x5;
run;
```

This is equivalent to the \text{REG} procedure in the SAS/STAT software:
The following example fits the second-order lagged regression model:

```
proc varmax data=two;
    model y1 y2 = x / xlag=2;
run;
```

This is equivalent to the REG procedure in the SAS/STAT software:

```
data three;
    set two;
    xlag1 = lag1(x);
    xlag2 = lag2(x);
run;
```

```
proc reg data=three;
    model y1 = x xlag1 xlag2;
    model y2 = x xlag1 xlag2;
run;
```

The following example fits the ordinary regression model with different regressors:

```
proc varmax data=one;
    model y1 = x1-x3, y2 = x2 x3;
run;
```

This is equivalent to the following SYSLIN procedure statements:

```
proc syslin data=one vardef=df sur;
    endogenous y1 y2;
    model y1 = x1-x3;
    model y2 = x2 x3;
run;
```

From the output in Figure 43.25 in the section “Getting Started: V ARMAX Procedure” on page 2982, you can see that the parameters, XL0_1_2, XL0_2_1, XL0_3_1, and XL0_3_2 associated with the exogenous variables, are not significant. The following example fits the V ARX(1,0) model with different regressors:

```
proc varmax data=grunfeld;
    model y1 = x1, y2 = x2, y3 / p=1 print=(estimates);
run;
```

**Figure 43.64** Parameter Estimates for the VARX(1, 0) Model

<table>
<thead>
<tr>
<th>XLag</th>
<th>Variable</th>
<th>x1</th>
<th>x2</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>y1</td>
<td>1.83231</td>
<td></td>
</tr>
<tr>
<td></td>
<td>y2</td>
<td></td>
<td>2.42110</td>
</tr>
<tr>
<td></td>
<td>y3</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Chapter 43: The VARMAX Procedure

As you can see in Figure 43.64, the symbol ‘_’ in the elements of matrix corresponds to endogenous variables that do not take the denoted exogenous variables.

### Seasonal Dummies and Time Trends

You can use the NSEASON= option to introduce seasonal dummies into the model, and the TREND= option to introduce linear trend or both linear and quadratic trends into the model. The definition of the seasonal dummies and trends starts from the first observation after skipping the presample and the observations that have missing values. The size of the presample is \( \max(p, s) \), where \( p \) is the maximum number of lags of AR terms and \( s \) is the maximum number of lags of exogenous variables; that is, the presample contains \( \{y_{-l+1}, x_{-l+1}, \ldots, y_0, x_0\} \), where \( l = \max(p, s) \).

The following statements fit a bivariate VARX(1, 2) model that has four seasonal periods and both linear and quadratic time trends:

```sas
data One;
  format date date9.;
  do obs = 1 to 100;
    date=intnx('quarter','01Jan1990'd,obs-1);
    y1 = normal(1); y2 = normal(1); x = normal(1);
    output;
  end;
run;

proc varmax data=One;
  model y1 y2 = x / nseason=4 xlag=2 p=1 trend=quad;
run;
```

In the following statements, the seasonal dummies and time trends are explicitly defined in the data set, together with the lags of dependent and exogenous variables, and then the equivalent model is fit by the REG procedure in SAS/STAT software:

```sas
data Two;
  set one;
  y1lag1 = lag(y1); y2lag1 = lag(y2);
  xlag1 = lag(x); xlag2 = lag2(x);
  if (obs>2) then do;
    ltrend = obs - 2;
    qtrend = ltrend * ltrend;
    const = 1;
    if (mod(ltrend-2,4)=0) then sd1 = 1;
    else sd1 = 0;
    if (mod(ltrend-3,4)=0) then sd2 = 1;
    else sd2 = 0;
    if (mod(ltrend-4,4)=0) then sd3 = 1;
    else sd3 = 0;
  end;
run;

proc reg data=Two(firstobs=3);
  model y1 = const sd1 sd2 sd3 ltrend qtrend x xlag1 xlag2 y1lag1 y2lag1 / noint;
```

Bayesian VAR and VARX Modeling

Consider the VAR($p$) model

$$y_t = \delta + \Phi_1 y_{t-1} + \cdots + \Phi_p y_{t-p} + \epsilon_t$$

or

$$y = (X \otimes I_k)\beta + e$$

When the parameter vector $\beta$ has a prior multivariate normal distribution with known mean $\beta^*$ and covariance matrix $V_\beta$, the prior density is written as

$$f(\beta) = \left(\frac{1}{2\pi}\right)^{k^2 p/2} |V_\beta|^{-1/2} \exp\left[-\frac{1}{2}(\beta - \beta^*)' V_\beta^{-1} (\beta - \beta^*)\right]$$

The likelihood function for the Gaussian process becomes

$$\ell(\beta|y) = \left(\frac{1}{2\pi}\right)^{kT/2} |I_T \otimes \Sigma|^{-1/2} \times \exp\left[-\frac{1}{2}(y - (X \otimes I_k)\beta)'(I_T \otimes \Sigma^{-1})(y - (X \otimes I_k)\beta)\right]$$
Therefore, the posterior density is derived as
\[
f(\beta | y) \propto \exp[-\frac{1}{2}(\beta - \bar{\beta})' \Sigma_{\beta}^{-1} (\beta - \bar{\beta})]
\]
where the posterior mean is
\[
\bar{\beta} = [V_{\beta}^{-1} + (X'X \otimes \Sigma^{-1})]^{-1} [V_{\beta}^{-1} \beta^* + (X' \otimes \Sigma^{-1})y]
\]
and the posterior covariance matrix is
\[
\Sigma_{\beta} = [V_{\beta}^{-1} + (X'X \otimes \Sigma^{-1})]^{-1}
\]
In practice, the prior mean $\beta^*$ and the prior variance $V_{\beta}$ need to be specified. If all the parameters are considered to shrink toward zero, the null prior mean should be specified. According to Litterman (1986), the prior variance can be given by
\[
v_{ij}(l) = \begin{cases} 
(\lambda/l)^2 & \text{if } i = j \\
(\lambda_\theta \sigma_{ii}/l\sigma_{jj})^2 & \text{if } i \neq j 
\end{cases}
\]
where $v_{ij}(l)$ is the prior variance of the $(i, j)$ element of $\Phi_l$, $\lambda$ is the prior standard deviation of the diagonal elements of $\Phi_l$, $\theta$ is a constant in the interval $(0, 1)$, and $\sigma_{ii}^2$ is the $i$th diagonal element of $\Sigma$. The deterministic terms have diffused prior variance. In practice, you replace the $\sigma_{ii}^2$ by the diagonal element of the ML estimator of $\Sigma$ in the nonconstrained model.

For example, for a bivariate BVAR(2) model,
\[
\begin{align*}
y_{1t} &= 0 + \phi_{1,11} y_{1,t-1} + \phi_{1,12} y_{2,t-1} + \phi_{1,21} y_{1,t-2} + \phi_{1,22} y_{2,t-2} + \epsilon_{1t} \\
y_{2t} &= 0 + \phi_{2,11} y_{1,t-1} + \phi_{2,21} y_{2,t-1} + \phi_{2,22} y_{1,t-2} + \phi_{2,22} y_{2,t-2} + \epsilon_{2t}
\end{align*}
\]
with the prior covariance matrix
\[
V_{\beta} = \text{Diag} \left( \infty, \lambda^2, (\lambda_\theta \sigma_1/\sigma_2)^2, (\lambda/2)^2, (\lambda_\theta \sigma_1/2\sigma_2)^2, (\lambda_\theta \sigma_2/\sigma_1)^2, \lambda^2, (\lambda_\theta \sigma_2/2\sigma_1)^2, (\lambda/2)^2 \right)
\]
For the Bayesian estimation of integrated systems, the prior mean is set to the first lag of each variable equal to one in its own equation and all other coefficients at zero. For example, for a bivariate BVAR(2) model,
\[
\begin{align*}
y_{1t} &= 0 + 1 y_{1,t-1} + 0 y_{2,t-1} + 0 y_{1,t-2} + 0 y_{2,t-2} + \epsilon_{1t} \\
y_{2t} &= 0 + 0 y_{1,t-1} + 1 y_{2,t-1} + 0 y_{1,t-2} + 0 y_{2,t-2} + \epsilon_{2t}
\end{align*}
\]
Forecasting of BVAR Modeling

The mean squared error (MSE) is used to measure forecast accuracy (Litterman 1986). The MSE of the \( s \)-step-ahead forecast is

\[
\text{MSE}_s = \frac{1}{J-s+1} \sum_{j=1}^{J-s+1} (A_{t_j} - F_{t_j}^s)^2
\]

where \( J \) is the number specified by NREP= option, \( t_j \) is the time index of the observation to be forecasted in repetition \( j \), \( A_{t_j} \) is the actual value at time \( t_j \), and \( F_{t_j}^s \) is the forecast made \( s \) periods earlier. If there are not enough observations, some MSEs might not be calculated.

Bayesian VARX Modeling

The Bayesian vector autoregressive model with exogenous variables is called the BVARX(\( p, s \)) model. The form of the BVARX(\( p, s \)) model can be written as

\[
y_t = \delta + \sum_{i=1}^{p} \Phi_i y_{t-i} + \sum_{i=0}^{s} \Theta_i^* x_{t-i} + \epsilon_t
\]

The parameter estimates can be obtained by representing the general form of the multivariate linear model,

\[
y = (X \otimes I_k) \beta + \epsilon
\]

The prior means for the AR coefficients are the same as those specified in BVAR(\( p \)). The prior means for the exogenous coefficients are set to zero.

Some examples of the Bayesian VARX model are as follows:

```plaintext
model y1 y2 = x1 / p=1 xlag=1 prior;
model y1 y2 = x1 / p=(1 3) xlag=1 nocurrentx
prior=(lambda=0.9 theta=0.1);
```

VARMA and VARMAX Modeling

A zero-mean VARMA(\( p, q \)) process is written as

\[
y_t = \sum_{i=1}^{p} \Phi_i y_{t-i} + \epsilon_t - \sum_{i=1}^{q} \Theta_i \epsilon_{t-i}
\]

or

\[
\Phi(B)y_t = \Theta(B)\epsilon_t
\]

where \( \Phi(B) = I_k - \sum_{i=1}^{p} \Phi_i B^i \) and \( \Theta(B) = I_k - \sum_{i=1}^{q} \Theta_i B^i \).
Stationarity and Invertibility

For stationarity and invertibility of the VARMA process, the roots of $|\Phi(z)| = 0$ and $|\Theta(z)| = 0$ are outside the unit circle.

Parameter Estimation

Under the assumption of normality of the $\epsilon_t$ with zero-mean vector and nonsingular covariance matrix $\Sigma$, the conditional (approximate) log-likelihood function of a zero-mean VARMA($p,q$) model is considered.

Define $Y = (y_1, \ldots, y_T)'$ and $E = (\epsilon_1, \ldots, \epsilon_T)'$ with $B^i Y = (y_{1-i}, \ldots, y_{T-i})'$ and $B^i E = (\epsilon_{1-i}, \ldots, \epsilon_{T-i})'$; define $y = \text{vec}(Y')$ and $e = \text{vec}(E')$. Then

$$y - \sum_{i=1}^{p} (I_T \otimes \Phi_i) B^i y = e - \sum_{i=1}^{q} (I_T \otimes \Theta_i) B^i e$$

where $B^i y = \text{vec}[(B^i Y)']$ and $B^i e = \text{vec}[(B^i E)']$.

Then, the conditional (approximate) log-likelihood function can be written as (Reinsel 1997)

$$\ell = -\frac{T}{2} \log |\Sigma| - \frac{1}{2} \sum_{t=1}^{T} \epsilon'_t \Sigma^{-1} \epsilon_t$$

$$= -\frac{T}{2} \log |\Sigma| - \frac{1}{2} w' \Theta^{-1} (I_T \otimes \Sigma^{-1}) \Theta^{-1} w$$

where $w = y - \sum_{i=1}^{p} (I_T \otimes \Phi_i) B^i y$ and $\Theta$ is such that $e - \sum_{i=1}^{q} (I_T \otimes \Theta_i) B^i e = \Theta e$. You can specify METHOD=CML in the MODEL statement to apply conditional maximum likelihood estimation.

For the exact log-likelihood function of a VARMA model, the VARMA model is transformed into the equivalent state space form and then the Kalman filtering method is applied.

The state space form of the zero-mean VARMA($p,q$) model consists of a state equation

$$z_t = F z_{t-1} + G \epsilon_t$$

and an observation equation

$$y_t = H z_t$$

where

$$z_t = (y'_t, y'_{t-1}, \ldots, y'_{t-(v-1)}, \epsilon'_t, \epsilon_{t-1}, \ldots, \epsilon'_{t-(q-1)})'$$

$$F = \begin{bmatrix} \Phi_1 & \cdots & \Phi_{v-1} & \Phi_v & -\Theta_1 & \cdots & -\Theta_{q-1} & -\Theta_q \\ I_k & \cdots & 0 & 0 & 0 & \cdots & 0 & 0 \\ \vdots & \ddots & 0 \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 \cdots & I_k & 0 & 0 & \cdots & 0 & 0 \\ 0 \cdots & 0 & 0 \cdots & 0 & 0 \cdots & 0 & 0 \\ 0 \cdots & 0 & 0 & I_k \cdots & 0 & 0 \\ \vdots & \ddots & 0 \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 \cdots & 0 & 0 & 0 \cdots & I_k & 0 \end{bmatrix}, \quad G = \begin{bmatrix} I_k \\ 0_{k(v-1)\times k} \\ I_k \\ 0_{k(q-1)\times k} \end{bmatrix}$$
VARMA and VARMAX Modeling

and

\[ H = [I_k, 0_{k(v+q-1)\times k}] \]

where \( v = \max(p, 1) \) and \( \Phi_i = 0 \) for \( i > p \).

The Kalman filtering approach is used to evaluate the likelihood function. The updating equation is

\[ \hat{z}_t|t = \hat{z}_{t|t-1} + K_t \epsilon_{t|t-1} \]

where

\[ K_t = P_{t|t-1} H'[HP_{t|t-1}H']^{-1} \]

The prediction equation is

\[ \hat{z}_{t-1} = F \hat{z}_{t-1|t-1}, \quad P_{t|t-1} = FP_{t-1|t-1}F' + G \Sigma G' \]

where \( P_{t|t} = [I - K_t H]P_{t|t-1} \) for \( t = 1, 2, \ldots, n \).

The log-likelihood function can be expressed as

\[ \ell = -\frac{1}{2} \sum_{t=1}^{T} \left[ \log |\Sigma_{t|t-1}| + (y_t - \hat{y}_{t|t-1})' \Sigma_{t|t-1}^{-1} (y_t - \hat{y}_{t|t-1}) \right] \]

where \( \hat{y}_{t|t-1} \) and \( \Sigma_{t|t-1} \) are determined recursively from the Kalman filtering method. To construct the likelihood function from Kalman filtering, you obtain \( \hat{y}_{t|t-1} = H \hat{z}_{t|t-1}, \hat{\epsilon}_{t|t-1} = y_t - \hat{y}_{t|t-1} \), and \( \Sigma_{t|t-1} = HP_{t|t-1}H' \).

When you specify METHOD=ML in the MODEL statement, the exact log likelihood is evaluated and used in the maximum likelihood estimation.

Define the vector \( \beta \) as

\[ \beta = (\phi_1', \ldots, \phi_p', \theta_1', \ldots, \theta_q', \text{vech}(\Sigma))' \]

where \( \phi_i = \text{vec}(\Phi_i) \) and \( \theta_i = \text{vec}(\Theta_i) \). All elements of \( \beta \) are estimated through the preceding (conditional) maximum likelihood method. The estimates of \( \Phi_i, \theta_i, i = 1, \ldots, p, \) and \( \Theta_i, i = 1, \ldots, q, \) are output in the ParameterEstimates ODS table. The estimates of the covariance matrix (\( \Sigma \)) are output in the CovarianceParameterEstimates ODS table. If you specify the OUTTEST=, OUTCOV, PRINT=(COVB), or PRINT=(CORRB) option, you can see all elements of \( \beta \), including the covariance matrix \( \Sigma \), in the parameter estimates, covariance of parameter estimates, or correlation of parameter estimates. You can also apply the BOUND, INITIAL, RESTRICT, and TEST statements to any elements of \( \beta \), including the covariance matrix \( \Sigma \). For more information, see the syntax of the corresponding statement.

The (conditional) log-likelihood equations are solved by iterative numerical methods such as quasi-Newton optimization. The starting values for the AR and MA parameters are obtained from the least squares estimates. Although the small-sample properties of CML estimates might not be as good as the ML estimates, the CML method is much faster than the ML method. Depending on the sample size and number of parameters to be estimated, the CML method can be hundreds or even thousands of times faster than the ML method. In the following example code, the CML method is about 100 times faster than the ML method, with very similar estimation and forecast results:
proc iml;
phi = (0.9 * I(4)) // (-0.7* I(4));
theta = 0.8 * I(4);
sig = I(4);
/* to simulate the vector time series */
call varmasim(y,phi,theta) sigma=sig n=400 seed=2;
   cn = {'y1' 'y2' 'y3' 'y4'};
create simul6 from y[colname=cn];
append from y;
close;
quit;

proc varmax data=simul6;
   model y1 y2 y3 y4 / noint p=2 q=1 method=cml;
nloptions pall maxit=5000 tech=qn;
   output out=ocml back=12 lead=24;
run;

proc varmax data=simul6;
   model y1 y2 y3 y4 / noint p=2 q=1 method=ml;
nloptions pall maxit=5000 tech=qn;
   output out=oml back=12 lead=24;
run;

Asymptotic Distribution of the Parameter Estimates

Under the assumptions of stationarity and invertibility for the VARMA model and the assumption that $\epsilon_t$ is a white noise process, $\hat{\beta}$ is a consistent estimator for $\beta$ and $\sqrt{T}(\hat{\beta} - \beta)$ converges in distribution to the multivariate normal $N(0, V^{-1})$ as $T \to \infty$, where $V$ is the asymptotic information matrix of $\beta$.

Asymptotic Distributions of Impulse Response Functions

Defining the vector $\beta$

$\beta = (\phi'_1, \ldots, \phi'_p, \theta'_1, \ldots, \theta'_q)'$

the asymptotic distribution of the impulse response function for a VARMA($p, q$) model is

$\sqrt{T}\text{vec}(\Psi_j - \Psi_j') \overset{d}{\to} N(0, G_j \Sigma_\beta G_j') \quad j = 1, 2, \ldots$

where $\Sigma_\beta$ is the covariance matrix of the parameter estimates and

$G_j = \frac{\partial \text{vec}(\Psi_j)}{\partial \beta'} = \sum_{i=0}^{j-1} H'(A')^{j-1-i} \otimes JA'J'$

where $H = [I_k, 0, \ldots, 0, I_k, 0, \ldots, 0]'$ is a $k(p + q) \times k$ matrix with the second $I_k$ following after $p$ block matrices; $J = [I_k, 0, \ldots, 0]$ is a $k \times k(p + q)$ matrix; $A$ is a $k(p + q) \times k(p + q)$ matrix,

$A = \begin{bmatrix}
A_{11} & A_{12} \\
A_{21} & A_{22}
\end{bmatrix}$
where

\[
A_{11} = \begin{bmatrix}
\Phi_1 & \Phi_2 & \cdots & \Phi_{p-1} & \Phi_p \\
I_k & 0 & \cdots & 0 & 0 \\
0 & I_k & \cdots & 0 & 0 \\
\vdots & \vdots & \ddots & \vdots & \vdots \\
0 & 0 & \cdots & I_k & 0
\end{bmatrix}
\quad A_{12} = \begin{bmatrix}
-\Theta_1 & -\Theta_{q-1} & -\Theta_q \\
0 & \cdots & 0 & 0 \\
0 & \cdots & 0 & 0 \\
\vdots & \vdots & \ddots & \vdots \\
0 & \cdots & 0 & 0
\end{bmatrix}
\]

\(A_{21}\) is a \(kq \times kp\) zero matrix, and

\[
A_{22} = \begin{bmatrix}
0 & 0 & \cdots & 0 & 0 \\
I_k & 0 & \cdots & 0 & 0 \\
0 & I_k & \cdots & 0 & 0 \\
\vdots & \vdots & \ddots & \vdots & \vdots \\
0 & 0 & \cdots & I_k & 0
\end{bmatrix}
\]

An Example of a VARMA(1,1) Model

Consider a VARMA(1,1) model with mean zero,

\[
y_t = \Phi_1 y_{t-1} + \epsilon_t - \Theta_1 \epsilon_{t-1}
\]

where \(\epsilon_t\) is the white noise process with a mean zero vector and the positive-definite covariance matrix \(\Sigma\).

The following IML procedure statements simulate a bivariate vector time series from this model to provide test data for the VARMAX procedure:

```
proc iml;
sig = {1.0 0.5, 0.5 1.25};
phi = {1.2 -0.5, 0.6 0.3};
theta = {0.5 -0.2, 0.1 0.3};
/* to simulate the vector time series */
call varmasim(y,phi,theta) sigma=sig n=100 seed=34657;
run;
```

The following statements fit a VARMA(1,1) model to the simulated data. You specify the order of the autoregressive model by using the P= option and specify the order of moving-average model by using the Q= option. You specify the quasi-Newton optimization in the NLOPTIONS statement as an optimization method.

```
proc varmax data=simul3;
nloptions tech=qn;
model y1 y2 / p=1 q=1 noint print=(estimates);
run;
```
Figure 43.66 shows the initial values of parameters. The initial values were estimated by using the least squares method.

**Figure 43.66** Start Parameter Estimates for the VARMA(1, 1) Model

<table>
<thead>
<tr>
<th>N Parameter</th>
<th>Estimate</th>
<th>Gradient Objective Function</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 AR1_1_1</td>
<td>0.964299</td>
<td>-2.357098</td>
</tr>
<tr>
<td>2 AR1_2_1</td>
<td>0.481620</td>
<td>-3.773499</td>
</tr>
<tr>
<td>3 AR1_1_2</td>
<td>-0.363819</td>
<td>1.865051</td>
</tr>
<tr>
<td>4 AR1_2_2</td>
<td>0.457378</td>
<td>-10.778568</td>
</tr>
<tr>
<td>5 MA1_1_1</td>
<td>0.244355</td>
<td>-2.552198</td>
</tr>
<tr>
<td>6 MA1_2_1</td>
<td>-0.034093</td>
<td>2.716227</td>
</tr>
<tr>
<td>7 MA1_1_2</td>
<td>-0.006261</td>
<td>-0.147004</td>
</tr>
<tr>
<td>8 MA1_2_2</td>
<td>0.444636</td>
<td>0.141839</td>
</tr>
<tr>
<td>9 COV1_1</td>
<td>1.353584</td>
<td>2.765550</td>
</tr>
<tr>
<td>10 COV1_2</td>
<td>0.415649</td>
<td>-1.389416</td>
</tr>
<tr>
<td>11 COV2_2</td>
<td>1.445260</td>
<td>2.581735</td>
</tr>
</tbody>
</table>

Figure 43.67 shows the default option settings for the quasi-Newton optimization technique.

**Figure 43.67** Default Criteria for the quasi-Newton Optimization

- Minimum Iterations: 0
- Maximum Iterations: 200
- Maximum Function Calls: 2000
- ABSGCONV Gradient Criterion: 0.00001
- GCONV Gradient Criterion: 1E-8
- ABSFCONV Function Criterion: 0
- FCONV Function Criterion: 2.220446E-16
- FCONV2 Function Criterion: 0
-FSIZE Parameter: 0
- ABSXCONV Parameter Change Criterion: 0
- XCONV Parameter Change Criterion: 0
- XSIZE Parameter: 0
- ABSCONV Function Criterion: -1.34078E154
- Line Search Method: 2
- Starting Alpha for Line Search: 1
- Line Search Precision LSPRECISION: 0.4
- DAMPSTEP Parameter for Line Search: .
- Singularity Tolerance (SINGULAR): 1E-8
Figure 43.68 shows the iteration history of parameter estimates.

**Figure 43.68** Iteration History of Parameter Estimates

<table>
<thead>
<tr>
<th>Iteration</th>
<th>Restarts</th>
<th>Function Calls</th>
<th>Active Constraints</th>
<th>Objective Function</th>
<th>Objective Function Change</th>
<th>Max Abs Gradient Element</th>
<th>Step Size</th>
<th>Slope of Search Direction</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>3</td>
<td>0</td>
<td>121.22330</td>
<td>0.1526</td>
<td>5.2001</td>
<td>0.00384</td>
<td>-78.688</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>5</td>
<td>0</td>
<td>120.97740</td>
<td>0.2459</td>
<td>6.2584</td>
<td>3.214</td>
<td>-0.156</td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>6</td>
<td>0</td>
<td>120.58286</td>
<td>0.3945</td>
<td>4.1004</td>
<td>0.948</td>
<td>-0.648</td>
</tr>
<tr>
<td>4</td>
<td>0</td>
<td>7</td>
<td>0</td>
<td>120.43152</td>
<td>0.1513</td>
<td>3.7834</td>
<td>1.000</td>
<td>-0.346</td>
</tr>
<tr>
<td>5</td>
<td>0</td>
<td>8</td>
<td>0</td>
<td>120.32909</td>
<td>0.1016</td>
<td>6.3797</td>
<td>1.000</td>
<td>-0.243</td>
</tr>
<tr>
<td>6</td>
<td>0</td>
<td>10</td>
<td>0</td>
<td>120.26832</td>
<td>0.0616</td>
<td>3.1048</td>
<td>0.407</td>
<td>-0.304</td>
</tr>
<tr>
<td>7</td>
<td>0</td>
<td>12</td>
<td>0</td>
<td>120.23311</td>
<td>0.0352</td>
<td>1.0747</td>
<td>0.983</td>
<td>-0.0731</td>
</tr>
<tr>
<td>8</td>
<td>0</td>
<td>14</td>
<td>0</td>
<td>120.22264</td>
<td>0.0105</td>
<td>0.6370</td>
<td>1.518</td>
<td>-0.0127</td>
</tr>
<tr>
<td>9</td>
<td>0</td>
<td>15</td>
<td>0</td>
<td>120.21560</td>
<td>0.00704</td>
<td>1.3563</td>
<td>4.650</td>
<td>-0.0056</td>
</tr>
<tr>
<td>10</td>
<td>0</td>
<td>16</td>
<td>0</td>
<td>120.21281</td>
<td>0.00279</td>
<td>1.2963</td>
<td>2.102</td>
<td>-0.0084</td>
</tr>
<tr>
<td>11</td>
<td>0</td>
<td>17</td>
<td>0</td>
<td>120.20951</td>
<td>0.00330</td>
<td>0.1634</td>
<td>1.139</td>
<td>-0.0061</td>
</tr>
<tr>
<td>12</td>
<td>0</td>
<td>19</td>
<td>0</td>
<td>120.20896</td>
<td>0.000542</td>
<td>0.1349</td>
<td>2.591</td>
<td>-0.0004</td>
</tr>
<tr>
<td>13</td>
<td>0</td>
<td>21</td>
<td>0</td>
<td>120.20884</td>
<td>0.000123</td>
<td>0.0662</td>
<td>1.883</td>
<td>-0.0001</td>
</tr>
<tr>
<td>14</td>
<td>0</td>
<td>22</td>
<td>0</td>
<td>120.20875</td>
<td>0.000093</td>
<td>0.1399</td>
<td>4.120</td>
<td>-0.0001</td>
</tr>
<tr>
<td>15</td>
<td>0</td>
<td>24</td>
<td>0</td>
<td>120.20871</td>
<td>0.000037</td>
<td>0.00917</td>
<td>1.073</td>
<td>-0.0001</td>
</tr>
<tr>
<td>16</td>
<td>0</td>
<td>26</td>
<td>0</td>
<td>120.20871</td>
<td>1.643E-6</td>
<td>0.00858</td>
<td>2.115</td>
<td>-155E-8</td>
</tr>
<tr>
<td>17</td>
<td>0</td>
<td>27</td>
<td>0</td>
<td>120.20871</td>
<td>7.704E-7</td>
<td>0.00543</td>
<td>5.409</td>
<td>-759E-9</td>
</tr>
</tbody>
</table>

Figure 43.69 shows the final parameter estimates.

**Figure 43.69** Results of Parameter Estimates for the VARMA(1, 1) Model

**The VARMAX Procedure**

<table>
<thead>
<tr>
<th>N</th>
<th>Parameter</th>
<th>Estimate</th>
<th>Gradient Objective Function</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>AR1_1_1</td>
<td>1.020117</td>
<td>0.003641</td>
</tr>
<tr>
<td>2</td>
<td>AR1_2_1</td>
<td>0.393557</td>
<td>0.000140</td>
</tr>
<tr>
<td>3</td>
<td>AR1_1_2</td>
<td>-0.388708</td>
<td>0.001311</td>
</tr>
<tr>
<td>4</td>
<td>AR1_2_2</td>
<td>0.551644</td>
<td>0.002479</td>
</tr>
<tr>
<td>5</td>
<td>MA1_1_1</td>
<td>0.330598</td>
<td>0.000131</td>
</tr>
<tr>
<td>6</td>
<td>MA1_2_1</td>
<td>-0.166999</td>
<td>0.000086321</td>
</tr>
<tr>
<td>7</td>
<td>MA1_1_2</td>
<td>-0.032507</td>
<td>-0.001133</td>
</tr>
<tr>
<td>8</td>
<td>MA1_2_2</td>
<td>0.587232</td>
<td>-0.000523</td>
</tr>
<tr>
<td>9</td>
<td>COV1_1</td>
<td>1.253624</td>
<td>0.005429</td>
</tr>
<tr>
<td>10</td>
<td>COV1_2</td>
<td>0.382094</td>
<td>-0.001152</td>
</tr>
<tr>
<td>11</td>
<td>COV2_2</td>
<td>1.322424</td>
<td>-0.000535</td>
</tr>
</tbody>
</table>

Figure 43.70 shows the AR coefficient matrix in terms of lag 1, the MA coefficient matrix in terms of lag 1, the parameter estimates, and their significance, which is one indication of how well the model fits the data.
The fitted VARMA(1,1) model with estimated standard errors in parentheses is given as

\[
y_t = \begin{pmatrix} 1.01846 & -0.38682 \\ 0.10256 & 0.09644 \\ 0.39182 & 0.55281 \\ 0.10062 & 0.08422 \end{pmatrix} y_{t-1} + \begin{pmatrix} 0.32292 & -0.02160 \\ 0.14524 & 0.14203 \\ -0.16501 & 0.58576 \\ 0.15704 & 0.14115 \end{pmatrix} \epsilon_{t-1} - \begin{pmatrix} 0.30060 & -0.03251 \\ -0.16700 & 0.58723 \end{pmatrix}
\]
and

\[ \epsilon_t \sim \text{iid } N(0, \begin{pmatrix} 1.25202 & 0.37950 \\ 0.17697 & 0.13401 \\ 0.37950 & 1.31315 \\ 0.13401 & 0.18610 \end{pmatrix}) \]

**VARMAX Modeling**

A general VARMAX \((p, q, s)\) process is written as

\[ y_t = \delta_t + \sum_{i=1}^{p} \Phi_i y_{t-i} + \epsilon_t - \sum_{i=1}^{q} \Theta_i \epsilon_{t-i} \]

or

\[ \Phi(B)y_t = \delta_t + \Theta(B)\epsilon_t \]

where \(\Phi(B) = I_k - \sum_{i=1}^{p} \Phi_i B^i\) and \(\Theta(B) = I_k - \sum_{i=1}^{q} \Theta_i B^i\). The vector \(\delta_t\) consists of all possible deterministic terms, namely constant, seasonal dummies, linear trend, quadratic trend, and exogenous variables. The vector \(\delta_t = \Delta c_t\), where \(c_t = (D'_t x'_t \ldots x'_{t-s})'; D_t = (1 d_{t,1} \ldots d_{t,n_s-1} t t^2)'; d_{t,i}, i = 1, \ldots, n_s - 1,\) are seasonal dummies and \(n_s\) is based on the NSEASON= option; \(\Delta = (A \Theta^*_0 \ldots \Theta^*_s); A\) is the parameter matrix corresponding to \(D_t\) and \(\Theta_t^*\) for \(x_{t-i}, i = 0, \ldots, s\).

The state space form of the VARMAX \((p,q,s)\) model consists of a state equation

\[ z_t = F z_{t-1} + w_t + G \epsilon_t \]

and an observation equation

\[ y_t = H z_t \]

where

\[ z_t = (y'_t, y'_{t-1}, \ldots, y'_{t-(v-1)}, \epsilon'_t, \epsilon_{t-1}, \ldots, \epsilon'_{t-(q-1)}, c'_{t+1})' \]

\[ F = \begin{bmatrix} \Phi_1 & \cdots & \Phi_{v-1} & \Phi_v & -\Theta_1 & \cdots & -\Theta_{q-1} & -\Theta_q & \Delta \\ I_k & \cdots & 0 & 0 & 0 & \cdots & 0 & 0 & 0 \\ \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & \cdots & I_k & 0 & 0 & \cdots & 0 & 0 & 0 \\ 0 & \cdots & 0 & 0 & \cdots & 0 & 0 & 0 & 0 \\ 0 & \cdots & 0 & \cdots & I_k & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & \cdots & 0 & \cdots & \cdots & \vdots & \cdots & \vdots & \vdots \\ 0 & \cdots & 0 & \cdots & \cdots & \cdots & I_k & 0 & 0 \end{bmatrix}, \quad G = \begin{bmatrix} I_k \\ 0_{k(v-1) \times k} \\ I_k \\ 0_{k(q-1) \times k} \\ 0_{n \times k} \end{bmatrix} \]

and

\[ H = [I_k, 0_{(k+v-1)+u} \times k] \]
where $v = \max(p, 1)$, $\Phi_t = 0$ for $i > p$, and $u$ is the dimension of $c_t$.

Kalman filtering is used to evaluate the likelihood function. The updating equation is
\[
\hat{z}_{t|t} = \hat{z}_{t|t-1} + K_t \epsilon_{t|t-1}
\]
where
\[
K_t = P_{t|t-1} H' [HP_{t|t-1} H']^{-1}
\]
The prediction equation is
\[
\hat{z}_{t-1} = F \hat{z}_{t-1|t-1} + w_t, \quad P_{t|t-1} = FP_{t|t-1} F' + G \Sigma G'
\]
where $P_{t|t} = [I - K_t H] P_{t|t-1}$ for $t = 1, 2, \ldots, n$.

The log-likelihood function can be expressed as
\[
\ell = -\frac{1}{2} \sum_{t=1}^{T} [\log |\Sigma_{t|t-1}| + (y_t - \hat{y}_{t|t-1})' \Sigma_{t|t-1}^{-1} (y_t - \hat{y}_{t|t-1})]
\]
where $\hat{y}_{t|t-1}$ and $\Sigma_{t|t-1}$ are determined recursively from Kalman filtering. To construct the likelihood function from Kalman filtering, you obtain $\hat{y}_{t|t-1} = H \hat{z}_{t|t-1}$, $\epsilon_{t|t-1} = y_t - \hat{y}_{t|t-1}$, and $\Sigma_{t|t-1} = HP_{t|t-1} H'$.

In the preceding state space form of a VARMAX model, the exogenous variables are treated as determined terms, which implies that the values of the exogenous variables must be provided to forecast the out-of-sample dependent variables. If you do not have the future values of the exogenous variables, either you predict the exogenous variables in a separate model, or you express both the exogenous variables and the dependent variables in one combined model and predict them together (Reinsel 1997).

The dimension of the state space vector of the Kalman filtering method for the VARMAX($p,q,s$) model might be large, so it might take a lot of time and memory for computing.

Two examples of VARMAX modeling follow:

```
model y1 y2 = x1 / q=1;
nloptions tech=qn;

model y1 y2 = x1 / p=1 q=1 xlag=1 nocurrentx;
nloptions tech=qn;
```

## Model Diagnostic Checks

### Multivariate Model Diagnostic Checks

#### Log Likelihood

The log-likelihood function for the fitted model is reported in the LogLikelihood ODS table. The log-likelihood functions for different models are defined as follows:

- For VARMAX models that are estimated through the (conditional) maximum likelihood method, see the section “VARMA and VARMAX Modeling” on page 3101.
For Bayesian VAR and VARX models, see the section “Bayesian VAR and VARX Modeling” on page 3099.

For (Bayesian) vector error correction models, see the section “Vector Error Correction Modeling” on page 3115.

For multivariate GARCH models, see the section “Multivariate GARCH Modeling” on page 3135.

For VARFIMA and VARFIMAX models, see the section “VARFIMA and VARFIMAX Modeling” on page 3146.

For VAR and VARX models that are estimated through the least squares (LS) method, the log likelihood is defined as

$$\ell = -\frac{1}{2}(T \log |\tilde{\Sigma}| + kT)$$

where $\tilde{\Sigma}$ is the maximum likelihood estimate of the innovation covariance matrix, $k$ is the number of dependent variables, and $T$ is the number of observations used in the estimation.

**Information Criteria**

The information criteria include Akaike’s information criterion (AIC), the corrected Akaike’s information criterion (AICC), the final prediction error criterion (FPE), the Hannan-Quinn criterion (HQC), and the Schwarz Bayesian criterion (SBC, also referred to as BIC). These criteria are defined as

$$\text{AIC} = -2\ell + 2r$$
$$\text{AICC} = -2\ell + 2rT/(T - r - 1)$$
$$\text{FPE} = \left(\frac{T + r_b}{T - r_b}\right)^k |\tilde{\Sigma}|$$
$$\text{HQC} = -2\ell + 2r \log(\log(T))$$
$$\text{SBC} = -2\ell + r \log(T)$$

where $\ell$ is the log likelihood, $r$ is the total number of parameters in the model, $k$ is the number of dependent variables, $T$ is the number of observations that are used to estimate the model, $r_b$ is the number of parameters in each mean equation, and $\tilde{\Sigma}$ is the maximum likelihood estimate of $\Sigma$. As suggested by Burnham and Anderson (2004) for least squares estimation, the total number of parameters, $r$, must include the parameters in the innovation covariance matrix. When comparing models, choose the model that has the smallest criterion values.

For an example of the output, see Figure 43.4 earlier in this chapter.

**Portmanteau Statistic**

The portmanteau statistic, $Q_s$, is used to test whether correlation remains on the model residuals. The null hypothesis is that the residuals are uncorrelated. Let $C_\varepsilon(l)$ be the residual cross-covariance matrices, $\hat{\rho}_\varepsilon(l)$ be the residual cross-correlation matrices as

$$C_\varepsilon(l) = T^{-1} \sum_{t=1}^{T-l} \varepsilon_t \varepsilon'_{t+l}$$
and
\[ \hat{\rho}_e(l) = \hat{V}^{-1/2}C(l)\hat{V}^{-1/2} \quad \text{and} \quad \hat{\rho}_e(-l) = \hat{\rho}_e(l)' \]

where \( \hat{V} = \text{Diag}(\hat{\sigma}_{11}^2, \ldots, \hat{\sigma}_{kk}^2) \) and \( \hat{\sigma}_{ii}^2 \) are the diagonal elements of \( \hat{\Sigma} \). The multivariate portmanteau test defined in Hosking (1980) is
\[ Q_s = T^2 \sum_{l=1}^{s} (T - l)^{-1} \text{tr}\{\hat{\rho}_e(l)\hat{\rho}_e(0)^{-1}\hat{\rho}_e(-l)\hat{\rho}_e(0)^{-1}\} \]

The statistic \( Q_s \) has approximately the chi-square distribution with \( k^2(s - p - q) \) degrees of freedom. An example of the output is displayed in Figure 43.7.

Univariate Model Diagnostic Checks

There are various ways to perform diagnostic checks for a univariate model. For more information, see the section “Testing for Nonlinear Dependence: Heteroscedasticity Tests” on page 403 in Chapter 8, “The AUTOREG Procedure.” An example of the output is displayed in Figure 43.8 and Figure 43.9.

- Durbin-Watson (DW) statistics: The DW test statistics test for the first order autocorrelation in the residuals.
- Jarque-Bera normality test: This test is helpful in determining whether the model residuals represent a white noise process. This tests the null hypothesis that the residuals have normality.
- \( F \) tests for autoregressive conditional heteroscedastic (ARCH) disturbances: \( F \) test statistics test for the heteroscedastic disturbances in the residuals. This tests the null hypothesis that the residuals have equal covariances.
- \( F \) tests for AR disturbance: These test statistics are computed from the residuals of the univariate AR(1), AR(1,2), AR(1,2,3), and AR(1,2,3,4) models to test the null hypothesis that the residuals are uncorrelated.

Cointegration

This section briefly introduces the concepts of cointegration (Johansen 1995a).

Definition 1. (Engle and Granger 1987): If a series \( y_t \) with no deterministic components can be represented by a stationary and invertible ARMA process after differencing \( d \) times, the series is integrated of order \( d \), that is, \( y_t \sim I(d) \).

Definition 2. (Engle and Granger 1987): If all elements of the vector \( y_t \) are \( I(d) \) and there exists a cointegrating vector \( \beta \neq 0 \) such that \( \beta'y_t \sim I(d - b) \) for any \( b > 0 \), the vector process is said to be cointegrated \( CI(d, b) \).
A simple example of a cointegrated process is the following bivariate system:

\[
\begin{align*}
y_{1t} &= \gamma y_{2t} + \epsilon_{1t} \\
y_{2t} &= y_{2,t-1} + \epsilon_{2t}
\end{align*}
\]

with \(\epsilon_{1t}\) and \(\epsilon_{2t}\) being uncorrelated white noise processes. In the second equation, \(y_{2t}\) is a random walk, \(\Delta y_{2t} = \epsilon_{2t}, \Delta \equiv 1 - B\). Differencing the first equation results in

\[
\Delta y_{1t} = \gamma \Delta y_{2t} + \Delta \epsilon_{1t} = \gamma \epsilon_{2t} + \epsilon_{1t} - \epsilon_{1,t-1}
\]

Thus, both \(y_{1t}\) and \(y_{2t}\) are \(I(1)\) processes, but the linear combination \(\gamma y_{1t} - y y_{2t}\) is stationary. Hence \(y_t = (y_{1t}, y_{2t})'\) is cointegrated with a cointegrating vector \(\beta = (1, -\gamma)'\).

In general, if the vector process \(y_t\) has \(k\) components, then there can be more than one cointegrating vector \(\beta'\). It is assumed that there are \(r\) linearly independent cointegrating vectors with \(r < k\), which make the \(k \times r\) matrix \(\beta\). The rank of matrix \(\beta\) is \(r\), which is called the \textit{cointegration rank} of \(y_t\).

**Common Trends**

This section briefly discusses the implication of cointegration for the moving-average representation. Let \(y_t\) be cointegrated \(CI(1, 1)\), then \(\Delta y_t\) has the Wold representation:

\[
\Delta y_t = \delta + \Psi(B)e_t
\]

where \(e_t\) is iid(0, \(\Sigma\)), \(\Psi(B) = \sum_{j=0}^{\infty} \Psi_j B^j\) with \(\Psi_0 = I_k\), and \(\sum_{j=0}^{\infty} j|\Psi_j| < \infty\).

Assume that \(e_t = 0\) if \(t \leq 0\) and \(y_0\) is a nonrandom initial value. Then the difference equation implies that

\[
y_t = y_0 + \delta t + \Psi(1) \sum_{i=0}^{t} e_i + \Psi^*(B)e_t
\]

where \(\Psi^*(B) = (1 - B)^{-1}(\Psi(B) - \Psi(1))\) and \(\Psi^*(B)\) is absolutely summable.

Assume that the rank of \(\Psi(1)\) is \(m = k - r\). When the process \(y_t\) is cointegrated, there is a cointegrating \(k \times r\) matrix \(\beta\) such that \(\beta'y_t\) is stationary.

Premultiplying \(y_t\) by \(\beta'\) results in

\[
\beta'y_t = \beta'y_0 + \beta'\Psi^*(B)e_t
\]

because \(\beta'\Psi(1) = 0\) and \(\beta'\delta = 0\).

Stock and Watson (1988) showed that the cointegrated process \(y_t\) has a common trends representation derived from the moving-average representation. Since the rank of \(\Psi(1)\) is \(m = k - r\), there is a \(k \times r\) matrix \(H_1\) with rank \(r\) such that \(\Psi(1)H_1 = 0\). Let \(H_2\) be a \(k \times m\) matrix with rank \(m\) such that \(H_2'H_1 = 0\); then \(A = C(1)H_2\) has rank \(m\). The \(H = (H_1, H_2)\) has rank \(k\). By construction of \(H\),

\[
\Psi(1)H = [0, A] = AS_m
\]
where \( S_m = (0_{m \times r}, I_m) \). Since \( \beta' \Psi(1) = 0 \) and \( \beta' \delta = 0 \), \( \delta \) lies in the column space of \( \Psi(1) \) and can be written

\[
\delta = \Psi(1) \tilde{\delta}
\]

where \( \tilde{\delta} \) is a \( k \)-dimensional vector. The common trends representation is written as

\[
y_t = y_0 + \Psi(1) \tilde{\delta} t + \sum_{i=0}^{t} \epsilon_i + \Psi^*(B) \epsilon_t
\]

\[
y_t = y_0 + \Psi(1) H [H^{-1} \tilde{\delta} t + H^{-1} \sum_{i=0}^{t} \epsilon_i] + a_t
\]

\[
y_t = y_0 + \tau_t + a_t
\]

and

\[
\tau_t = \pi + \tau_{t-1} + v_t
\]

where \( a_t = \Psi^*(B) \epsilon_t, \pi = S_m H^{-1} \tilde{\delta}, \tau_t = S_m [H^{-1} \tilde{\delta} t + H^{-1} \sum_{i=0}^{t} \epsilon_i], \) and \( v_t = S_m H^{-1} \epsilon_t \).

Stock and Watson showed that the common trends representation expresses \( y_t \) as a linear combination of \( m \) random walks (\( \tau_t \)) with drift \( \pi \) plus \( I(0) \) components (\( a_t \)).

### Test for the Common Trends

Stock and Watson (1988) proposed statistics for common trends testing. The null hypothesis is that the \( k \)-dimensional time series \( y_t \) has \( m \) common stochastic trends, where \( m \leq k \) and the alternative is that it has \( s \) common trends, where \( s < m \). The test procedure of \( m \) versus \( s \) common stochastic trends is performed based on the first-order serial correlation matrix of \( y_t \). Let \( \beta \perp \) be a \( k \times m \) matrix orthogonal to the cointegrating matrix such that \( \beta \perp \beta = 0 \) and \( \beta \perp \beta \perp = I_m \). Let \( z_t = \beta' y_t \) and \( w_t = \beta' \perp y_t \). Then

\[
w_t = \beta' y_0 + \beta' \perp \tilde{\delta} t + \beta' \perp \Psi(1) \sum_{i=0}^{t} \epsilon_i + \beta' \perp \Psi^*(B) \epsilon_t
\]

Combining the expression of \( z_t \) and \( w_t \),

\[
\begin{bmatrix}
z_t \\
w_t
\end{bmatrix} = \begin{bmatrix}
\beta' y_0 \\
\beta' \perp y_0
\end{bmatrix} + \begin{bmatrix}
0 \\
\beta' \delta
\end{bmatrix} t + \begin{bmatrix}
0 \\
\beta' \perp \Psi(1)
\end{bmatrix} \sum_{i=1}^{t} \epsilon_i + \begin{bmatrix}
\beta' \Psi^*(B) \\
\beta' \perp \Psi^*(B)
\end{bmatrix} \epsilon_t
\]

The Stock-Watson common trends test is performed based on the component \( w_t \) by testing whether \( \beta' \perp \Psi(1) \) has rank \( m \) against rank \( s \).

The following statements perform the Stock-Watson test for common trends:
proc iml;
   sig = 100*i(2);
   phi = {-0.2 0.1, 0.5 0.2, 0.8 0.7, -0.4 0.6};
   call varmasim(y,phi) sigma=sig n=100 initial=0
     seed=45876;
   cn = {'y1' 'y2'};
   create simul2 from y[colname=cn];
   append from y;
quit;

data simul2;
   set simul2;
   date = intnx( 'year', '01jan1900'd, _n_-1 );
   format date year4. ;
run;

proc varmax data=simul2;
   model y1 y2 / p=2 cointtest=(sw);
run;

In Figure 43.71, the first column is the null hypothesis that $y_t$ has $m \leq k$ common trends; the second column is the alternative hypothesis that $y_t$ has $s < m$ common trends; the third column contains the eigenvalues used for the test statistics; the fourth column contains the test statistics using AR($p$) filtering of the data. The table shows the output of the case $p = 2$.

![Figure 43.71 Common Trends Test (COINTTEST=(SW) Option)](image)

The test statistic for testing for 2 versus 1 common trends is more negative ($-35.1$) than the critical value ($-23.0$). Therefore, the test rejects the null hypothesis, which means that the series has a single common trend.

**Vector Error Correction Modeling**

This section discusses the implication of cointegration for the autoregressive representation.

Consider the vector autoregressive process that has Gaussian errors defined by

$$y_t = \sum_{i=1}^{p} \Phi_i y_{t-i} + \epsilon_t$$
or

\[ \Phi(B) y_t = \epsilon_t \]

where the initial values, \( y_{-p+1}, \ldots, y_0 \), are fixed and \( \epsilon_t \sim N(0, \Sigma) \). The AR operator \( \Phi(B) \) can be re-expressed as

\[ \Phi(B) = \Phi^*(B)(1 - B) + \Phi(1)B \]

where

\[ \Phi(1) = I_k - \Phi_1 - \Phi_2 - \cdots - \Phi_p, \quad \Phi^*(B) = I_k - \sum_{i=1}^{p-1} \Phi_i^* B^i, \quad \Phi_i^* = -\sum_{j=i+1}^{p} \Phi_j \]

The vector error correction model (VECM), also called the vector equilibrium correction model, is defined as

\[ \Phi^*(B)(1 - B)y_t = \alpha \beta'y_{t-1} + \epsilon_t \]

or

\[ \Delta y_t = \alpha \beta'y_{t-1} + \sum_{i=1}^{p-1} \Phi_i^* \Delta y_{t-i} + \epsilon_t \]

where \( \alpha \beta' = -\Phi(1) \).

**Granger Representation Theorem**

Engle and Granger (1987) define

\[ \Pi(z) \equiv (1 - z)I_k - \alpha \beta'z - \sum_{i=1}^{p-1} \Phi_i^*(1 - z)z^i \]

and the following assumptions hold:

1. \( |\Pi(z)| = 0 \Rightarrow |z| > 1 \) or \( z = 1 \).
2. The number of unit roots, \( z = 1 \), is exactly \( k - r \).
3. \( \alpha \) and \( \beta \) are \( k \times r \) matrices, and their ranks are both \( r \).

Then \( y_t \) has the representation

\[ y_t = C \sum_{i=1}^{t} \epsilon_i + C^*(B) \epsilon_t + y_0^* \]

where the Granger representation coefficient, \( C \), is

\[ C = \beta_\perp \left[ \alpha'_\perp \Phi(1) \beta_\perp \right]^{-1} \alpha'_\perp \]
where the full-rank $k \times (k - r)$ matrix $\beta_\perp$ is orthogonal to $\beta$ and the full-rank $k \times (k - r)$ matrix $\alpha_\perp$ is orthogonal to $\alpha$. $C^*(B)\epsilon_t = \sum_{j=1}^{\infty} C^*_j \epsilon_{t-j}$ is an $I(0)$ process, and $y^*_0$ depends on the initial values.

The Granger representation coefficient $C$ can be defined only when the $k \times (k - r)$ matrix $\hat{\alpha}_\perp$ is invertible.

One motivation for the VECM($p$) form is to consider the relation $\beta'y_t = \epsilon$ as defining the underlying economic relations. Assume that agents react to the disequilibrium error $\beta'y_t - \epsilon$ through the adjustment coefficient $\alpha$ to restore equilibrium. The cointegrating vector, $\beta$, is sometimes called the long-run parameter.

Consider a vector error correction model that has a deterministic term, $D_t$, which can contain a constant, a linear trend, and seasonal dummy variables. Exogenous variables can also be included in the model. The model has the form

$$
\Delta y_t = \Pi y_{t-1} + \sum_{i=1}^{p-1} \Phi^*_i \Delta y_{t-i} + AD_t + \sum_{i=0}^{s} \Theta^*_i x_{t-i} + \epsilon_t
$$

where $\Pi = \alpha \beta'$.

The alternative vector error correction representation considers the error correction term at lag $t - p$ and is written as

$$
\Delta y_t = \sum_{i=1}^{p-1} \Phi^*_i \Delta y_{t-i} + \Pi \beta'y_{t-p} + AD_t + \sum_{i=0}^{s} \Theta^*_i x_{t-i} + \epsilon_t
$$

If the matrix $\Pi$ has a full rank ($r = k$), all components of $y_t$ are $I(0)$. On the other hand, $y_t$ are stationary in difference if $\text{rank}(P) = 0$. When the rank of the matrix $\Pi$ is $r < k$, there are $k - r$ linear combinations that are nonstationary and $r$ stationary cointegrating relations. Note that the linearly independent vector $z_t = \beta'y_t$ is stationary and this transformation is not unique unless $r = 1$. There does not exist a unique cointegrating matrix $\beta$ because the coefficient matrix $\Pi$ can also be decomposed as

$$
\Pi = \alpha M M^{-1} \beta' = \alpha^* \beta^*'
$$

where $M$ is an $r \times r$ nonsingular matrix.

**Test for Cointegration**

The cointegration rank test determines the linearly independent columns of $\Pi$. Johansen and Juselius proposed the cointegration rank test by using the reduced rank regression (Johansen 1988, 1995b; Johansen and Juselius 1990).

**Different Specifications of Deterministic Trends**

When you construct the VECM($p$) form from the VAR($p$) model, the deterministic terms in the VECM($p$) form can differ from those in the VAR($p$) model. When there are deterministic cointegrated relationships among variables, deterministic terms in the VAR($p$) model are not present in the VECM($p$) form. On the other hand, if there are stochastic cointegrated relationships in the VAR($p$) model, deterministic terms appear in the VECM($p$) form via the error correction term or as an independent term in the VECM($p$) form. There are five different specifications of deterministic trends in the VECM($p$) form.
• **Case 1:** There is no separate drift in the VECM$(p)$ form.

\[ \Delta y_t = \alpha \beta' y_{t-1} + \sum_{i=1}^{p-1} \Phi_i^* \Delta y_{t-i} + \epsilon_t \]

• **Case 2:** There is no separate drift in the VECM$(p)$ form, but a constant enters only via the error correction term.

\[ \Delta y_t = \alpha (\beta', \beta_0) (y_{t-1}', 1) + \sum_{i=1}^{p-1} \Phi_i^* \Delta y_{t-i} + \epsilon_t \]

• **Case 3:** There is a separate drift and no separate linear trend in the VECM$(p)$ form.

\[ \Delta y_t = \alpha \beta' y_{t-1} + \sum_{i=1}^{p-1} \Phi_i^* \Delta y_{t-i} + \delta_0 + \epsilon_t \]

• **Case 4:** There is a separate drift and no separate linear trend in the VECM$(p)$ form, but a linear trend enters only via the error correction term.

\[ \Delta y_t = \alpha (\beta', \beta_1) (y_{t-1}', t) + \sum_{i=1}^{p-1} \Phi_i^* \Delta y_{t-i} + \delta_0 + \delta_1 t + \epsilon_t \]

• **Case 5:** There is a separate linear trend in the VECM$(p)$ form.

\[ \Delta y_t = \alpha \beta' y_{t-1} + \sum_{i=1}^{p-1} \Phi_i^* \Delta y_{t-i} + \delta_0 + \delta_1 t + \epsilon_t \]

First, focus on Cases 1, 3, and 5 to test the null hypothesis that there are at most $r$ cointegrating vectors. Let

\[
\begin{align*}
Z_{0t} &= \Delta y_t \\
Z_{1t} &= y_{t-1} \\
Z_{2t} &= [\Delta y'_{t-1}, \ldots, \Delta y'_{t-p+1}, D_t]' \\
Z_0 &= [Z_{01}, \ldots, Z_{0T}]' \\
Z_1 &= [Z_{11}, \ldots, Z_{1T}]' \\
Z_2 &= [Z_{21}, \ldots, Z_{2T}]'
\end{align*}
\]

where $D_t$ can be empty for Case 1, 1 for Case 3, and $(1, t)$ for Case 5.

In Case 2, $Z_{1t}$ and $Z_{2t}$ are defined as

\[
\begin{align*}
Z_{1t} &= [y'_{t-1}, 1]' \\
Z_{2t} &= [\Delta y'_{t-1}, \ldots, \Delta y'_{t-p+1}]'
\end{align*}
\]
In Case 4, $Z_{1t}$ and $Z_{2t}$ are defined as

$$
Z_{1t} = [y'_{t-1}, t]' \\
Z_{2t} = [\Delta y'_{t-1}, \ldots, \Delta y'_{t-p+1}, 1]' 
$$

Let $\Psi$ be the matrix of parameters consisting of $\Phi_1^*, \ldots, \Phi_{p-1}^*, A, \Theta_0^*, \ldots, \Theta_s^*$, where parameter $A$ corresponds with the regressors $D_t$. Then the VECM($p$) form is rewritten in these variables as

$$
Z_{0t} = \alpha \beta' Z_{1t} + \Psi Z_{2t} + \epsilon_t
$$

The log-likelihood function is given by

$$
\ell = -\frac{kT}{2} \log 2\pi - \frac{T}{2} \log |\Sigma| - \frac{1}{2} \sum_{t=1}^{T} (Z_{0t} - \alpha \beta' Z_{1t} - \Psi Z_{2t})' \Sigma^{-1} (Z_{0t} - \alpha \beta' Z_{1t} - \Psi Z_{2t})
$$

The residuals, $R_{0t}$ and $R_{1t}$, are obtained by regressing $Z_{0t}$ and $Z_{1t}$ on $Z_{2t}$, respectively. The regression equation of residuals is

$$
R_{0t} = \alpha \beta' R_{1t} + \hat{\epsilon}_t
$$

The crossproducts matrices are computed

$$
S_{ij} = \frac{1}{T} \sum_{t=1}^{T} R_{it} R'_{jt}, \quad i, j = 0, 1
$$

Then the maximum likelihood estimator for $\beta$ is obtained from the eigenvectors that correspond to the $r$ largest eigenvalues of the following equation:

$$
|\lambda S_{11} - S_{10} S^{-1}_{00} S_{01}| = 0
$$

The eigenvalues of the preceding equation are squared canonical correlations between $R_{0t}$ and $R_{1t}$, and the eigenvectors that correspond to the $r$ largest eigenvalues are the $r$ linear combinations of $y_{t-1}$, which have the largest squared partial correlations with the stationary process $\Delta y_t$ after correcting for lags and deterministic terms. Such an analysis calls for a reduced rank regression of $\Delta y_t$ on $y_{t-1}$ corrected for $(\Delta y_{t-1}, \ldots, \Delta y_{t-p+1}, D_t)$, as discussed by Anderson (1951). Johansen (1988) suggests two test statistics to test the null hypothesis that there are at most $r$ cointegrating vectors

$$
H_0 : \lambda_i = 0 \text{ for } i = r + 1, \ldots, k
$$
Trace Test

The trace statistic for testing the null hypothesis that there are at most \( r \) cointegrating vectors is as follows:

\[
\lambda_{\text{trace}} = -T \sum_{i=r+1}^{k} \log(1 - \lambda_i)
\]

The asymptotic distribution of this statistic is given by

\[
\text{tr}\left\{ \int_0^1 (dW)\hat{W}' \left( \int_0^1 \hat{W}\hat{W}'\,dr \right)^{-1} \int_0^1 \hat{W}(dW)' \right\}
\]

where \( \text{tr}(A) \) is the trace of a matrix \( A \), \( W \) is the \( k - r \) dimensional Brownian motion, and \( \hat{W} \) is the Brownian motion itself, or the demeaned or detrended Brownian motion according to the different specifications of deterministic trends in the vector error correction model.

Maximum Eigenvalue Test

The maximum eigenvalue statistic for testing the null hypothesis that there are at most \( r \) cointegrating vectors is as follows:

\[
\lambda_{\text{max}} = -T \log(1 - \lambda_{r+1})
\]

The asymptotic distribution of this statistic is given by

\[
\max\{ \int_0^1 (dW)\hat{W}' \left( \int_0^1 \hat{W}\hat{W}'\,dr \right)^{-1} \int_0^1 \hat{W}(dW)' \}
\]

where \( \max(A) \) is the maximum eigenvalue of a matrix \( A \). Osterwald-Lenum (1992) provided detailed tables of the critical values of these statistics.

The following statements use the JOHANSEN option to compute the Johansen cointegration rank trace test of integrated order 1:

```plaintext
proc varmax data=simul2;
   model y1 y2 / p=2 cointest=(johansen=(normalize=y1));
run;
```

Figure 43.72 shows the output based on the model specified in the MODEL statement. An intercept term is assumed. In the “Cointegration Rank Test Using Trace” table, the column Drift in ECM indicates that there is no separate drift in the error correction model, and the column Drift in Process indicates that the process has a constant drift before differencing. The “Cointegration Rank Test Using Trace” table shows the trace statistics and \( p \)-values based on Case 3, and the “Cointegration Rank Test Using Trace under Restriction” table shows the trace statistics and \( p \)-values based on Case 2. For a specified significance level, such as 5\%, the output indicates that the null hypothesis that the series are not cointegrated (H0: Rank = 0) can be rejected, because the \( p \)-values for both Case 2 and Case 3 are less than 0.05. The output also shows that the null hypothesis that the series are cointegrated with rank 1 (H0: Rank = 1) cannot be rejected for either Case 2 or Case 3, because the \( p \)-values for these tests are both greater than 0.05.
Figure 43.72  Cointegration Rank Test (COINTTEST=(JOHANSEN=) Option)

The VARMAX Procedure

<table>
<thead>
<tr>
<th>Cointegration Rank Test Using Trace</th>
<th>H0: Rank</th>
<th>H1: Rank</th>
<th>Eigenvalue</th>
<th>Trace</th>
<th>Pr &gt; Trace</th>
<th>Drift in ECM</th>
<th>Drift in Process</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.4644</td>
<td>61.7522</td>
<td>&lt;.0001</td>
<td>Constant</td>
<td>Linear</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0.0056</td>
<td>0.5552</td>
<td>0.4559</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Cointegration Rank Test Using Trace Under Restriction</th>
<th>H0: Rank</th>
<th>H1: Rank</th>
<th>Eigenvalue</th>
<th>Trace</th>
<th>Pr &gt; Trace</th>
<th>Drift in ECM</th>
<th>Drift in Process</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.5209</td>
<td>76.3788</td>
<td>&lt;.0001</td>
<td>Constant</td>
<td>Constant</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0.0426</td>
<td>4.2680</td>
<td>0.3741</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Figure 43.73 shows which result, either Case 2 (the hypothesis H0) or Case 3 (the hypothesis H1), is appropriate depending on the significance level. Since the cointegration rank is chosen to be 1 by the result in Figure 43.72, look at the last row that corresponds to rank=1. Since the p-value is 0.054, the Case 2 cannot be rejected at the significance level 5%, but it can be rejected at the significance level 10%. For modeling of the two Case 2 and Case 3, see Figure 43.76 and Figure 43.77.

Figure 43.73  Cointegration Rank Test, Continued

<table>
<thead>
<tr>
<th>Hypothesis of the Restriction</th>
<th>Drift in ECM</th>
<th>Drift in Process</th>
</tr>
</thead>
<tbody>
<tr>
<td>H0(Case 2)</td>
<td>Constant</td>
<td>Constant</td>
</tr>
<tr>
<td>H1(Case 3)</td>
<td>Constant</td>
<td>Linear</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Hypothesis Test of the Restriction</th>
<th>Restricted Eigenvalue</th>
<th>DF</th>
<th>Chi-Square</th>
<th>Pr &gt; ChiSq</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0.4644</td>
<td>2</td>
<td>14.63</td>
<td>0.0007</td>
</tr>
<tr>
<td>1</td>
<td>0.0056</td>
<td>1</td>
<td>3.71</td>
<td>0.0540</td>
</tr>
</tbody>
</table>

Figure 43.74 shows the estimates of long-run parameter (Beta) and adjustment coefficients (Alpha) based on Case 3.

Figure 43.74  Cointegration Rank Test, Continued

<table>
<thead>
<tr>
<th>Beta</th>
<th>Variable</th>
<th>1</th>
<th>2</th>
</tr>
</thead>
<tbody>
<tr>
<td>y1</td>
<td>1.00000</td>
<td>1.00000</td>
<td></td>
</tr>
<tr>
<td>y2</td>
<td>-2.04869</td>
<td>-0.02854</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Alpha</th>
<th>Variable</th>
<th>1</th>
<th>2</th>
</tr>
</thead>
<tbody>
<tr>
<td>y1</td>
<td>-0.46421</td>
<td>-0.00502</td>
<td></td>
</tr>
<tr>
<td>y2</td>
<td>0.17535</td>
<td>-0.01275</td>
<td></td>
</tr>
</tbody>
</table>
Using the NORMALIZE= option, the first row of the “Beta” table has 1. Considering that the cointegration rank is 1, the long-run relationship of the series is

$$\begin{align*}
\beta' y_t &= \begin{bmatrix} 1 & -2.04869 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} \\
&= y_{1t} - 2.04869y_{2t} \\
y_{1t} &= 2.04869y_{2t}
\end{align*}$$

Figure 43.75 shows the estimates of long-run parameter (Beta) and adjustment coefficients (Alpha) based on Case 2.

**Figure 43.75 Cointegration Rank Test, Continued**

<table>
<thead>
<tr>
<th>Beta Under Restriction</th>
<th>1</th>
<th>2</th>
</tr>
</thead>
<tbody>
<tr>
<td>y1</td>
<td>1.00000</td>
<td>1.00000</td>
</tr>
<tr>
<td>y2</td>
<td>-2.04366</td>
<td>-2.75773</td>
</tr>
<tr>
<td>1</td>
<td>6.75919</td>
<td>101.37051</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Alpha Under Restriction</th>
<th>1</th>
<th>2</th>
</tr>
</thead>
<tbody>
<tr>
<td>y1</td>
<td>-0.48015</td>
<td>0.01091</td>
</tr>
<tr>
<td>y2</td>
<td>0.12538</td>
<td>0.03722</td>
</tr>
</tbody>
</table>

Considering that the cointegration rank is 1, the long-run relationship of the series is

$$\begin{align*}
\beta' y_t &= \begin{bmatrix} 1 & -2.04366 & 6.75919 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ 1 \end{bmatrix} \\
&= y_{1t} - 2.04366y_{2t} + 6.75919 \\
y_{1t} &= 2.04366y_{2t} - 6.75919
\end{align*}$$

**Estimation of Vector Error Correction Model**

The preceding log-likelihood function is maximized for

$$\begin{align*}
\hat{\beta} &= S_{11}^{-1/2} [v_1, \ldots, v_r] \\
\hat{\alpha} &= S_{01} \hat{\beta}' (\hat{\beta} S_{11} \hat{\beta})^{-1} \\
\hat{\Gamma} &= \hat{\alpha} \hat{\beta}' \\
\hat{\psi}' &= (Z_2'Z_2)^{-1}Z_2'(Z_0 - Z_1 \hat{\Gamma}') \\
\hat{\Sigma} &= (Z_0 - Z_2 \hat{\psi}' - Z_1 \hat{\Gamma}')(Z_0 - Z_2 \hat{\psi}' - Z_1 \hat{\Gamma}') / T
\end{align*}$$

The estimators of the orthogonal complements of $\alpha$ and $\beta$ are

$$\hat{\beta}_{\perp} = S_{11} [v_{r+1}, \ldots, v_k]$$
and
\[ \hat{\alpha}_\perp = S_{01}^{-1}S_{01}[v_{r+1}, \ldots, v_k] \]

Let \( \vartheta \) denote the parameter vector \( (\text{vec}(\alpha), \Psi)', \text{vech}(\Sigma)' \)' of the \( \tilde{\alpha} \). The covariance of parameter estimates \( \hat{\vartheta} \) is obtained as the inverse of the negative Hessian matrix \( H \equiv \frac{\partial^2 L}{\partial \vartheta \partial \vartheta'} \). Because \( \tilde{\Pi} = \hat{\alpha} \hat{\beta}' \), the variance of \( \tilde{\Pi} \) and the covariance between \( \tilde{\Pi} \) and \( \hat{\vartheta} \) are calculated as follows:

\[
\begin{align*}
\text{cov}(\text{vec}(\tilde{\Pi}), \text{vec}(\tilde{\Pi})) &= (\hat{\beta} \otimes I_k)\text{cov}(\text{vec}(\hat{\alpha}), \text{vec}(\hat{\alpha}))(\hat{\beta} \otimes I_k)' \\
\text{cov}(\text{vec}(\tilde{\Pi}), \hat{\vartheta}) &= (\hat{\beta} \otimes I_k)\text{cov}(\text{vec}(\hat{\alpha}), \hat{\vartheta})
\end{align*}
\]

For Case 2 (Case 4), because the coefficient vector \( \hat{\delta}_0 (\hat{\delta}_1) \) for the constant term (the linear trend term) is the product of \( \hat{\alpha} \) and \( \hat{\beta}_0 (\hat{\beta}_1) \), the variance of \( \hat{\delta}_0 (\hat{\delta}_1) \) and the covariance between \( \hat{\delta}_0 (\hat{\delta}_1) \) and \( \hat{\vartheta} \) are calculated as follows:

\[
\begin{align*}
\text{cov}(\hat{\delta}_i, \hat{\delta}_i) &= (\hat{\beta}_i' \otimes I_k)\text{cov}(\text{vec}(\hat{\alpha}), \text{vec}(\hat{\alpha}))(\hat{\beta}_i' \otimes I_k)', \; i = 0 \text{ or } 1 \\
\text{cov}(\hat{\delta}_i, \hat{\vartheta}) &= (\hat{\beta}_i' \otimes I_k)\text{cov}(\text{vec}(\hat{\alpha}), \hat{\vartheta}), \; i = 0 \text{ or } 1
\end{align*}
\]

The following statements are examples of fitting the five different cases of the vector error correction models mentioned in the previous section.

For fitting Case 1,

```
model y1 y2 / p=2 noint;
  cointeg rank=1 normalize=y1;
```

For fitting Case 2,

```
model y1 y2 / p=2;
  cointeg rank=1 normalize=y1 ectrend;
```

For fitting Case 3,

```
model y1 y2 / p=2;
  cointeg rank=1 normalize=y1;
```

For fitting Case 4,

```
model y1 y2 / p=2 trend=linear;
  cointeg rank=1 normalize=y1 ectrend;
```

For fitting Case 5,

```
model y1 y2 / p=2 trend=linear;
  cointeg rank=1 normalize=y1;
```

In the previous example, the output from the COINTTEST=(JOHANSEN) option shown in Figure 43.73 indicates that you can fit the model by using either Case 2 or Case 3 because the test of the restriction was not significant at the 0.05 level, but was significant at the 0.10 level. Following both models are fit to show the differences in the displayed output. Figure 43.76 is for Case 2, and Figure 43.77 is for Case 3.
For Case 2,

```r
proc varmax data=simul2;
   model y1 y2 / p=2 print=(estimates);
   cointeg rank=1 normalize=y1 ectrend;
run;
```

**Figure 43.76** Parameter Estimation with the ECTREND Option

### The VARMAX Procedure

<table>
<thead>
<tr>
<th>Parameter Alpha * Beta Estimates</th>
</tr>
</thead>
<tbody>
<tr>
<td>Variable</td>
</tr>
<tr>
<td>----------</td>
</tr>
<tr>
<td>y1</td>
</tr>
<tr>
<td>y2</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>AR Coefficients of Differenced Lag</th>
</tr>
</thead>
<tbody>
<tr>
<td>DIF Lag</td>
</tr>
<tr>
<td>---------</td>
</tr>
<tr>
<td>1</td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Model Parameter Estimates</th>
</tr>
</thead>
<tbody>
<tr>
<td>Equation Parameter Estimate Standard Error t Value Pr &gt;</td>
</tr>
<tr>
<td>---------------------------</td>
</tr>
<tr>
<td>D_y1</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>D_y2</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>

**Figure 43.76** can be reported as follows:

\[
\Delta y_t = \begin{bmatrix} -0.48015 & 0.98126 & -3.24543 \\ 0.12538 & -0.25624 & 0.84748 \end{bmatrix} \begin{bmatrix} y_{1,t-1} \\ y_{2,t-1} \end{bmatrix} + \begin{bmatrix} -0.72759 & -0.77463 \\ 0.38982 & -0.55173 \end{bmatrix} \Delta y_{t-1} + \epsilon_t
\]

The keyword “EC” in the “Model Parameter Estimates” table means that the ECTREND option is used for fitting the model.
For fitting Case 3,

```sas
proc varmax data=simul2;
    model y1 y2 / p=2 print=(estimates);
    cointeg rank=1 normalize=y1;
run;
```

**Figure 43.77** Parameter Estimation without the ECTREND Option

### The VARMAX Procedure

<table>
<thead>
<tr>
<th>Parameter Alpha * Beta' Estimates</th>
<th>y1</th>
<th>y2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Variable</td>
<td></td>
<td></td>
</tr>
<tr>
<td>y1</td>
<td>-0.46421</td>
<td>0.95103</td>
</tr>
<tr>
<td>y2</td>
<td>0.17535</td>
<td>-0.35923</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>AR Coefficients of Differenced Lag</th>
<th>DIF Lag</th>
<th>Variable</th>
<th>y1</th>
<th>y2</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>y1</td>
<td>-0.74052</td>
<td>-0.76305</td>
<td></td>
</tr>
<tr>
<td></td>
<td>y2</td>
<td>0.34820</td>
<td>-0.51194</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Model Parameter Estimates</th>
<th>Equation</th>
<th>Parameter</th>
<th>Estimate</th>
<th>Standard Error</th>
<th>t Value</th>
<th>Pr &gt;</th>
<th>Variable</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>CONST1</td>
<td>-2.60825</td>
<td>1.32398</td>
<td>-1.97</td>
<td>0.0518</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>AR1_1_1</td>
<td>-0.46421</td>
<td>0.05474</td>
<td>-8.48</td>
<td>&lt;0.001</td>
<td>y1(t-1)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>AR1_1_2</td>
<td>0.95103</td>
<td>0.11215</td>
<td>8.48</td>
<td>&lt;0.001</td>
<td>y2(t-2)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>AR2_1_1</td>
<td>-0.74052</td>
<td>0.05060</td>
<td>-14.63</td>
<td>&lt;0.001</td>
<td>D_y1(t-1)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>AR2_1_2</td>
<td>-0.76305</td>
<td>0.05352</td>
<td>-14.26</td>
<td>&lt;0.001</td>
<td>D_y2(t-1)</td>
</tr>
<tr>
<td>D_y2</td>
<td>CONST2</td>
<td>3.43005</td>
<td>1.39587</td>
<td>2.46</td>
<td>0.0159</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>AR1_2_1</td>
<td>0.17535</td>
<td>0.05771</td>
<td>3.04</td>
<td>0.0031</td>
<td>y1(t-1)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>AR1_2_2</td>
<td>-0.35923</td>
<td>0.11824</td>
<td>-3.04</td>
<td>0.0031</td>
<td>y2(t-1)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>AR2_2_1</td>
<td>0.34820</td>
<td>0.05335</td>
<td>6.53</td>
<td>&lt;0.001</td>
<td>D_y1(t-1)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>AR2_2_2</td>
<td>-0.51194</td>
<td>0.05643</td>
<td>-9.07</td>
<td>&lt;0.001</td>
<td>D_y2(t-1)</td>
<td></td>
</tr>
</tbody>
</table>

**Figure 43.77** can be reported as follows:

\[
\Delta y_t = \begin{bmatrix}
-0.46421 & 0.95103 \\
0.17535 & -0.35923 \\
\end{bmatrix} y_{t-1} + \begin{bmatrix}
-0.74052 & -0.76305 \\
0.34820 & -0.51194 \\
\end{bmatrix} \Delta y_{t-1} + \begin{bmatrix}
-2.60825 \\
3.43005 \\
\end{bmatrix} + \epsilon_t
\]

**A Test for the Long-Run Relations**

Consider the example with the variables \(m_t\) log real money, \(y_t\) log real income, \(i_t^d\) deposit interest rate, and \(i_t^b\) bond interest rate. It seems a natural hypothesis that in the long-run relation, money and income have equal coefficients with opposite signs. This can be formulated as the hypothesis that the cointegrated relation contains only \(m_t\) and \(y_t\) through \(m_t - y_t\). For the analysis, you can express these restrictions in the
parameterization of \( H \) such that \( \beta = H \phi \), where \( H \) is a known \( k \times s \) matrix and \( \psi \) is the \( s \times r (r \leq s < k) \) parameter matrix to be estimated. For this example, \( H \) is given by

\[
H = \begin{bmatrix}
1 & 0 & 0 \\
-1 & 0 & 0 \\
0 & 1 & 0 \\
0 & 0 & 1 
\end{bmatrix}
\]

**Restriction** \( H_0: \beta = H\phi \)

When the linear restriction \( \beta = H\phi \) is given, it implies that the same restrictions are imposed on all cointegrating vectors. You obtain the maximum likelihood estimator of \( \beta \) by reduced rank regression of \( \Delta y_t \) on \( H y_{t-1} \) corrected for \( (\Delta y_{t-1}, \ldots, \Delta y_{t-p+1}, D_t) \), solving the following equation,

\[
|\rho H' S_{11} H - H' S_{10} S_{00}^{-1} S_{01} H| = 0
\]

for the eigenvalues \( 1 > \rho_1 > \cdots > \rho_s > 0 \) and eigenvectors \( (v_1, \ldots, v_s) \), \( S_{ij} \) given in the preceding section. Then choose \( \hat{\phi} = (v_1, \ldots, v_r) \) that corresponds to the \( r \) largest eigenvalues, and the \( \hat{\beta} \) is \( H\hat{\phi} \).

The test statistic for \( H_0: \beta = H\phi \) is given by

\[
T \sum_{i=1}^{r} \log\{(1 - \rho_i)/(1 - \lambda_i)\} \xrightarrow{d} \chi^2_{r(k-s)}
\]

If the series has no deterministic trend, the constant term should be restricted by \( \alpha' \delta_0 = 0 \) as in Case 2. Then \( H \) is given by

\[
H = \begin{bmatrix}
1 & 0 & 0 & 0 \\
-1 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 \\
0 & 0 & 1 & 0 \\
0 & 0 & 0 & 1 
\end{bmatrix}
\]

The following statements test that \( 2 \beta_1 + \beta_2 = 0 \):

```plaintext
proc varmax data=simul2;
   model y1 y2 / p=2;
   cointeg rank=1 h=(1,-2) normalize=y1;
run;
```

**Figure 43.78** shows the results of testing \( H_0: 2\beta_1 + \beta_2 = 0 \). The input \( H \) matrix is \( H = (1 - 2)' \). The adjustment coefficient is reestimated under the restriction, and the test indicates that you cannot reject the null hypothesis.
Test for the Weak Exogeneity and Restrictions of Alpha

Consider a vector error correction model:

$$
\Delta y_t = \alpha \beta' y_{t-1} + \sum_{i=1}^{p-1} \Phi_i^* \Delta y_{t-i} + AD_t + \epsilon_t
$$

Divide the process $y_t$ into $(y_{1t}', y_{2t}')'$ with dimension $k_1$ and $k_2$ and the $\Sigma$ into

$$
\Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}
$$

Similarly, the parameters can be decomposed as follows:

$$
\alpha = \begin{bmatrix} \alpha_1 \\ \alpha_2 \end{bmatrix}, \Phi_i = \begin{bmatrix} \Phi_{1i}^* \\ \Phi_{2i}^* \end{bmatrix}, A = \begin{bmatrix} A_1 \\ A_2 \end{bmatrix}
$$

Then the VECM($p$) form can be rewritten by using the decomposed parameters and processes:

$$
\begin{bmatrix} \Delta y_{1t} \\ \Delta y_{2t} \end{bmatrix} = \begin{bmatrix} \alpha_1 \\ \alpha_2 \end{bmatrix} \beta' y_{t-1} + \sum_{i=1}^{p-1} \begin{bmatrix} \Phi_{1i}^* \\ \Phi_{2i}^* \end{bmatrix} \Delta y_{t-i} + \begin{bmatrix} A_1 \\ A_2 \end{bmatrix} D_t + \begin{bmatrix} \epsilon_{1t} \\ \epsilon_{2t} \end{bmatrix}
$$

The conditional model for $y_{1t}$ given $y_{2t}$ is

$$
\Delta y_{1t} = \omega \Delta y_{2t} + (\alpha_1 - \omega \alpha_2) \beta' y_{t-1} + \sum_{i=1}^{p-1} (\Phi_{1i}^* - \omega \Phi_{2i}^*) \Delta y_{t-i} + (A_1 - \omega A_2) D_t + \epsilon_{1t} - \omega \epsilon_{2t}
$$
and the marginal model of $y_{2t}$ is

$$\Delta y_{2t} = \alpha_2 \beta' y_{t-1} + \sum_{i=1}^{p-1} \Phi_{22}^* \Delta y_{t-i} + A_2 D_t + \epsilon_{2t}$$

where $\omega = \Sigma_{12} \Sigma_{22}^{-1}$.

The test of weak exogeneity of $y_{2t}$ for the parameters $(\alpha_1, \beta)$ determines whether $\alpha_2 = 0$. Weak exogeneity means that there is no information about $\beta$ in the marginal model or that the variables $y_{2t}$ do not react to a disequilibrium.

**Restriction** $H_0: \alpha = J \psi$

Consider the null hypothesis $H_0: \alpha = J \psi$, where $J$ is a $k \times m$ matrix with $r \leq m < k$.

From the previous residual regression equation

$$R_{0t} = \alpha \beta' R_{1t} + \hat{\epsilon}_t = J \psi \beta' R_{1t} + \hat{\epsilon}_t$$

you can obtain

$$\bar{J}' R_{0t} = \psi \beta' R_{1t} + \bar{J}' \hat{\epsilon}_t$$

$$J_{\perp}' R_{0t} = J_{\perp}' \hat{\epsilon}_t$$

where $\bar{J} = J(J'J)^{-1}$ and $J_{\perp}$ is orthogonal to $J$ such that $J_{\perp}' J = 0$.

Define

$$\Sigma_{JJ_{\perp}} = \bar{J}' \Sigma J_{\perp} \text{ and } \Sigma_{J_{\perp} J_{\perp}} = J_{\perp}' \Sigma J_{\perp}$$

and let $\omega = \Sigma_{JJ_{\perp}} \Sigma_{J_{\perp} J_{\perp}}^{-1}$. Then $\bar{J}' R_{0t}$ can be written as

$$\bar{J}' R_{0t} = \psi \beta' R_{1t} + \omega J_{\perp}' R_{0t} + \bar{J}' \hat{\epsilon}_t - \omega J_{\perp}' \hat{\epsilon}_t$$

Using the marginal distribution of $J_{\perp}' R_{0t}$ and the conditional distribution of $\bar{J}' R_{0t}$, the new residuals are computed as

$$\bar{R}_{Jt} = \bar{J}' R_{0t} - S_{JJ_{\perp}} S_{J_{\perp} J_{\perp}}^{-1} J_{\perp}' R_{0t}$$

$$\bar{R}_{1t} = R_{1t} - S_{1J_{\perp}} S_{J_{\perp} J_{\perp}}^{-1} J_{\perp}' R_{0t}$$

where

$$S_{JJ_{\perp}} = \bar{J}' S_{00} J_{\perp} \text{, } S_{J_{\perp} J_{\perp}} = J_{\perp}' S_{00} J_{\perp} \text{ and } S_{J_{\perp} 1} = J_{\perp}' S_{01}$$

In terms of $\bar{R}_{Jt}$ and $\bar{R}_{1t}$, the MLE of $\beta$ is computed by using the reduced rank regression. Let

$$S_{ij,J_{\perp}} = \frac{1}{T} \sum_{t=1}^{T} \bar{R}_{it} \bar{R}_{jt}', \text{ for } i, j = 1, J$$
Under the null hypothesis $H_0: \alpha = J \psi$, the MLE $\tilde{\beta}$ is computed by solving the equation

$$|\rho S_{11,J \perp} - S_{1J,J \perp} S_{J,J \perp}^{-1} S_{J1,J \perp}| = 0$$

Then $\tilde{\beta} = (v_1, \ldots, v_r)$, where the eigenvectors correspond to the $r$ largest eigenvalues and are normalized such that $\tilde{\beta}' S_{11,J \perp} \tilde{\beta} = I_r; \tilde{\alpha} = J S_{J1,J \perp} \tilde{\beta}$. The likelihood ratio test for $H_0: \alpha = J \psi$ is

$$T \sum_{i=1}^{r} \log \left\{ (1 - \rho_i)/(1 - \lambda_i) \right\} \xrightarrow{d} \chi^2_{r(k-m)}$$

For more information, see Theorem 6.1 in Johansen and Juselius (1990).

The test of weak exogeneity of $y_{2t}$ is a special case of the test $\alpha = J \psi$, considering $J = (I_{k1}, 0)'$. Consider the previous example with four variables ($m_t, y_t, i^b_t, i^d_t$). If $r = 1$, you formulate the weak exogeneity of $(y_t, i^b_t, i^d_t)$ for $m_t$ as $J = [0, I_3]'$ and the weak exogeneity of $i^d_t$ for $(m_t, y_t, i^b_t)$ as $J = [I_3, 0]'$.

The following statements test the weak exogeneity of other variables, assuming $r = 1$:

```r
proc varmax data=simul2;
   model y1 y2 / p=2;
   cointeg rank=1 exogeneity normalize=y1;
run;
```

Figure 43.79 shows that each variable is not the weak exogeneity of other variable.

**Figure 43.79** Testing of Weak Exogeneity (EXOGENEITY Option)

### The VARMAX Procedure

<table>
<thead>
<tr>
<th>Variable</th>
<th>Weak Exogeneity</th>
<th>DF</th>
<th>Chi-Square</th>
<th>Pr &gt; ChiSq</th>
</tr>
</thead>
<tbody>
<tr>
<td>y1</td>
<td></td>
<td>1</td>
<td>53.46</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>y2</td>
<td></td>
<td>1</td>
<td>8.76</td>
<td>0.0031</td>
</tr>
</tbody>
</table>

### General Tests and Restrictions on Parameters

The previous sections discuss some special forms of tests on $\beta$ and $\alpha$, namely the long-run relations that are expressed in the form $H_0: \beta = H_\phi$, the weak exogeneity test, and the null hypotheses on $\alpha$ in the form $H_0: \alpha = J \psi$. In fact, with the help of the RESRICT and BOUND statements, you can estimate the models that have linear restrictions on any parameters to be estimated, which means that you can implement the likelihood ratio (LR) test for any linear relationship between the parameters.

The restricted error correction model must be estimated through numerical optimization. You might need to use the NLOPTIONS statement to try different options for the optimizer and the INITIAL statement to try different starting points. This is essentially important because the $\alpha$ and $\beta$ are usually not identifiable.

You can also use the TEST statement to apply the Wald test for any linear relationships between parameters that are not long-run. Even more, you can test the constraints on $\Pi(= \alpha \beta')$ and $\delta_0(= \alpha \beta_0)$ in Case 2 or $\delta_1(= \alpha \beta_1)$ in Case 4 when the constant term or linear trend is restricted to the error correction term.
For more information and examples, see the section “Example 43.3: Analysis of Restricted Cointegrated Systems” on page 3189.

**Forecasting of the VECM**

Consider the cointegrated moving-average representation of the differenced process of $y_t$

$$\Delta y_t = \delta + \Psi(B)\epsilon_t$$

Assume that $y_0 = 0$. The linear process $y_t$ can be written as

$$y_t = \delta t + \sum_{i=1}^{t} \sum_{j=0}^{t-i} \Psi_j \epsilon_i$$

Therefore, for any $l > 0$,

$$y_{t+l} = \delta(t + l) + \sum_{i=1}^{t} \sum_{j=0}^{t+l-i} \Psi_j \epsilon_{t+i}$$

The $l$-step-ahead forecast is derived from the preceding equation:

$$y_{t+l | t} = (t + l) + \sum_{i=1}^{t} \sum_{j=0}^{t+l-i} \Psi_j \epsilon_i$$

Note that

$$\lim_{l \to \infty} \beta' y_{t+l | t} = 0$$

since $\lim_{l \to \infty} \sum_{j=0}^{t+l-i} \Psi_j = \Psi(1)$ and $\beta' \Psi(1) = 0$. The long-run forecast of the cointegrated system shows that the cointegrated relationship holds, although there might exist some deviations from the equilibrium status in the short-run. The covariance matrix of the predict error $\epsilon_{t+l | t} = y_{t+l} - y_{t+l | t}$ is

$$\Sigma(l) = \sum_{i=1}^{l} [(\sum_{j=0}^{l-i} \Psi_j) \Sigma(\sum_{j=0}^{l-i} \Psi_j')]$$

When the linear process is represented as a VECM($p$) model, you can obtain

$$\Delta y_t = \Pi y_{t-1} + \sum_{j=1}^{p-1} \Phi_j^* \Delta y_{t-j} + \delta + \epsilon_t$$

The transition equation is defined as

$$z_t = Fz_{t-1} + e_t$$
where $z_t = (y'_{t-1}, \Delta y'_t, \Delta y'_{t-1}, \ldots, \Delta y'_{t-p+2})'$ is a state vector and the transition matrix is

$$
F = \begin{bmatrix}
I_k & I_k & 0 & \cdots & 0 \\
\Pi & (\Pi + \Phi_1^*) & \Phi_2^* & \cdots & \Phi_{p-1}^* \\
0 & I_k & 0 & \cdots & 0 \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
0 & 0 & \cdots & I_k & 0
\end{bmatrix}
$$

where $0$ is a $k \times k$ zero matrix. The observation equation can be written

$$
y_t = \delta_t + Hz_t
$$

where $H = [I_k, I_k, 0, \ldots, 0]$.

The $l$-step-ahead forecast is computed as

$$
y_{t+l|t} = \delta(t + l) + HF^l z_t
$$

### Cointegration with Exogenous Variables

The error correction model with exogenous variables can be written as follows:

$$
\Delta y_t = \alpha \beta' y_{t-1} + \sum_{i=1}^{p-1} \Phi_i^* \Delta y_{t-i} + AD_t + \sum_{i=0}^{s} \Theta_i^* x_{t-i} + \epsilon_t
$$

The following statements demonstrate how to fit VECMX($p, s$), where $p = 2$ and $s = 1$ from the P=2 and XLAG=1 options:

```plaintext
proc varmax data=simul3;
  model y1 y2 = x1 / p=2 xlag=1;
  cointeg rank=1;
run;
```

The following statements demonstrate how to BVECXML(2,1):

```plaintext
proc varmax data=simul3;
  model y1 y2 = x1 / p=2 xlag=1
  prior=(lambda=0.9 theta=0.1);
  cointeg rank=1;
run;
```

### I(2) Model

The VARX($p,s$) model can be written in the error correction form:

$$
\Delta y_t = \alpha \beta' y_{t-1} + \sum_{i=1}^{p-1} \Phi_i^* \Delta y_{t-i} + AD_t + \sum_{i=0}^{s} \Theta_i^* x_{t-i} + \epsilon_t
$$
Let $\Phi^* = I_k - \sum_{i=1}^{p-1} \Phi_i^*$. If $\alpha$ and $\beta$ have full-rank $r$, and $\text{rank}(\alpha^\prime \Phi^* \beta) = k - r$, then $y_t$ is an $I(1)$ process.

If the condition $\text{rank}(\alpha^\prime \Phi^* \beta) = k - r$ fails and $\alpha^\prime \Phi^* \beta$ has reduced-rank $\alpha^\prime_1 \Phi^* \beta_\perp = \xi \eta'$ where $\xi$ and $\eta$ are $(k - r) \times s$ matrices with $s \leq k - r$, then $\alpha_\perp$ and $\beta_\perp$ are defined as $k \times (k - r)$ matrices of full rank such that $\alpha^\prime \alpha_\perp = 0$ and $\beta^\prime \beta_\perp = 0$.

If $\xi$ and $\eta$ have full-rank $s$, then the process $y_t$ is $I(2)$, which has the implication of $I(2)$ model for the moving-average representation.

$$y_t = B_0 + B_1 t + C_2 \sum_{j=1}^{t} \sum_{i=1}^{j} \epsilon_i + C_1 \sum_{i=1}^{t} \epsilon_i + C_0(B) \epsilon_t$$

The matrices $C_1$, $C_2$, and $C_0(B)$ are determined by the cointegration properties of the process, and $B_0$ and $B_1$ are determined by the initial values. For more information, see Johansen (1995b).

The implication of the $I(2)$ model for the autoregressive representation is given by

$$\Delta^2 y_t = \Pi y_{t-1} - \Phi^* \Delta y_{t-1} + \sum_{i=1}^{p-2} \Psi_i \Delta^2 y_{t-i} + AD_t + \sum_{i=0}^{s} \Theta_i^* x_{t-i} + \epsilon_t$$

where $\Psi_i = -\sum_{j=i+1}^{p-1} \Phi_i^*$ and $\Phi^* = I_k - \sum_{i=1}^{p-1} \Phi_i^*$.

**Test for $I(2)$**

The $I(2)$ cointegrated model is given by the following parameter restrictions:

$$H_{r,s}: \Pi = \alpha^\prime \beta' \text{ and } \alpha^\prime_1 \Phi^* \beta_\perp = \xi \eta'$$

where $\xi$ and $\eta$ are $(k - r) \times s$ matrices with $0 \leq s \leq k - r$. Let $H_0^0$ represent the $I(1)$ model where $\alpha$ and $\beta$ have full-rank $r$, let $H_{r,s}^0$ represent the $I(2)$ model where $\xi$ and $\eta$ have full-rank $s$, and let $H_{r,s}$ represent the $I(2)$ model where $\xi$ and $\eta$ have rank $\leq s$. Table 43.6 shows the relation between the $I(1)$ models and the $I(2)$ models.

<table>
<thead>
<tr>
<th>Table 43.6</th>
<th>Relation between the $I(1)$ and $I(2)$ Models</th>
</tr>
</thead>
<tbody>
<tr>
<td>$r \backslash k - r - s$</td>
<td>1</td>
</tr>
<tr>
<td>0</td>
<td>$H_{00}$</td>
</tr>
<tr>
<td>1</td>
<td>$H_{10}$</td>
</tr>
<tr>
<td>$\vdots$</td>
<td>$\vdots$</td>
</tr>
<tr>
<td>$k-1$</td>
<td>$H_{k-1,0}$</td>
</tr>
</tbody>
</table>

Johansen (1995b) proposed the two-step procedure to analyze the $I(2)$ model. In the first step, the values of $(r, \alpha, \beta)$ are estimated using the reduced rank regression analysis, performing the regression analysis $\Delta^2 y_t$, and...
\( \Delta y_{t-1} \), and \( y_{t-1} \) on \( \Delta^2 y_{t-1}, \ldots, \Delta^2 y_{t-p+2} \), and \( D_t \). This gives residuals \( R_{0t}, R_{1t}, \) and \( R_{2t} \), and residual product moment matrices

\[
M_{ij} = \frac{1}{T} \sum_{t=1}^{T} R_{it} R'_{jt} \quad \text{for} \quad i, j = 0, 1, 2
\]

Perform the reduced rank regression analysis \( \Delta^2 y_t \) on \( y_{t-1} \) corrected for \( \Delta y_{t-1}, \Delta^2 y_{t-1}, \ldots, \Delta^2 y_{t-p+2} \), and \( D_t \), and solve the eigenvalue problem of the equation

\[
|\lambda M_{22.1} - M_{20.1} M_{01}^{-1} M_{02.1}| = 0
\]

where \( M_{ij.1} = M_{ij} - M_{i1} M_{11}^{-1} M_{1j} \) for \( i, j = 0, 2 \).

In the second step, if \( (r, \alpha, \beta) \) are known, the values of \( (s, \xi, \eta) \) are determined using the reduced rank regression analysis, regressing \( \hat{\alpha} \perp \Delta^2 y_t \) on \( \hat{\beta} \perp \Delta y_{t-1} \) corrected for \( \Delta^2 y_{t-1}, \ldots, \Delta^2 y_{t-p+2}, D_t, \) and \( \beta' \Delta y_{t-1} \).

The reduced rank regression analysis reduces to the solution of an eigenvalue problem for the equation

\[
|\rho M_{\beta \perp \beta \perp \beta} - M_{\beta \perp \alpha \perp \beta} M_{\alpha \perp \alpha \perp \beta} M_{\alpha \perp \beta \perp \beta}| = 0
\]

where

\[
M_{\beta \perp \beta \perp \beta} = \beta' (M_{11} - M_{11} \beta (\beta' M_{11} \beta)^{-1} \beta' M_{11}) \beta
\]
\[
M_{\beta \perp \alpha \perp \beta} = \alpha' (M_{01} - M_{01} \beta (\beta' M_{11} \beta)^{-1} \beta' M_{11}) \beta
\]
\[
M_{\alpha \perp \alpha \perp \beta} = \alpha' (M_{00} - M_{01} \beta (\beta' M_{11} \beta)^{-1} \beta' M_{10}) \alpha
\]

where \( \hat{\alpha} = \alpha (\alpha' \alpha)^{-1} \).

The solution gives eigenvalues \( 1 > \rho_1 > \cdots > \rho_s > 0 \) and eigenvectors \( (v_1, \ldots, v_s) \). Then, the ML estimators are

\[
\hat{\eta} = (v_1, \ldots, v_s)
\]
\[
\hat{\xi} = M_{\alpha \perp \beta \perp \beta} \hat{\eta}
\]

The likelihood ratio test for the reduced rank model \( H_{r,s} \) with rank \( \leq s \) in the model \( H_{r,k-r} = H_r^0 \) is given by

\[
Q_{r,s} = -T \sum_{i=s+1}^{k-r} \log(1 - \rho_i), \quad s = 0, \ldots, k - r - 1
\]

The following statements simulate an I(2) process and compute the rank test to test for cointegrated order 2:

```plaintext
proc iml;
    alpha = { 1, 1 };  * alphaOrthogonal = { 1, -1 };
    beta = { 1, -0.5 };  * betaOrthogonal = { 1, 2 };
    * alphaOrthogonal` * phiStar * betaOrthogonal = 0;
    phiStar = { 1 0, 0 0.5 };
    A1 = 2 * I(2) + alpha * beta` - phiStar;
```
A2 = phiStar - I(2);
phi = A1 // A2;
sig = I(2);

/* to simulate the vector time series */
call varmasim(y,phi) sigma=sig n=200 seed=2;
cn = {'y1' 'y2'};
create simul4 from y[colname=cn];
append from y;
close;
quit;

proc varmax data=simul4;
model y1 y2 /noint p=2 cointtest=(johansen=(iorder=2));
run;

The last two columns in Figure 43.80 explain the cointegration rank test with integrated order 1. For a
specified significance level, such as 5%, the output indicates that the null hypothesis that the series are not
cointegrated (H0: r = 0) is rejected, because the p-value for this test, shown in the column Pr > Trace of I(1),
is less than 0.05. The results also indicate that the null hypothesis that there is a cointegrated relationship
with cointegration rank 1 (H0: r = 1) cannot be rejected at the 5% significance level, because the p-value
for the test statistic, 0.7961, is greater than 0.05. Because of this latter result, the rows in the table that are
associated with r = 1 are further examined. The test statistic, 0.0257, tests the null hypothesis that the series
are cointegrated order 2. The p-value that is associated with this test is 0.8955, which indicates that the null
hypothesis cannot be rejected at the 5% significance level.

Figure 43.80  Cointegrated I(2) Test (IORDER= Option)

The VARMAX Procedure

<table>
<thead>
<tr>
<th>Cointegration Rank Test for I(2)</th>
<th>2</th>
<th>1</th>
<th>Pr &gt; Trace of I(1)</th>
<th>Pr &gt; Trace of I(1)</th>
</tr>
</thead>
<tbody>
<tr>
<td>r-k-r-s</td>
<td></td>
<td></td>
<td>Trace of I(1)</td>
<td>Trace of I(1)</td>
</tr>
<tr>
<td>0</td>
<td>575.3784</td>
<td>1.1833</td>
<td>215.3011</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Pr &gt; Trace of I(2)</td>
<td>0.0000</td>
<td>0.3223</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>0.0257</td>
<td>0.0986</td>
<td>0.7961</td>
<td></td>
</tr>
<tr>
<td>Pr &gt; Trace of I(2)</td>
<td>0.8955</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Vector Error Correction Model in ARMA Form

The vector error correction model in ARMA form (the VEC-ARMA model) introduces MA terms and is
defined as follows:

\[ \Delta y_t = \alpha \beta' y_{t-1} + \sum_{i=1}^{p-1} \Phi_i \Delta y_{t-i} + \epsilon_t - \sum_{i=1}^{q} \Theta_i \epsilon_{t-i} \]

The determined terms and the exogenous variables can also be introduced into the model. Similar to the
VECM that has only AR terms, the constant term is constrained in the error correction term in Case 2 and the
linear trend term is similarly constrained in Case 4.

The model is estimated through the maximum likelihood method. The log likelihood of the model is defined as
\[ \ell = -\frac{T}{2} \log \|\Sigma\| - \frac{1}{2} \sum_{t=1}^{T} \epsilon_t' \Sigma^{-1} \epsilon_t \]

where

\[ \epsilon_t = \Delta y_t - \alpha \beta' y_{t-1} - \sum_{i=1}^{p-1} \Phi_i' \Delta y_{t-i} + \sum_{i=1}^{q} \Theta_i \epsilon_{t-i} \]

conditional on the presample \( \{y_0, \ldots, y_{1-p}\} \), and \( \epsilon_s = 0, s \leq 0 \).

You can specify a VEC-ARMA(2,1) model with cointegration rank 2 on the three-dimensional time series by the following statements:

```plaintext
model y1-y3 / p=2 q=1;
cointeg rank=2;
```

For more information about modeling the cointegrated VARMA processes, see Lütkepohl (2007, Chapter 14).

### Multivariate GARCH Modeling

Stochastic volatility modeling is important in many areas, particularly in finance. To study the volatility of time series, GARCH models are widely used because they provide a good approach to conditional variance modeling.

**BEKK Representation**

Engle and Kroner (1995) propose a general multivariate GARCH model and call it a BEKK representation. Let \( \mathcal{F}(t-1) \) be the sigma field generated by the past values of \( \epsilon_t \), and let \( H_t \) be the conditional covariance matrix of the \( k \)-dimensional random vector \( \epsilon_t \). Let \( H_t \) be measurable with respect to \( \mathcal{F}(t-1) \); then the multivariate GARCH model can be written as

\[ \epsilon_t | \mathcal{F}(t-1) \sim N(0, H_t) \]

\[ H_t = C + \sum_{i=1}^{q} A_i' \epsilon_{t-i} \epsilon_{t-i}' A_i + \sum_{i=1}^{p} G_i' H_{t-i} G_i \]

where \( C, A_i, \) and \( G_i \) are \( k \times k \) parameter matrices.

Consider the bivariate GARCH(1,1) model

\[ H_t = \begin{bmatrix} c_{11} & c_{12} \\ c_{12} & c_{22} \end{bmatrix} + \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}' \begin{bmatrix} \epsilon_{1,t-1}^2 & \epsilon_{1,t-1} \epsilon_{2,t-1} \\ \epsilon_{2,t-1} \epsilon_{1,t-1} & \epsilon_{2,t-1}^2 \end{bmatrix} \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \\
+ \begin{bmatrix} g_{11} & g_{12} \\ g_{21} & g_{22} \end{bmatrix}' H_{t-1} \begin{bmatrix} g_{11} & g_{12} \\ g_{21} & g_{22} \end{bmatrix} \]

or, representing the univariate model,
For the BEKK representation of the bivariate GARCH(1,1) model, the SAS statements are

```sas
model y1 y2;
garch q=1 p=1 form=bekk;
```

The multistep forecast of the conditional covariance matrix, \( H_{t+h|t} \), \( h = 1, 2, \ldots \), is obtained recursively through the formula

\[
H_{t+h|t} = C + \sum_{i=1}^{h-1} A_i' H_{t+h-i|t} A_i + \sum_{i=h}^{q} A_i' \epsilon_{t+h-i} \epsilon_{t+h-i}' A_i + \sum_{i=1}^{p} G_i' H_{t+h-i|t} G_i
\]

where \( H_{s|t} = H_s \) for \( s \leq t \).

**CCC Representation**

Bollerslev (1990) proposes a multivariate GARCH model with time-varying conditional variances and covariances but constant conditional correlations.

The conditional covariance matrix \( H_t \) consists of

\[
H_t = D_t S D_t
\]

where \( D_t \) is a \( k \times k \) stochastic diagonal matrix with element \( \sigma_{i,t} \) and \( S \) is a \( k \times k \) time-invariant correlation matrix with the typical element \( s_{ij} \).

The element of \( H_t \) is

\[
h_{ij,t} = s_{ij} \sigma_{i,t} \sigma_{j,t} \quad i, j = 1, \ldots, k
\]

Note that \( h_{ii,t} = \sigma_{i,t}^2 \), \( i = 1, \ldots, k \).

If you specify CORRCONSTANT=EXPECT, the element \( s_{ij} \) of the time-invariant correlation matrix \( S \) is

\[
s_{ij} = \frac{1}{T} \sum_{t=1}^{T} \frac{\epsilon_{i,t}}{\sqrt{h_{ii,t}}} \frac{\epsilon_{j,t}}{\sqrt{h_{jj,t}}}
\]

where \( T \) is the sample size.
By default, or when you specify SUBFORM=GARCH, $\sigma_{i,t}^2$ follows a univariate GARCH process,

$$\sigma_{i,t}^2 = c_i + \sum_{l=1}^{q} a_{ii,l}\varepsilon_{i,t-l}^2 + \sum_{l=1}^{p} g_{ii,l}\sigma_{i,t-l}^2 \quad i = 1, \ldots, k$$

As shown in many empirical studies, positive and negative innovations have different impacts on future volatility. There is a long list of variations of univariate GARCH models that consider the asymmetricity. Four typical variations follow:

- exponential GARCH (EGARCH) model (Nelson and Cao 1992)
- quadratic GARCH (QGARCH) model (Engle and Ng 1993)
- threshold GARCH (TGARCH) model (Glosten, Jaganathan, and Runkle 1993; Zakoian 1994)
- power GARCH (PGARCH) model (Ding, Granger, and Engle 1993)

For more information about the asymmetric GARCH models, see Engle and Ng (1993). You can choose the type of GARCH model of interest by specifying the SUBFORM= option.

The EGARCH model was proposed by Nelson (1991). Nelson and Cao (1992) argue that the nonnegativity constraints in the GARCH model are too restrictive. The GARCH model, implicitly or explicitly, imposes the nonnegative constraints on the parameters, whereas these parameters have no restrictions in the EGARCH model. In the EGARCH model, the conditional variance is an asymmetric function of lagged disturbances,

$$\ln(\sigma_{i,t}^2) = c_i + \sum_{l=1}^{q} a_{ii,l}\left(b_{ii,l}\varepsilon_{i,t-l} + \frac{\varepsilon_{i,t-l}}{\sigma_{i,t-l}} - \sqrt{\frac{2}{\pi}}\right) + \sum_{l=1}^{p} g_{ii,l}\ln(\sigma_{i,t-l}^2) \quad i = 1, \ldots, k$$

In the QGARCH model, the lagged errors’ centers are shifted from zero to some constant values,

$$\sigma_{i,t}^2 = c_i + \sum_{l=1}^{q} a_{ii,l}(\varepsilon_{i,t-l} - b_{ii,l})^2 + \sum_{l=1}^{p} g_{ii,l}\sigma_{i,t-l}^2 \quad i = 1, \ldots, k$$

In the TGARCH model, each lagged squared error has an extra slope coefficient,

$$\sigma_{i,t}^2 = c_i + \sum_{l=1}^{q}(a_{ii,l} + 1_{\varepsilon_{i,t-l} < 0}b_{ii,l})\varepsilon_{i,t-l}^2 + \sum_{l=1}^{p} g_{ii,l}\sigma_{i,t-l}^2 \quad i = 1, \ldots, k$$

where the indicator function $1_{\varepsilon_{i,t} < 0}$ is one if $\varepsilon_{i,t} < 0$ and zero otherwise.
The PGARCH model not only considers the asymmetric effect but also provides a way to model the long memory property in the volatility,

\[ \sigma_{i,t}^{2\lambda_i} = c_i + \sum_{l=1}^{q} a_{i,i,l} (|\epsilon_{i,t-l}| - b_{i,i,l} \epsilon_{i,t-l})^{2\lambda_i} + \sum_{l=1}^{p} g_{i,i;l} \sigma_{i,t-l}^{2\lambda_i} \quad i = 1, \ldots, k \]

where \( \lambda_i > 0 \) and \( |b_{i,i,l}| \leq 1, l = 1, \ldots, q, i = 1, \ldots, k \).

Note that the implemented TGARCH model is also well known as GJR-GARCH (Glosten, Jaganathan, and Runkle 1993), which is similar to the threshold GARCH model proposed by Zakoian (1994) but not exactly the same. In Zakoian’s model, the conditional standard deviation is a linear function of the past values of the white noise. Zakoian’s model can be regarded as a special case of the PGARCH model when \( \lambda_i = 1/2 \).

The following formulas are recursively implemented to obtain the multistep forecast of conditional error variance \( \sigma_{i,t+h|t}^{2}, i = 1, \ldots, k \) and \( h = 1, 2, \ldots \):

- for the GARCH\((p, q)\) model:
  \[ \sigma_{i,t+h|t}^{2} = c_i + \sum_{l=1}^{h-1} a_{i,i,l} \sigma_{i,t+h-l|t}^{2} + \sum_{l=h}^{q} a_{i,i,l} \epsilon_{i,t+h-l|t}^{2} + \sum_{l=1}^{p} g_{i,i;l} \sigma_{i,t+h-l|t}^{2} \]

- for the EGARCH\((p, q)\) model:
  \[ \ln(\sigma_{i,t+h|t}^{2}) = c_i + \sum_{l=h}^{q} a_{i,i,l} \left( b_{i,i,l} \frac{\epsilon_{i,t+h-l|t}}{\sigma_{i,t+h-l|t}} + \frac{\epsilon_{i,t+h-l|t}}{\sigma_{i,t+h-l|t}} - \sqrt{\frac{2}{\pi}} \right) + \sum_{l=1}^{p} g_{i,i;l} \ln(\sigma_{i,t+h-l|t}^{2}) \]

- for the QGARCH\((p, q)\) model:
  \[ \sigma_{i,t+h|t}^{2} = c_i + \sum_{l=1}^{h-1} a_{i,i,l} (\sigma_{i,t+h-l|t}^{2} + b_{i,i,l}^{2}) + \sum_{l=h}^{q} a_{i,i,l} (\epsilon_{i,t+h-l|t} + b_{i,i,l})^{2} \]
  \[ + \sum_{l=1}^{p} g_{i,i;l} \sigma_{i,t+h-l|t}^{2} \]

- for the TGARCH\((p, q)\) model:
  \[ \sigma_{i,t+h|t}^{2} = c_i + \sum_{l=1}^{h-1} (a_{i,i,l} + b_{i,i,l}/2) \sigma_{i,t+h-l|t}^{2} + \sum_{l=h}^{q} (a_{i,i,l} + 1 \epsilon_{i,t-l} < 0 b_{i,i,l}) \epsilon_{i,t-l}^{2} \]
  \[ + \sum_{l=1}^{p} g_{i,i;l} \sigma_{i,t+h-l|t}^{2} \]
for the PGARCH\((p, q)\) model:

\[
\sigma_{i,t+h|t}^{2\lambda_i} = c_i + \sum_{l=1}^{h-1} a_{i,l}((1 + b_{i,l})^{2\lambda_i} + (1 - b_{i,l})^{2\lambda_i})\sigma_{i,t+h-l|t}/2
\]

\[
+ \sum_{l=h}^{q} a_{i,l}(|\epsilon_{i,t-l}| - b_{i,l}|\epsilon_{i,t-l}|)^{2\lambda_i} + \sum_{l=1}^{p} g_{i,l}\sigma_{i,t+h-1|t}^{2\lambda_i}
\]

In the preceding equations, \(\sigma_{i,s|t} = \sigma_{i,s}\) for \(s \leq t\). Then, the multistep forecast of conditional covariance matrix \(H_{t+h|t}, h = 1, 2, \ldots\), is calculated by

\[
H_{t+h|t} = D_{t+h|t}SD_{t+h|t}
\]

where \(D_{t+h|t}\) is the diagonal matrix with element \(\sigma_{i,t+h|t}, i = 1, \ldots, k\).

### DCC Representation

Engle (2002) proposes a parsimonious parametric multivariate GARCH model that has time-varying conditional covariances and correlations.

The conditional covariance matrix \(H_t\) consists of

\[
H_t = D_t \Gamma_t D_t
\]

where \(D_t\) is a \(k \times k\) stochastic diagonal matrix with the element \(\sigma_{i,t}\) and \(\Gamma_t\) is a \(k \times k\) time-varying matrix with the typical element \(\rho_{ij,t}\).

The element of \(H_t\) is

\[
h_{i,j,t} = \rho_{ij,t}\sigma_{i,t}\sigma_{j,t} \quad i, j = 1, \ldots, k
\]

Note that \(h_{ii,t} = \sigma_{i,t}^2, i = 1, \ldots, k\).

As in the CCC GARCH model, you can choose the type of GARCH model of interest by specifying the \text{SUBFORM=} option.

In the GARCH model,

\[
\sigma_{i,t}^2 = c_i + \sum_{l=1}^{q} a_{i,l}\epsilon_{i,t-l}^2 + \sum_{l=1}^{p} g_{i,l}\sigma_{i,t-l}^2 \quad i = 1, \ldots, k
\]

In the EGARCH model, the conditional variance is an asymmetric function of lagged disturbances,

\[
\ln(\sigma_{i,t}^2) = c_i + \sum_{l=1}^{q} a_{i,l}\epsilon_{i,t-l} + \frac{|\epsilon_{i,t-l}| - \sqrt{2/\pi}}{\sigma_{i,t-l}} + \sum_{l=1}^{p} g_{i,l}\ln(\sigma_{i,t-l}^2) \quad i = 1, \ldots, k
\]
In the QGARCH model, the lagged errors’ centers are shifted from zero to some constant values,

\[
\sigma_{i,t}^2 = c_i + \sum_{l=1}^{q} a_{i,l}(\epsilon_{i,t-l} - b_{i,l})^2 + \sum_{l=1}^{p} g_{i,l}\sigma_{i,t-1}^2 \quad i = 1, \ldots, k
\]

In the TGARCH model, each lagged squared error has an extra slope coefficient,

\[
\sigma_{i,t}^2 = c_i + \sum_{l=1}^{q} (a_{i,l} + \epsilon_{i,t-l}b_{i,l})\epsilon_{i,t-l}^2 + \sum_{l=1}^{p} g_{i,l}\sigma_{i,t-1}^2 \quad i = 1, \ldots, k
\]

where the indicator function \(1_{\epsilon_{i,t} < 0}\) is one if \(\epsilon_{i,t} < 0\) and zero otherwise.

The PGARCH model not only considers the asymmetric effect but also provides another way to model the long memory property in the volatility,

\[
\sigma_{i,t}^{2\lambda_i} = c_i + \sum_{l=1}^{q} a_{i,l}(|\epsilon_{i,t-l}| - b_{i,l}\epsilon_{i,t-l})^{2\lambda_i} + \sum_{l=1}^{p} g_{i,l}\sigma_{i,t-1}^{2\lambda_i} \quad i = 1, \ldots, k
\]

where \(\lambda_i > 0\) and \(|b_{i,l}| \leq 1, l = 1, \ldots, q; i = 1, \ldots, k\).

The conditional correlation estimator \(\rho_{i,j,t}\) is

\[
\rho_{i,j,t} = \frac{q_{i,j,t}}{\sqrt{q_{i,i,t}q_{j,j,t}}} \quad i, j = 1, \ldots, k
\]

\[
q_{i,j,t} = (1 - \alpha - \beta)s_{ij} + \alpha \frac{\epsilon_{i,t-1} - \epsilon_{j,t-1}}{\sigma_{i,t-1} \sigma_{j,t-1}} + \beta q_{i,j,t-1}
\]

where \(s_{ij}\) is the element of \(S\), the unconditional correlation matrix.

If you specify CORRCONSTANT=EXPECT, the element \(s_{ij}\) of the unconditional correlation matrix \(S\) is

\[
s_{ij} = \frac{1}{T} \sum_{t=1}^{T} \frac{\epsilon_{i,t} \epsilon_{j,t}}{\sigma_{i,t} \sigma_{j,t}}
\]

where \(T\) is the sample size.

As shown in the CCC GARCH models, the following formulas are recursively implemented to obtain the multistep forecast of conditional error variance \(\sigma_{i,t+h|t}^2, i = 1, \ldots, k\) and \(h = 1, 2, \ldots\):

- for the GARCH\((p, q)\) model:

\[
\sigma_{i,t+h|t}^2 = c_i + \sum_{l=1}^{h-1} a_{i,l} \sigma_{i,t+h-l|t}^2 + \sum_{l=h}^{q} a_{i,l} \epsilon_{i,t+h-l|t}^2 + \sum_{l=1}^{p} g_{i,l} \sigma_{i,t+h-1|t}^2
\]
Multivariate GARCH Modeling

For the EGARCH\((p, q)\) model:

\[
\ln(\sigma_{i,t+h|t}^2) = c_i + \sum_{l=h}^{q} a_{i,i,l} \left( b_{i,i,l} + \frac{\epsilon_{i,t+h-l}}{\sigma_{i,t+h-l}} + \frac{\epsilon_{i,t+h-l}}{\sigma_{i,t+h-l}} - \sqrt{\frac{2}{\pi}} \right) + \sum_{l=1}^{p} g_{i,i,l} \ln(\sigma_{i,t+h-l|t}^2)
\]

For the QGARCH\((p, q)\) model:

\[
\sigma_{i,t+h|t}^2 = c_i + \sum_{l=1}^{h-1} a_{i,i,l} (\sigma_{i,t+h-l|t}^2 + b_{i,i,l}^2) + \sum_{l=h}^{q} a_{i,i,l} (\epsilon_{i,t+h-l} - b_{i,i,l})^2 \\
+ \sum_{l=1}^{p} g_{i,i,l} \sigma_{i,t+h-l|t}^2
\]

For the TGARCH\((p, q)\) model:

\[
\sigma_{i,t+h|t}^2 = c_i + \sum_{l=1}^{h-1} a_{i,i,l} + b_{i,i,l}/2)\sigma_{i,t+h-1|t}^2 + \sum_{l=h}^{q} a_{i,i,l} + 1\epsilon_{i,t-l}<0b_{i,i,l})\epsilon_{i,t-l}^2 \\
+ \sum_{l=1}^{p} g_{i,i,l} \sigma_{i,t+h-1|t}^2
\]

For the PGARCH\((p, q)\) model:

\[
\sigma_{i,t+h|t}^{2\lambda_i} = c_i + \sum_{l=1}^{h-1} a_{i,i,l} (1 + b_{i,i,l})^{2\lambda_i} + (1 - b_{i,i,l})^{2\lambda_i} \sigma_{i,t+h-l|t}^{2\lambda_i} / 2 \\
+ \sum_{l=h}^{q} a_{i,i,l} (\epsilon_{i,t-l} - b_{i,i,l} \epsilon_{i,t-l})^{2\lambda_i} + \sum_{l=1}^{p} g_{i,i,l} \sigma_{i,t+h-1|t}^{2\lambda_i}
\]

In the preceding equations, \(\sigma_{i,s|t} = \sigma_{i,s} \) for \(s \leq t\). Then, the multistep forecast of conditional covariance matrix \(H_{t+h|t}, h = 1, 2, \ldots\), is calculated by

\[
H_{t+h|t} = D_{t+h|t} \Gamma_{t+h|t} D_{t+h|t}
\]

where \(D_{t+h|t}\) is the diagonal matrix with element \(\sigma_{i,t+h|t}, i = 1, \ldots, k\), and \(\Gamma_{t+h|t}\) is the matrix with element \(\rho_{i,j,t+h|t}, i, j = 1, \ldots, k\).

\[
\rho_{i,j,t+h|t} = \begin{cases} \\
\frac{q_{i,i,t+h|t}}{\sqrt{q_{i,i,t+h|t}q_{j,j,t+h|t}}} \\
(1 - \alpha - \beta) s_{ij} + \alpha \frac{\epsilon_{i,t} \epsilon_{j,t}}{\sigma_{i,t} \sigma_{j,t}} + \beta q_{i,j,t} & h = 1 \\
(1 - \alpha - \beta) s_{ij} + \alpha q_{i,j,t+h-1|t} + \beta q_{i,j,t+h-1|t} & h > 1 
\end{cases}
\]
Estimation of GARCH Model

The log-likelihood function of the multivariate GARCH model is written without a constant term as

$$\ell = -\frac{1}{2} \sum_{t=1}^{T} [\log |H_t| + \epsilon_t' H_t^{-1} \epsilon_t]$$

where $\epsilon_t$ is calculated from the first-moment model (that is, the VARMAX model or VEC-ARMA model). The log-likelihood function is maximized by an iterative numerical method such as quasi-Newton optimization. The starting values for the regression parameters are obtained from the least squares estimates. The covariance of $\epsilon_t$ is used as the starting value for the GARCH constant parameters, and the starting values for the other GARCH parameters are either $10^{-6}$ or $10^{-3}$, depending on the GARCH model’s representation.

Prediction of Endogenous (Dependent) Variables

In multivariate GARCH models, the optimal (minimum MSE) $l$-step-ahead forecast of endogenous variables $y_{t+l|t}$ uses the same formula as shown in the section “Forecasting” on page 3082. However, the exogenous (independent) variables, if present, are always assumed to be nonstochastic (deterministic); that is, to predict the endogenous variables, you must specify the future values of the exogenous variables. The prediction error of the optimal $l$-step-ahead forecast is

$$e_{t+l|t} = y_{t+l} - y_{t+l|t} = \sum_{j=0}^{l-1} \Psi_j \epsilon_{t+l-j},$$

with zero mean and covariance matrix,

$$\Sigma_t(l) = \text{Cov}(e_{t+l|t}) = \sum_{j=0}^{l-1} \Psi_j H_{t+l-j|t} \Psi_j'$$

where $H_{t+h|t}$, $h = 1, \ldots, l$, is the $h$-step-ahead forecast of the conditional covariance matrix. As emphasized by the subscript $t$, $\Sigma_t(l)$ is time-dependent. In the OUT= data set, the forecast standard errors and prediction intervals are constructed according to $\Sigma_t(l)$. If you specify the COVPE option, the prediction error covariances that are output in the CovPredictError and CovPredictErrorbyVar ODS tables are based on the time-independent formula

$$\Sigma(l) = \sum_{j=0}^{l-1} \Psi_j \Sigma \Psi_j'$$

where $\Sigma$ is the unconditional covariance matrix of innovations. The decomposition of the prediction error covariances is also based on $\Sigma(l)$.

Covariance Stationarity

Define the multivariate GARCH process as

$$h_t = \sum_{i=1}^{\infty} G(B)^{i-1} [c + A(B) \eta_t]$$
where $h_t = \text{vec}(H_t)$, $c = \text{vec}(C_0)$, and $\eta_t = \text{vec}(\epsilon_t \epsilon_t')$. This representation is equivalent to a GARCH$(p,q)$ model by the following algebra:

$$h_t = c + A(B)\eta_t + \sum_{i=2}^{\infty} G(B)^{i-1}[c + A(B)\eta_t]$$

$$= c + A(B)\eta_t + G(B)\sum_{i=1}^{\infty} G(B)^{i-1}[tmbc + A(B)\eta_t]$$

$$= c + A(B)\eta_t + G(B)h_t$$

Defining $A(B) = \sum_{i=1}^{q}(A_i \otimes A_i)'B^i$ and $G(B) = \sum_{i=1}^{p}(G_i \otimes G_i)'B^i$ gives a BEKK representation.

The necessary and sufficient conditions for covariance stationarity of the multivariate GARCH process are that all the eigenvalues of $A(1) + G(1)$ are less than 1 in modulus.

**An Example of a VAR(1)–ARCH(1) Model**

The following DATA step simulates a bivariate vector time series to provide test data for the multivariate GARCH model:

```sas
data garch;
  retain seed 16587;
esq1 = 0; esq2 = 0;
ly1 = 0; ly2 = 0;
do i = 1 to 1000;
  ht = 6.25 + 0.5*esq1;
call rannor(seed,ehat);
e1 = sqrt(ht)*ehat;
ht = 1.25 + 0.7*esq2;
call rannor(seed,ehat);
e2 = sqrt(ht)*ehat;
y1 = 2 + 1.2*ly1 - 0.5*ly2 + e1;
y2 = 4 + 0.6*ly1 + 0.3*ly2 + e2;
if i>500 then output;
esq1 = e1*e1; esq2 = e2*e2;
ly1 = y1; ly2 = y2;
end;
keep y1 y2;
run;
```

The following statements fit a VAR(1)–ARCH(1) model to the data. For a VAR-ARCH model, you specify the order of the autoregressive model with the P=1 option in the MODEL statement and the Q=1 option in the GARCH statement. In order to produce the initial and final values of parameters, the TECH=QN option is specified in the NLOPTIONS statement.

```sas
proc varmax data=garch;
  model y1 y2 / p=1
    print=(roots estimates diagnose);
garch q=1;
nloptions tech=qn;
run;
```
Figure 43.81 through Figure 43.85 show the details of this example. Figure 43.81 shows the initial values of parameters.

**Figure 43.81** Start Parameter Estimates for the VAR(1)–ARCH(1) Model

<table>
<thead>
<tr>
<th>Parameter Estimate</th>
<th>Gradient Objective Function</th>
</tr>
</thead>
<tbody>
<tr>
<td>CONST1 2.249575</td>
<td>0.000082533</td>
</tr>
<tr>
<td>CONST2 3.902673</td>
<td>0.000401</td>
</tr>
<tr>
<td>AR1_1 1.231775</td>
<td>0.000105</td>
</tr>
<tr>
<td>AR1_2 0.576890</td>
<td>-0.004811</td>
</tr>
<tr>
<td>AR1_1 0.528405</td>
<td>0.000617</td>
</tr>
<tr>
<td>AR1_2 0.343714</td>
<td>0.001811</td>
</tr>
<tr>
<td>GCHC1 9.929763</td>
<td>0.151293</td>
</tr>
<tr>
<td>GCHC2 0.193163</td>
<td>-0.014305</td>
</tr>
<tr>
<td>ACH1_1 4.063245</td>
<td>0.370333</td>
</tr>
<tr>
<td>ACH1_1 0.001000</td>
<td>-0.667182</td>
</tr>
<tr>
<td>ACH1_2 0.001000</td>
<td>-0.668905</td>
</tr>
<tr>
<td>ACH1_2 0.001000</td>
<td>-0.734486</td>
</tr>
<tr>
<td>ACH1_2 -3.127035</td>
<td></td>
</tr>
</tbody>
</table>

Figure 43.82 shows the final parameter estimates.

**Figure 43.82** Results of Parameter Estimates for the VAR(1)–ARCH(1) Model

<table>
<thead>
<tr>
<th>Parameter Estimate</th>
<th>Gradient Objective Function</th>
</tr>
</thead>
<tbody>
<tr>
<td>CONST1 2.156865</td>
<td>0.000246</td>
</tr>
<tr>
<td>CONST2 4.048879</td>
<td>0.000105</td>
</tr>
<tr>
<td>AR1_1 1.224620</td>
<td>-0.001957</td>
</tr>
<tr>
<td>AR1_2 0.609651</td>
<td>0.000173</td>
</tr>
<tr>
<td>AR1_1 -0.534248</td>
<td>-0.000468</td>
</tr>
<tr>
<td>AR1_2 0.302599</td>
<td>-0.000375</td>
</tr>
<tr>
<td>GCHC1 8.238625</td>
<td>-0.000056090</td>
</tr>
<tr>
<td>GCHC2 -0.231183</td>
<td>-0.000021724</td>
</tr>
<tr>
<td>GCHC2 1.565459</td>
<td>0.000110</td>
</tr>
<tr>
<td>ACH1_1 0.374255</td>
<td>-0.000419</td>
</tr>
<tr>
<td>ACH1_2 0.035883</td>
<td>-0.000606</td>
</tr>
<tr>
<td>ACH1_1 0.057461</td>
<td>0.001636</td>
</tr>
<tr>
<td>ACH1_2 0.717897</td>
<td>-0.000149</td>
</tr>
</tbody>
</table>
Figure 43.83 shows the conditional variance by using the BEKK representation of the ARCH(1) model. The ARCH parameters are estimated as follows by the vectorized parameter matrices:

\[ \epsilon_t|\mathcal{F}(t-1) \sim N(0, H_t) \]

\[
H_t = \begin{bmatrix}
8.23863 & -0.23118 \\
-0.23118 & 1.56546
\end{bmatrix} + \begin{bmatrix}
0.37426 & 0.05746 \\
0.03588 & 0.71790
\end{bmatrix} \epsilon_{t-1} \epsilon_{t-1}' \begin{bmatrix}
0.37426 & 0.05746 \\
0.03588 & 0.71790
\end{bmatrix}
\]

**Figure 43.83** ARCH(1) Parameter Estimates for the VAR(1)–ARCH(1) Model

The **VARMAX Procedure**

<table>
<thead>
<tr>
<th>Type of Model</th>
<th>VAR(1)-ARCH(1)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Estimation Method</td>
<td>Maximum Likelihood Estimation</td>
</tr>
<tr>
<td>Representation Type</td>
<td>BEKK</td>
</tr>
</tbody>
</table>

**GARCH Model Parameter Estimates**

| Parameter | Estimate | Standard Error | t Value | Pr > |t| |
|-----------|----------|----------------|---------|------|---|
| GCCHC1_1  | 8.23863  | 0.72663        | 11.34   | 0.0001 |
| GCCHC1_2  | -0.23118 | 0.21434        | -1.08   | 0.2813 |
| GCHC2_2   | 1.56546  | 0.19407        | 8.07    | 0.0001 |
| ACH1_1_1  | 0.37426  | 0.07502        | 4.99    | 0.0001 |
| ACH1_2_1  | 0.03588  | 0.06974        | 0.51    | 0.6071 |
| ACH1_2_2  | 0.05746  | 0.02597        | 2.21    | 0.0274 |
| ACH1_2_2  | 0.71790  | 0.06895        | 10.41   | 0.0001 |

Figure 43.84 shows the AR parameter estimates and their significance.

The fitted VAR(1) model with the previous conditional covariance ARCH model is written as follows:

\[
y_t = \begin{bmatrix}
2.15687 \\
4.04888
\end{bmatrix} + \begin{bmatrix}
1.22462 & -0.53425 \\
0.60965 & 0.30260
\end{bmatrix} y_{t-1} + \epsilon_t
\]

**Figure 43.84** VAR(1) Parameter Estimates for the VAR(1)–ARCH(1) Model

| Equation | Parameter | Estimate | Standard Error | t Value | Pr > |t| Variable |
|----------|-----------|----------|----------------|---------|------|----------|
| y1       | CONST1    | 2.15687  | 0.21717        | 9.93    | 0.0001 | 1        |
|          | AR1_1_1   | 1.22462  | 0.02542        | 48.17   | 0.0001 | y1(t-1)  |
|          | AR1_1_2   | -0.53425 | 0.02807        | -19.03  | 0.0001 | y2(t-1)  |
| y2       | CONST2    | 4.04888  | 0.10663        | 37.97   | 0.0001 | 1        |
|          | AR1_2_1   | 0.60965  | 0.01216        | 50.13   | 0.0001 | y1(t-1)  |
|          | AR1_2_2   | 0.30260  | 0.01491        | 20.30   | 0.0001 | y2(t-1)  |
Figure 43.85 shows the roots of the AR and ARCH characteristic polynomials. The eigenvalues have a modulus less than one.

**Figure 43.85** Roots for the VAR(1)–ARCH(1) Model

<table>
<thead>
<tr>
<th>Index</th>
<th>Real</th>
<th>Imaginary</th>
<th>Modulus</th>
<th>Radian</th>
<th>Degree</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.76361</td>
<td>0.33641</td>
<td>0.8344</td>
<td>0.4150</td>
<td>23.7762</td>
</tr>
<tr>
<td>2</td>
<td>0.76361</td>
<td>-0.33641</td>
<td>0.8344</td>
<td>-0.4150</td>
<td>-23.7762</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Index</th>
<th>Real</th>
<th>Imaginary</th>
<th>Modulus</th>
<th>Radian</th>
<th>Degree</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.52388</td>
<td>0.00000</td>
<td>0.5239</td>
<td>0.0000</td>
<td>0.0000</td>
</tr>
<tr>
<td>2</td>
<td>0.26661</td>
<td>0.00000</td>
<td>0.2666</td>
<td>0.0000</td>
<td>0.0000</td>
</tr>
<tr>
<td>3</td>
<td>0.26661</td>
<td>0.00000</td>
<td>0.2666</td>
<td>0.0000</td>
<td>0.0000</td>
</tr>
<tr>
<td>4</td>
<td>0.13569</td>
<td>0.00000</td>
<td>0.1357</td>
<td>0.0000</td>
<td>0.0000</td>
</tr>
</tbody>
</table>

**VARFIMA and VARFIMAX Modeling**

VAR and VARMA series are short-range dependent (SRD) in the sense that their autocovariance function dies out exponentially fast with the increasing lag. However, in many financial and macroeconomics applications, stationary yet persistent series arise, calling for models that have a slowly decaying autocovariance function and that are therefore more suitable to capture long-range dependence in the data.

The VARFIMA model captures both long-range and short-range dependence dynamics in a multivariate series. For a $k$-dimensional series $y_t = (y_{1t},\ldots,y_{kt})'$, $t = 1,\ldots,T$, the VARFIMA($p,D,q$) model is defined as

$$
\Phi(B)y_t = (I - B)^{-D}\Theta(B)\epsilon_t
$$

where $B$ and $I$ are the backshift and identity operators; $D = \text{diag}(d_j)$ $d_j \in (-1/2,1/2)$, are the LRD parameters of the component series $\{y_{jt}\}_{t \in \mathbb{Z}}$, $j = 1,\ldots,k$; and $\{\epsilon_t\}_{t \in \mathbb{Z}}$ is a $k$-dimensional white noise series with zero mean $\mathbb{E}\epsilon_t = 0$ and covariance $\mathbb{E}\epsilon_t\epsilon_t' = \Sigma$.

The fractional integration operator $(I - B)^{-D}$ allows for long memory in the series. On the other hand, $\Phi(z)$ and $\Theta(z)$, which are the typical autoregressive and moving average matrix polynomials of orders $p$ and $q$, respectively, capture the short-range dependence.

The VARFIMA($p,D,q$) series satisfies the multivariate long-range dependence definitions given in Kechagias and Pipiras (2015). Moreover, each component series $\{y_{jt}\}_{t \in \mathbb{Z}}$, $j = 1,\ldots,k$, satisfies the univariate time and frequency domain LRD definitions given in Beran et al. (2013). The following sections briefly review these definitions and show how you can detect long-range dependence in the data before fitting a VARFIMA model.

**Autocorrelation and Spectral Density of VARFIMA Series**

The diagonal components of the autocorrelation matrix function of a VARFIMA($p,D,q$) series satisfy the univariate LRD time domain definition

$$
\rho_l(n) \sim c_1n^{2d_i-1}, \quad i = 1,\ldots,k, \quad \text{as} \quad n \to \infty
$$
where $a_n \sim b_n$ implies that $\lim_{n \to \infty} a_n / b_n = 1$ and $c_1 > 0$. Similarly, the diagonal components of the spectral density matrix function of a VARFIMA($p, D, q$) series satisfy

$$f_i(\lambda) \sim c_2 \lambda^{-2d_i}, \ i = 1, \ldots, k, \ \text{as} \ \lambda \to 0^+$$

for some $c_2 > 0$.

To obtain preliminary estimates of the LRD parameters, you can plot the logged periodogram values against the log of the Fourier frequencies $\lambda_j = 2\pi j / T$, $j = 1, \ldots, T/2$, and then fit a line for frequencies near 0. The slope of this line is expected to be equal to $-2d_i$ (the exponent in the right-hand side of the preceding relation). The following statements demonstrate this procedure for a synthetic VARFIMA(1, D, 1) series with $T = 2000$ and true parameters $d_1 = 0.4, \ d_2 = 0.3, \ \Phi_{11} = \Sigma_{11} = \Sigma_{22} = 3, \ \Sigma_{12} = 0.5, \ \Phi_{11} = 0.8, \ \Phi_{12} = 0.3, \ \Phi_{21} = -0.2, \ \Phi_{22} = 0.1, \ \Theta_{11} = 0.2, \ \Theta_{12} = 0.4, \ \Theta_{21} = 0, \ \text{and} \ \Theta_{22} = 0.3$:

```sas
data VARFIMA1D1;
  time = _N_;  
  input y1 y2;
  datalines;
1.495250048 2.694910375
... more lines ...  
3.12049851 5.330308391
7.732287586 1.665071247
; /* Compute the two periodograms */
proc spectra data = VARFIMA1D1 out = spectra;
  var y1 y2;
run;
/* Convert to log scale */
data logspectra;
  set spectra(firstobs=2);
  /* Compute Fourier frequencies */
  j = _N_;  
  pi = constant('pi');
  logfreq = log(2*pi*j/2000);
  logpdg1 = log(P_01);
  logpdg2 = log(P_02);
  /* Introduce weights where regression will be performed */
  wt = (1<= j <=100);
  keep wt logfreq logpdg1 logpdg2;
run;
/* Regression for log-periodogram of y1*/
proc autoreg data = logspectra(obs = 100);
  model logpdg1 = logfreq;
run;
```
/* Regression for log-periodogram of y1*/
proc autoreg data = logspectra(obs = 100);
   model logpdg2 = logfreq;
run;

The output from the two regressions is shown in Figure 43.86 and Figure 43.87.

**Figure 43.86** Regression Estimates for y1

<table>
<thead>
<tr>
<th>Parameter Estimates</th>
</tr>
</thead>
<tbody>
<tr>
<td>Variable</td>
</tr>
<tr>
<td>Intercept</td>
</tr>
<tr>
<td>logfreq</td>
</tr>
</tbody>
</table>

**Figure 43.87** Regression Estimates for y2

<table>
<thead>
<tr>
<th>Parameter Estimates</th>
</tr>
</thead>
<tbody>
<tr>
<td>Variable</td>
</tr>
<tr>
<td>Intercept</td>
</tr>
<tr>
<td>logfreq</td>
</tr>
</tbody>
</table>

The following statements produce log-log plots of the two periodograms along with the regression lines:

/*Plot the periodograms in log-log scale*/
ods graphics on;

proc sgplot data = logspectra;
   series x = logfreq y = logpdg1 / lineattrs = (pattern = solid);
   reg y = logpdg1 x = logfreq / normarkers weight = wt lineattrs =
     (thickness = 1 color = 'red');
   inset "Slope = -0.905" / position = topright textattrs = (color = 'red');
   xaxis label = 'log-frequency';
   yaxis label = 'log-periodogram';
   title 'Log-periodogram of y1';
run;

proc sgplot data = logspectra;
   series x = logfreq y = logpdg2 / lineattrs = (pattern = solid);
   reg y = logpdg2 x = logfreq / normarkers weight = wt lineattrs =
     (thickness = 1 color = 'red');
   inset "Slope = -0.523" / position = topright textattrs = (color = 'red');
   xaxis label = 'log-frequency';
   yaxis label = 'log-periodogram';
   title 'Log-periodogram of y2';
run;

The final plots are shown in Figure 43.88.
Dividing the slopes by 2 and removing the negative signs yields preliminary estimates for the LRD parameters, $\hat{d}_1 = 0.45$ and $\hat{d}_2 = 0.26$.

**Estimation**

Estimation of all the parameters in the VARFIMA model is performed using the conditional likelihood Durbin-Levinson (CLDL) algorithm of Tsay (2010). This method uses the multivariate Durbin-Levinson algorithm, whose order of complexity is $O(T^2)$, making it computationally feasible for small or medium sample sizes.

The initial values of the LRD parameters are obtained by the semiparametric estimator of Geweke and Porter-Hudak (1983). The initial values of the AR and MA parameters are obtained from least squares estimation on the fractionally differenced series $(I - B)^D y_t$. The LRD parameters are restricted in the range $(-1/2, 1/2)$. If an initial LRD parameter estimate is outside this range, then the chosen starting value is either $-1/2 + 10^{-6}$ or $1/2 - 10^{-6}$ for negative or positive initial semiparametric estimates, respectively.

**Forecasting**

One-step-ahead and multi-step-ahead forecasts for the VARFIMA series are based on a finite past. However, the $h$-step-ahead forecast errors for $h > 1$ are based on the infinite past except for VARFIMA series that have only MA components. In the latter case, the forecast errors are also based on a finite past.

The following statements plot the $h$-step-ahead forecasts, $h = 1, \ldots, 36$, for a bivariate synthetic VARFIMA($1, D, 1$) series with $T = 400$ and true parameters $d_1 = 0.4, d_2 = 0.3, \Phi_{11} = \Sigma_{11} = \Sigma_{22} = 3, \Sigma_{12} = 0.5, \Phi_{11} = 0.8, \Phi_{12} = 0.3, \Phi_{22} = -0.2, \Theta_{11} = 0.1, \Theta_{12} = 0.2, \Theta_{21} = 0.4, \Theta_{22} = 0$, and $\Theta_{22} = 0.3$. The statements also specify initial values for $d_1$ and $d_2$ close to the true parameter values.

```plaintext
data VARFIMA1D1N4;
  time = _N_;  
  input y1 y2;
  datalines;
  0.55596529  2.114409393
  -1.842925215  3.415027987
  ... more lines ...
```
Chapter 43: The VARMAX Procedure

-2.86707489 1.147627529
-0.195787414 0.820107072

proc varmax data = VARFIMA1D1N4 plots = (forecasts);
  model y1 y2 / noint fi p=1 q=1;
  initial d(1) = 0.45, d(2) = 0.25;
  output out = forec back = 36 lead = 36;
run;

The BACK option in the preceding SAS statements is used to specify the point where the historical data ends and multi-step-ahead forecasting begins. Note that the BACK option does not affect estimation. The latter is performed using the whole data set, even when you specify the BACK option.

Impulse Response Functions

The impulse response functions of the VARFIMA series are calculated using the methodology of Chung (2001). The following statements produce the first 12 simple, accumulated and orthogonal impulse response functions and their corresponding standard errors for the VARFIMA(1, D, 1) series of the preceding example.

```
proc varmax data = VARFIMA1D1N4 plots = (impulse);
  model y1 y2 / noint fi p=1 q=1 print = (impulse = (all));
run;
```

VARFIMAX Modeling

The VARFIMAX(p, D, q, s) series is defined as

\[
\Phi(B)y_t + \Theta^*(B)x_t = (I - B)^{-D}\Theta(B)\epsilon_t
\]

where \(x_t = (x_{1t}, \ldots, x_{rt})\), \(t = 1, \ldots, T\), is an \(r\)-dimensional time series vector of exogenous variables and \(\Theta^*(z)\) is the order \(s\) matrix polynomial defined as \(\Theta^*(z) = \Theta_0^* + \Theta_1^*z + \cdots + \Theta_s^*z^s\) for some \(k \times r\) real matrices \(\Theta_i^*, i = 1, \ldots, s\).

The following statements estimate a bivariate VARFIMAX(1, D, 1, 0) model:

```sas
model y1 y2 = x1 / fi p=1 q=1;
```

Figure 43.89 Plot of the Two Series and \(h\)-Step-Ahead Forecasts, \(h = 1, \ldots, 36\)
Output Data Sets

The VARMAX procedure can create the OUT=, OUTTEST=, OUTHT=, and OUTSTAT= data sets. In general, if processing fails, the output is not recorded or is set to missing in the relevant output data set, and appropriate error and/or warning messages are recorded in the log.

OUT= Data Set

The OUT= data set contains the forecast values produced by the OUTPUT statement. The following output variables can be created:

- the BY variables
- the ID variable
- the MODEL statement dependent (endogenous) variables. These variables contain the actual values from the input data set.
- FORi, numeric variables that contain the forecasts. The FORi variables contain the forecasts for the ith endogenous variable in the MODEL statement list. Forecasts are one-step-ahead predictions until the end of the data or until the observation specified by the BACK= option. Multistep forecasts can be computed after that point based on the LEAD= option.
- RESi, numeric variables that contain the residual for the forecast of the ith endogenous variable in the MODEL statement list. For multistep forecast observations, the actual values are missing and the RESi variables contain missing values.
- STDi, numeric variables that contain the standard deviation for the forecast of the ith endogenous variable in the MODEL statement list. The values of the STDi variables can be used to construct univariate confidence limits for the corresponding forecasts.
- LCIi, numeric variables that contain the lower confidence limits for the corresponding forecasts of the ith endogenous variable in the MODEL statement list.
- UCIi, numeric variables that contain the upper confidence limits for the corresponding forecasts of the ith endogenous variable in the MODEL statement list.

The OUT= data set contains the values shown in Table 43.7 and Table 43.8 for a bivariate case.

<table>
<thead>
<tr>
<th>Obs</th>
<th>ID Variable</th>
<th>y1</th>
<th>FOR1</th>
<th>RES1</th>
<th>STD1</th>
<th>LCI1</th>
<th>UCI1</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>date</td>
<td>y11</td>
<td>f11</td>
<td>r11</td>
<td>σ11</td>
<td>l11</td>
<td>u11</td>
</tr>
<tr>
<td>2</td>
<td>date</td>
<td>y12</td>
<td>f12</td>
<td>r12</td>
<td>σ11</td>
<td>l12</td>
<td>u12</td>
</tr>
<tr>
<td>...</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Consider the following example:

```plaintext
proc varmax data=simul1 noprint;
    id date interval=year;
    model y1 y2 / p=1 noint;
    output out=out lead=5;
run;
proc print data=out(firstobs=98);
run;
```

The output in Figure 43.90 shows part of the results of the OUT= data set for the preceding example.

**Figure 43.90  OUT= Data Set**

<table>
<thead>
<tr>
<th>Log-periodogram of y2</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Obs</strong></td>
</tr>
<tr>
<td>-------</td>
</tr>
<tr>
<td>98</td>
</tr>
<tr>
<td>99</td>
</tr>
<tr>
<td>100</td>
</tr>
<tr>
<td>101</td>
</tr>
<tr>
<td>102</td>
</tr>
<tr>
<td>103</td>
</tr>
<tr>
<td>104</td>
</tr>
<tr>
<td>105</td>
</tr>
</tbody>
</table>

**OUTEST= Data Set**

The OUTEST= data set contains estimation results of the fitted model produced by the VARMAX statement. The following output variables can be created:

- **BY variables**
- **NAME**, a character variable that contains the name of the endogenous (dependent) variables or the name of the parameters for the covariance of the matrix of the parameter estimates if you specify the OUTCOV option
- **TYPE**, a character variable that contains the value EST for parameter estimates, the value STD for standard error of parameter estimates, and the value COV for the covariance of the matrix of the parameter estimates if you specify the OUTCOV option
• **CONST**, a numeric variable that contains the estimates of constant parameters and their standard errors

• **SEASON\_i**, a numeric variable that contains the estimates of seasonal dummy parameters and their standard errors, where \(i = 1, \ldots, (nseason - 1)\), and \(nseason\) is based on the NSEASON= option

• **LTREND**, a numeric variable that contains the estimates of linear trend parameters and their standard errors

• **QTREND**, a numeric variable that contains the estimates of quadratic trend parameters and their standard errors

• **XL\_l\_i**, numeric variables that contain the estimates of exogenous parameters and their standard errors, where \(l\) is the lag \(l\)th coefficient matrix and \(i = 1, \ldots, r\), where \(r\) is the number of exogenous variables

• **AR\_l\_i**, numeric variables that contain the estimates of autoregressive parameters and their standard errors, where \(l\) is the lag \(l\)th coefficient matrix and \(i = 1, \ldots, k\), where \(k\) is the number of endogenous variables

• **MA\_l\_i**, numeric variables that contain the estimates of moving-average parameters and their standard errors, where \(l\) is the lag \(l\)th coefficient matrix and \(i = 1, \ldots, k\), where \(k\) is the number of endogenous variables

• **COV\_i**, numeric variables that contain the estimates of the covariance of innovations parameters when the maximum likelihood method is applied, where \(i = 1, \ldots, k\)

• **DCCAB**, a numeric variable that contains the estimates of \(\alpha\) or \(\beta\) in the correlation equation for DCC representation and their standard errors

• **CCC\_i**, numeric variables that contain the estimates of the conditional constant correlation parameters for CCC representation, where \(i = 2, \ldots, k\)

• **DCCS\_i**, numeric variables that contain the estimates of the unconditional correlation parameters for DCC representation, where \(i = 2, \ldots, k\)

• **GCHC\_i**, numeric variables that contain the estimates of the constant parameters of the covariance matrix and their standard errors, where \(i = 1, \ldots, k\) for BEKK representation, \(k\) is the number of endogenous variables, and \(i = 1\) for CCC and DCC representations

• **ACH\_l\_i**, numeric variables that contain the estimates of the ARCH parameters of the covariance matrix and their standard errors, where \(l\) is the lag \(l\)th coefficient matrix and \(i = 1, \ldots, k\) for BEKK, CCC, and DCC representations, where \(k\) is the number of endogenous variables

• **EACH\_l\_i**, numeric variables that contain the estimates of the exponential ARCH parameters of the covariance matrix and their standard errors, where \(l\) is the lag \(l\)th coefficient matrix and \(i = 1, \ldots, k\) for CCC and DCC representations, where \(k\) is the number of endogenous variables

• **PACH\_l\_i**, numeric variables that contain the estimates of the power ARCH parameters of the covariance matrix and their standard errors, where \(l\) is the lag \(l\)th coefficient matrix and \(i = 1, \ldots, k\) for CCC and DCC representations, where \(k\) is the number of endogenous variables

• **QACH\_l\_i**, numeric variables that contain the estimates of the quadratic ARCH parameters of the covariance matrix and their standard errors, where \(l\) is the lag \(l\)th coefficient matrix and \(i = 1, \ldots, k\) for CCC and DCC representations, where \(k\) is the number of endogenous variables
The OUTEST= data set contains the values shown in Table 43.9 for a bivariate case.

<table>
<thead>
<tr>
<th>Obs</th>
<th>NAME</th>
<th>TYPE</th>
<th>CONST</th>
<th>AR1_1</th>
<th>AR1_2</th>
<th>AR2_1</th>
<th>AR2_2</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>y1</td>
<td>EST</td>
<td>δ1</td>
<td>φ1,11</td>
<td>φ1,12</td>
<td>φ2,11</td>
<td>φ2,12</td>
</tr>
<tr>
<td>2</td>
<td>STD</td>
<td>se(δ1)</td>
<td>se(φ1,11)</td>
<td>se(φ1,12)</td>
<td>se(φ2,11)</td>
<td>se(φ2,12)</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>y2</td>
<td>EST</td>
<td>δ2</td>
<td>φ1,21</td>
<td>φ1,22</td>
<td>φ2,21</td>
<td>φ2,22</td>
</tr>
<tr>
<td>4</td>
<td>STD</td>
<td>se(δ2)</td>
<td>se(φ1,21)</td>
<td>se(φ1,22)</td>
<td>se(φ2,21)</td>
<td>se(φ2,22)</td>
<td></td>
</tr>
</tbody>
</table>

Consider the following example:

```plaintext
proc varmax data=simul2 outest=est;
  model y1 y2 / p=2 noint noprint;
  cointeg rank=1 normalize=y1;
run;
proc print data=est;
run;
```

The output in Figure 43.91 shows the results of the OUTEST= data set.

**Figure 43.91 OUTEST= Data Set**

<table>
<thead>
<tr>
<th>Log-periodogram of y2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Obs</td>
</tr>
<tr>
<td>-----</td>
</tr>
<tr>
<td>1</td>
</tr>
<tr>
<td>2</td>
</tr>
<tr>
<td>3</td>
</tr>
<tr>
<td>4</td>
</tr>
</tbody>
</table>

**OUTHT= Data Set**

The OUTHT= data set contains predictions of conditional covariance matrices of innovations of the fitted GARCH model that the GARCH statement produces. The following output variables can be created:
• the BY variables, if BY-group processing is performed
• the ID variable, if the ID statement is specified
• $H_{ij}$, numeric variables that contain the prediction of covariance, where $1 \leq i \leq j \leq k$, where $k$ is the number of dependent variables

The OUTHT= data set contains the values shown in Table 43.10 for a bivariate case.

<table>
<thead>
<tr>
<th>Table 43.10</th>
<th>OUTHT= Data Set</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Obs</strong></td>
<td><strong>H1_1</strong></td>
</tr>
<tr>
<td>1</td>
<td>h111</td>
</tr>
<tr>
<td>2</td>
<td>h112</td>
</tr>
<tr>
<td>:</td>
<td>:</td>
</tr>
</tbody>
</table>

The OUTHT= data set has the same number of observations as the OUT= data set. Both the OUTHT= and OUT= data sets include any observations at the beginning of the data set that are skipped because of the DIF=, DIFY=, DIFX=, P=, or XLAG= option and include the predicted observations at the end of the data set, which correspond with the LEAD= specification. If you specify an ID statement together with the OUTHT= and OUT= options, then the values of the ID variable in the two data sets correspond with one another.

Consider the following example of the OUTHT= option:

```plaintext
data garch;
  set garch;
  date = intnx( 'month', '01may1972'd, _n_-1 );
  format date yymms.;
run;

proc varmax data=garch;
  id date interval=month;
  model y1 y2 / p=1;
  garch q=1 outht=ht;
  output out=og lead=6;
run;

proc print data=og(obs=8);
  var date y1 for1 std1 lci1 uci1 y2 for2 std2 lci2 uci2;
run;

proc print data=ht(obs=8);
run;

proc print data=og(firstobs=499);
  var date y1 for1 std1 lci1 uci1 y2 for2 std2 lci2 uci2;
run;

proc print data=ht(firstobs=499);
run;
```
The output in Figure 43.92 and Figure 43.93 shows the first eight observations in the OUT= and OUTHT= data sets, respectively. The first observation is skipped in the GARCH model estimation because of the P=1 option, resulting in the missing values in the first observations in the OUT= and OUTHT= data sets. The output in Figure 43.94 and Figure 43.95 shows the last eight observations in the OUT= and OUTHT= data sets, respectively. In the OUT= data set, the standard deviations of the forecast of dependent variables are time-variant. The last six observations in OUTHT= data set are the multistep forecast of conditional covariance matrices of innovations.

**Figure 43.92** First Part of OUT= Data Set

Log-periodogram of y2

<table>
<thead>
<tr>
<th>Obs</th>
<th>date</th>
<th>y1</th>
<th>FOR1</th>
<th>STD1</th>
<th>LCI1</th>
<th>UCI1</th>
<th>y2</th>
<th>FOR2</th>
<th>STD2</th>
<th>LCI2</th>
<th>UCI2</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1972/05</td>
<td>-4.4005</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>1.83794</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td>2</td>
<td>1972/06</td>
<td>-8.0533</td>
<td>-4.2140</td>
<td>3.10387</td>
<td>-10.2975</td>
<td>1.86947</td>
<td>1.59720</td>
<td>1.92227</td>
<td>1.92885</td>
<td>-1.85820</td>
<td>5.70274</td>
</tr>
<tr>
<td>3</td>
<td>1972/07</td>
<td>-10.8362</td>
<td>-8.5587</td>
<td>3.21511</td>
<td>-14.8602</td>
<td>-2.25720</td>
<td>1.51833</td>
<td>0.37752</td>
<td>1.33100</td>
<td>1.298623</td>
<td>2.23118</td>
</tr>
<tr>
<td>5</td>
<td>1972/09</td>
<td>-7.8272</td>
<td>-4.3716</td>
<td>3.63437</td>
<td>-11.4949</td>
<td>2.75160</td>
<td>-0.03774</td>
<td>-0.9637</td>
<td>1.44118</td>
<td>2.92702</td>
<td>2.72829</td>
</tr>
<tr>
<td>8</td>
<td>1972/12</td>
<td>-8.0182</td>
<td>-7.5245</td>
<td>2.87208</td>
<td>-13.1537</td>
<td>-1.89535</td>
<td>0.43513</td>
<td>-0.65343</td>
<td>1.61823</td>
<td>-3.82511</td>
<td>2.51825</td>
</tr>
</tbody>
</table>

**Figure 43.93** First Part of OUTHT= Data Set

Log-periodogram of y2

<table>
<thead>
<tr>
<th>Obs</th>
<th>date</th>
<th>h1_1</th>
<th>h1_2</th>
<th>h2_2</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1972/05</td>
<td>.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>1972/06</td>
<td>9.6340</td>
<td>0.14073</td>
<td>3.72045</td>
</tr>
<tr>
<td>3</td>
<td>1972/07</td>
<td>10.3369</td>
<td>0.42643</td>
<td>1.77155</td>
</tr>
<tr>
<td>4</td>
<td>1972/08</td>
<td>8.8538</td>
<td>-1.19603</td>
<td>3.07876</td>
</tr>
<tr>
<td>5</td>
<td>1972/09</td>
<td>13.2086</td>
<td>1.36328</td>
<td>2.07699</td>
</tr>
<tr>
<td>6</td>
<td>1972/10</td>
<td>9.9058</td>
<td>-0.02914</td>
<td>1.58995</td>
</tr>
<tr>
<td>7</td>
<td>1972/11</td>
<td>8.3757</td>
<td>-0.29722</td>
<td>1.59728</td>
</tr>
<tr>
<td>8</td>
<td>1972/12</td>
<td>8.2489</td>
<td>-0.12736</td>
<td>2.61868</td>
</tr>
</tbody>
</table>

**Figure 43.94** Last Part of OUT= Data Set

Log-periodogram of y2

<table>
<thead>
<tr>
<th>Obs</th>
<th>date</th>
<th>y1</th>
<th>FOR1</th>
<th>STD1</th>
<th>LCI1</th>
<th>UCI1</th>
<th>y2</th>
<th>FOR2</th>
<th>STD2</th>
<th>LCI2</th>
<th>UCI2</th>
</tr>
</thead>
<tbody>
<tr>
<td>500</td>
<td>2013/12</td>
<td>-10.2133</td>
<td>-8.6817</td>
<td>2.97211</td>
<td>-14.5070</td>
<td>-2.8565</td>
<td>2.88544</td>
<td>2.11833</td>
<td>1.28490</td>
<td>-0.4000</td>
<td>4.6367</td>
</tr>
</tbody>
</table>
The OUTSTAT= data set contains estimation results of the fitted model produced by the VARMAX statement.
The following output variables can be created. The subindex $i$ is 1, \ldots, k, where $k$ is the number of endogenous variables.

- the BY variables
- NAME, a character variable that contains the name of endogenous (dependent) variables
- SIGMA$_i$, numeric variables that contain the estimate of the innovation covariance matrix
- AICC, a numeric variable that contains the corrected Akaike’s information criterion value
- HQC, a numeric variable that contains the Hannan-Quinn’s information criterion value
- AIC, a numeric variable that contains the Akaike’s information criterion value
- SBC, a numeric variable that contains the Schwarz Bayesian’s information criterion value
- FPEC, a numeric variable that contains the final prediction error criterion value
- FValue, a numeric variable that contains the $F$ statistics
- PValue, a numeric variable that contains $p$-value for the $F$ statistics

If the JOHANSEN= option is specified, the following items are added:

- Eigenvalue, a numeric variable that contains eigenvalues for the cointegration rank test of integrated order 1
- RestrictedEigenvalue, a numeric variable that contains eigenvalues for the cointegration rank test of integrated order 1 when the NOINT option is not specified
- Beta$_i$, numeric variables that contain long-run effect parameter estimates, $\beta$
• Alpha_i, numeric variables that contain adjustment parameter estimates, α

If the JOHANSEN=(IORDER=2) option is specified, the following items are added:

• EValueI2_i, numeric variables that contain eigenvalues for the cointegration rank test of integrated order 2
• EValueI1, a numeric variable that contains eigenvalues for the cointegration rank test of integrated order 1
• Eta_i, numeric variables that contain the parameter estimates in integrated order 2, η
• Xi_i, numeric variables that contain the parameter estimates in integrated order 2, ξ

The OUTSTAT= data set contains the values shown Table 43.11 for a bivariate case.

Table 43.11 OUTSTAT= Data Set

<table>
<thead>
<tr>
<th>Obs</th>
<th>NAME</th>
<th>SIGMA_1</th>
<th>SIGMA_2</th>
<th>AICC</th>
<th>RSquare</th>
<th>FValue</th>
<th>PValue</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>y1</td>
<td>σ₁₁</td>
<td>σ₁₂</td>
<td>aicc</td>
<td>R²₁</td>
<td>F₁</td>
<td>prob₁</td>
</tr>
<tr>
<td>2</td>
<td>y2</td>
<td>σ₂₁</td>
<td>σ₂₂</td>
<td>.</td>
<td>R²₂</td>
<td>F₂</td>
<td>prob₂</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Obs</th>
<th>EValueI2_1</th>
<th>EValueI2_2</th>
<th>EValueI1</th>
<th>Beta_1</th>
<th>Beta_2</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>e₁₁</td>
<td>e₁₂</td>
<td>e₁</td>
<td>β₁₁</td>
<td>β₁₂</td>
</tr>
<tr>
<td>2</td>
<td>e₂₁</td>
<td>.</td>
<td>e₂</td>
<td>β₂₁</td>
<td>β₂₂</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Obs</th>
<th>Alpha_1</th>
<th>Alpha_2</th>
<th>Eta_1</th>
<th>Eta_2</th>
<th>Xi_1</th>
<th>Xi_2</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>α₁₁</td>
<td>α₁₂</td>
<td>η₁₁</td>
<td>η₁₂</td>
<td>ξ₁₁</td>
<td>ξ₁₂</td>
</tr>
<tr>
<td>2</td>
<td>α₂₁</td>
<td>α₂₂</td>
<td>η₂₁</td>
<td>η₂₂</td>
<td>ξ₂₁</td>
<td>ξ₂₂</td>
</tr>
</tbody>
</table>

Consider the following example:

```plaintext
proc varmax data=simul2 outstat=stat;
  model y1 y2 / p=2 noint noprint
    cointest=(johansen=(iorder=2));
  cointeg rank=1 normalize=y1;
run;

proc print data=stat;
run;
```

The output in Figure 43.96 shows the results of the OUTSTAT= data set.
The default printed output produced by the VARMAX procedure is described in the following list:

- descriptive statistics, which include the number of observations used, the names of the variables, their means and standard deviations (STD), their minimums and maximums, the differencing operations used, and the labels of the variables
- a type of model to fit the data and an estimation method
- a table of parameter estimates that shows the following for each parameter: the variable name for the left-hand side of equation, the parameter name, the parameter estimate, the approximate standard error, \( t \) value, the approximate probability \( (P_r > |t|) \), and the variable name for the right-hand side of equations in terms of each parameter
- the innovation covariance matrix
- the information criteria

If PRINT=ESTIMATES is specified, the VARMAX procedure prints the following list with the default printed output:

- the estimates of the constant vector (or seasonal constant matrix), the trend vector, the coefficient matrices of the distributed lags, the AR coefficient matrices, and the MA coefficient matrices
- the \( \alpha \) and \( \beta \) parameter estimates for the error correction model
- the schematic representation of parameter estimates

If PRINT=DIAGNOSE is specified, the VARMAX procedure prints the following list with the default printed output:

- the cross-covariance and cross-correlation matrices of the residuals
- the tables of test statistics for the hypothesis that the residuals of the model are white noise:
  - Durbin-Watson (DW) statistics
- $F$ test for autoregressive conditional heteroscedastic (ARCH) disturbances
- $F$ test for AR disturbance
- Jarque-Bera normality test
- portmanteau test

### ODS Table Names

The VARMAX procedure assigns a name to each table that it creates. You can use these names to reference the table when using the Output Delivery System (ODS) to select tables and create output data sets. These names are listed in Table 43.12.

<table>
<thead>
<tr>
<th>ODS Table Name</th>
<th>Description</th>
<th>Option</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>ODS Tables Created by the MODEL Statement</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AccumImpulse</td>
<td>Accumulated impulse response matrices</td>
<td>IMPULSE=(ACCUM)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>IMPULSE=(ALL)</td>
</tr>
<tr>
<td>AccumImpulsebyVar</td>
<td>Accumulated impulse response by variable</td>
<td>IMPULSE=(ACCUM)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>IMPULSE=(ALL)</td>
</tr>
<tr>
<td>AccumImpulseX</td>
<td>Accumulated transfer function matrices</td>
<td>IMPULSX=(ACCUM)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>IMPULSX=(ALL)</td>
</tr>
<tr>
<td>AccumImpulseXbyVar</td>
<td>Accumulated transfer function by variable</td>
<td>IMPULSX=(ACCUM)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>IMPULSX=(ALL)</td>
</tr>
<tr>
<td>Alpha</td>
<td>$\alpha$ coefficients</td>
<td>JOHANSEN=</td>
</tr>
<tr>
<td>AlphaInECM</td>
<td>$\alpha$ coefficients when RANK=$r$</td>
<td>PRINT=(ESTIMATES)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>with ECM=</td>
</tr>
<tr>
<td>AlphaOnDrift</td>
<td>$\alpha$ coefficients under the restriction of a deterministic term</td>
<td>JOHANSEN=</td>
</tr>
<tr>
<td>AlphaBetaInECM</td>
<td>$\Pi = \alpha\beta'$ coefficients when RANK=$r$</td>
<td>PRINT=(ESTIMATES)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>with ECM=</td>
</tr>
<tr>
<td>ANOVA</td>
<td>Univariate model diagnostic checks for the residuals</td>
<td>PRINT=DIAGNOSE</td>
</tr>
<tr>
<td>ARCoef</td>
<td>AR coefficients</td>
<td>PRINT=(ESTIMATES)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>with P=</td>
</tr>
<tr>
<td>ARRoots</td>
<td>Roots of AR characteristic polynomial</td>
<td>ROOTS with P=</td>
</tr>
<tr>
<td>Beta</td>
<td>$\beta$ coefficients</td>
<td>JOHANSEN=</td>
</tr>
<tr>
<td>BetaInECM</td>
<td>$b\beta$ coefficients when RANK=$r$</td>
<td>PRINT=(ESTIMATES)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>with ECM=</td>
</tr>
<tr>
<td>BetaOnDrift</td>
<td>$\beta$ coefficients under the restriction of a deterministic term</td>
<td>JOHANSEN=</td>
</tr>
<tr>
<td>CCCCorrConstant</td>
<td>Constant correlation matrix in the CCC GARCH model</td>
<td>CORRCONSTANT=EXPECT</td>
</tr>
<tr>
<td></td>
<td></td>
<td>with FORM=CCC</td>
</tr>
<tr>
<td>Constant</td>
<td>Constant estimates</td>
<td>Without NOINT</td>
</tr>
<tr>
<td>CorrB</td>
<td>Correlations of parameter estimates</td>
<td>CORRB</td>
</tr>
<tr>
<td>CorrResiduals</td>
<td>Correlations of residuals</td>
<td>PRINT=DIAGNOSE</td>
</tr>
<tr>
<td>ODS Table Name</td>
<td>Description</td>
<td>Option</td>
</tr>
<tr>
<td>--------------------------</td>
<td>------------------------------------------------------</td>
<td>-------------------------------</td>
</tr>
<tr>
<td>CorrResidualsbyVar</td>
<td>Correlations of residuals by variable</td>
<td>PRINT=DIAGNOSE</td>
</tr>
<tr>
<td>CorrResidualsGraph</td>
<td>Schematic representation of correlations of residuals</td>
<td>PRINT=DIAGNOSE</td>
</tr>
<tr>
<td>CorrXGraph</td>
<td>Schematic representation of sample correlations of independent series</td>
<td>CORRX</td>
</tr>
<tr>
<td>CorrYGraph</td>
<td>Schematic representation of sample correlations of dependent series</td>
<td>CORRY</td>
</tr>
<tr>
<td>CorrXLags</td>
<td>Correlations of independent series</td>
<td>CORRX</td>
</tr>
<tr>
<td>CorrXbyVar</td>
<td>Correlations of independent series by variable</td>
<td>CORRX</td>
</tr>
<tr>
<td>CorrYLags</td>
<td>Correlations of dependent series</td>
<td>CORRY</td>
</tr>
<tr>
<td>CorrYbyVar</td>
<td>Correlations of dependent series by variable</td>
<td>CORRY</td>
</tr>
<tr>
<td>CovarianceParameter-</td>
<td>Covariance parameter estimates</td>
<td>METHOD=ML without the PRIOR= option, or GARCH statement</td>
</tr>
<tr>
<td>Estimates</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CovB</td>
<td>Covariances of parameter estimates</td>
<td>COVB</td>
</tr>
<tr>
<td>CovInnovation</td>
<td>Covariances of the innovations</td>
<td>Default</td>
</tr>
<tr>
<td>CovPredictError</td>
<td>Covariance matrices of the prediction error</td>
<td>COVPE</td>
</tr>
<tr>
<td>CovPredictErrorbyVar</td>
<td>Covariances of the prediction error by variable</td>
<td>COVPE</td>
</tr>
<tr>
<td>CovResiduals</td>
<td>Covariances of residuals</td>
<td>PRINT=DIAGNOSE</td>
</tr>
<tr>
<td>CovResidualsbyVar</td>
<td>Covariances of residuals by variable</td>
<td>PRINT=DIAGNOSE</td>
</tr>
<tr>
<td>CovXLags</td>
<td>Covariances of independent series</td>
<td>COVX</td>
</tr>
<tr>
<td>CovXbyVar</td>
<td>Covariances of independent series by variable</td>
<td>COVX</td>
</tr>
<tr>
<td>CovYLags</td>
<td>Covariances of dependent series</td>
<td>COVY</td>
</tr>
<tr>
<td>CovYbyVar</td>
<td>Covariances of dependent series by variable</td>
<td>COVY</td>
</tr>
<tr>
<td>DCCCorrConstant</td>
<td>Unconditional correlation matrix in the DCC GARCH model</td>
<td>CORRCONSTANT=EXPECT with FORM=DCC</td>
</tr>
<tr>
<td>DecomposeCovPredictError</td>
<td>Decomposition of the prediction error covariances</td>
<td>DECOMPOSE</td>
</tr>
<tr>
<td>DecomposeCovPredictErrorbyVar</td>
<td>Decomposition of the prediction error covariances by variable</td>
<td>DECOMPOSE</td>
</tr>
<tr>
<td>DFTest</td>
<td>Dickey-Fuller test</td>
<td>DFTEST</td>
</tr>
<tr>
<td>DiagnostAR</td>
<td>Test the AR disturbance for the residuals</td>
<td>PRINT=DIAGNOSE</td>
</tr>
<tr>
<td>DiagnostWN</td>
<td>Test the ARCH disturbance and normality for the residuals</td>
<td>PRINT=DIAGNOSE</td>
</tr>
<tr>
<td>DynamicARCoef</td>
<td>AR coefficients of the dynamic model</td>
<td>DYNAMIC</td>
</tr>
<tr>
<td>DynamicConstant</td>
<td>Constant estimates of the dynamic model</td>
<td>DYNAMIC</td>
</tr>
<tr>
<td>DynamicCovInnovation</td>
<td>Covariances of the innovations of the dynamic model</td>
<td>DYNAMIC</td>
</tr>
</tbody>
</table>
Table 43.12  continued

<table>
<thead>
<tr>
<th>ODS Table Name</th>
<th>Description</th>
<th>Option</th>
</tr>
</thead>
<tbody>
<tr>
<td>DynamicLinearTrend</td>
<td>Linear trend estimates of the dynamic model</td>
<td>DYNAMIC</td>
</tr>
<tr>
<td>DynamicMACoeff</td>
<td>MA coefficients of the dynamic model</td>
<td>DYNAMIC</td>
</tr>
<tr>
<td>DynamicSConstant</td>
<td>Seasonal constant estimates of the dynamic model</td>
<td>DYNAMIC</td>
</tr>
<tr>
<td>DynamicParameter-Estimates</td>
<td>Parameter estimates table of the dynamic model</td>
<td>DYNAMIC</td>
</tr>
<tr>
<td>DynamicParameter-Graph</td>
<td>Schematic representation of the parameters of the dynamic model</td>
<td>DYNAMIC</td>
</tr>
<tr>
<td>DynamicQuadTrend</td>
<td>Quadratic trend estimates of the dynamic model</td>
<td>DYNAMIC</td>
</tr>
<tr>
<td>DynamicSeasonGraph</td>
<td>Schematic representation of the seasonal dummies of the dynamic model</td>
<td>DYNAMIC</td>
</tr>
<tr>
<td>DynamicXLagCoef</td>
<td>Dependent coefficients of the dynamic model</td>
<td>DYNAMIC</td>
</tr>
<tr>
<td>Hypothesis</td>
<td>Hypothesis of different deterministic terms in cointegration rank test</td>
<td>JOHANSEN=</td>
</tr>
<tr>
<td>HypothesisTest</td>
<td>Test hypothesis of different deterministic terms in cointegration rank test</td>
<td>JOHANSEN=</td>
</tr>
<tr>
<td>EigenvalueI2</td>
<td>Eigenvalues in integrated order 2</td>
<td>JOHANSEN= (IORDER=2)</td>
</tr>
<tr>
<td>Eta</td>
<td>$\eta$ coefficients</td>
<td>JOHANSEN= (IORDER=2)</td>
</tr>
<tr>
<td>InfiniteARRepresent</td>
<td>Infinite order ar representation</td>
<td>IARR</td>
</tr>
<tr>
<td>InfoCriteria</td>
<td>Information criteria</td>
<td>Default</td>
</tr>
<tr>
<td>LinearTrend</td>
<td>Linear trend estimates</td>
<td>TREND=</td>
</tr>
<tr>
<td>LogLikelihood</td>
<td>Log likelihood</td>
<td>Default</td>
</tr>
<tr>
<td>MACoeff</td>
<td>MA coefficients</td>
<td>Q=</td>
</tr>
<tr>
<td>MARoots</td>
<td>Roots of MA characteristic polynomial</td>
<td>ROOTS with Q=</td>
</tr>
<tr>
<td>MaxTest</td>
<td>Cointegration rank test using the maximum eigenvalue</td>
<td>JOHANSEN= (TYPE=MAX)</td>
</tr>
<tr>
<td>Minic</td>
<td>Tentative order selection</td>
<td>MINIC or MINIC=</td>
</tr>
<tr>
<td>ModelType</td>
<td>Type of model</td>
<td>Default</td>
</tr>
<tr>
<td>NObs</td>
<td>Number of observations</td>
<td>Default</td>
</tr>
<tr>
<td>OrthoImpulse</td>
<td>Orthogonalized impulse response matrices</td>
<td>IMPULSE=(ORTH) IMPULSE=(ALL)</td>
</tr>
<tr>
<td>OrthoImpulsebyVar</td>
<td>Orthogonalized impulse response by variable</td>
<td>IMPULSE=(ORTH) IMPULSE=(ALL)</td>
</tr>
<tr>
<td>ParameterEstimates</td>
<td>Parameter estimates table</td>
<td>Default</td>
</tr>
<tr>
<td>ParameterGraph</td>
<td>Schematic representation of the parameters</td>
<td>PRINT=ESTIMATES</td>
</tr>
<tr>
<td>PartialAR</td>
<td>Partial autoregression matrices</td>
<td>PARCOEF</td>
</tr>
<tr>
<td>PartialARGraph</td>
<td>Schematic representation of partial autoregression</td>
<td>PARCOEF</td>
</tr>
<tr>
<td>ODS Table Name</td>
<td>Description</td>
<td>Option</td>
</tr>
<tr>
<td>---------------------------</td>
<td>-------------------------------------------------------</td>
<td>-----------------------------</td>
</tr>
<tr>
<td>PartialCanCorr</td>
<td>Partial canonical correlation analysis</td>
<td>PCANCORR</td>
</tr>
<tr>
<td>PartialCorr</td>
<td>Partial cross-correlation matrices</td>
<td>PCORR</td>
</tr>
<tr>
<td>PartialCorrbyVar</td>
<td>Partial cross-correlations by variable</td>
<td>PCORR</td>
</tr>
<tr>
<td>PartialCorrGraph</td>
<td>Schematic representation of partial cross-correlations</td>
<td>PCORR</td>
</tr>
<tr>
<td>PortmanteauTest</td>
<td>Chi-square test table for residual cross-correlations</td>
<td>PRINT=DIAGNOSE</td>
</tr>
<tr>
<td>ProportionCovPredictError</td>
<td>Proportions of prediction error covariance decomposition</td>
<td>DECOMPOSE</td>
</tr>
<tr>
<td>ProportionCovPredictErrorbyVar</td>
<td>Proportions of prediction error covariance decomposition by variable</td>
<td>DECOMPOSE</td>
</tr>
<tr>
<td>RankTestI2</td>
<td>Cointegration rank test in integrated order 2</td>
<td>JOHANSEN=(IORDER=2)</td>
</tr>
<tr>
<td>RestrictMaxTest</td>
<td>Cointegration rank test using the maximum eigenvalue under the restriction of a deterministic term</td>
<td>JOHANSEN=(TYPE=MAX) without NOINT</td>
</tr>
<tr>
<td>RestrictTraceTest</td>
<td>Cointegration rank test using the trace under the restriction of a deterministic term</td>
<td>JOHANSEN=(TYPE=TRACE) without NOINT</td>
</tr>
<tr>
<td>QuadTrend</td>
<td>Quadratic trend estimates</td>
<td>TREND=QUAD</td>
</tr>
<tr>
<td>SeasonGraph</td>
<td>Schematic representation of the seasonal dummies</td>
<td>PRINT=ESTIMATES with NSEASON=</td>
</tr>
<tr>
<td>SConstant</td>
<td>Seasonal constant estimates</td>
<td>NSEASON=</td>
</tr>
<tr>
<td>SimpleImpulse</td>
<td>Impulse response matrices</td>
<td>IMPULSE=(SIMPLE)</td>
</tr>
<tr>
<td>SimpleImpulsebyVar</td>
<td>Impulse response by variable</td>
<td>IMPULSE=(ALL)</td>
</tr>
<tr>
<td>SimpleImpulseX</td>
<td>Impulse response matrices of transfer function</td>
<td>IMPULSX=(SIMPLE)</td>
</tr>
<tr>
<td>SimpleImpulseXbyVar</td>
<td>Impulse response of transfer function by variable</td>
<td>IMPULSX=(ALL)</td>
</tr>
<tr>
<td>Summary</td>
<td>Simple summary statistics</td>
<td>Default</td>
</tr>
<tr>
<td>SWTTest</td>
<td>Common trends test</td>
<td>SW=</td>
</tr>
<tr>
<td>TraceTest</td>
<td>Cointegration rank test using the trace</td>
<td>JOHANSEN=(TYPE=TRACE)</td>
</tr>
<tr>
<td>Xi</td>
<td>ξ coefficient matrix</td>
<td>JOHANSEN=</td>
</tr>
<tr>
<td>XLagCoef</td>
<td>Dependent coefficients</td>
<td>XLAG=</td>
</tr>
<tr>
<td>YWEstimates</td>
<td>Yule-Walker estimates</td>
<td>YW</td>
</tr>
</tbody>
</table>
### Table 43.12  continued

<table>
<thead>
<tr>
<th>ODS Table Name</th>
<th>Description</th>
<th>Option</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>ODS Tables Created by the GARCH Statement</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ARCHCoef</td>
<td>ARCH coefficients</td>
<td>Q=</td>
</tr>
<tr>
<td>GARCHCoef</td>
<td>GARCH coefficients</td>
<td>P=</td>
</tr>
<tr>
<td>GARCHConstant</td>
<td>GARCH constant estimates</td>
<td>PRINT=ESTIMATES</td>
</tr>
<tr>
<td>GARCHParameter- Estimates</td>
<td>GARCH parameter estimates table</td>
<td>Default</td>
</tr>
<tr>
<td>GARCHParameter- Graph</td>
<td>Schematic representation of the garch parameters</td>
<td>PRINT=ESTIMATES</td>
</tr>
<tr>
<td>GARCHRoots</td>
<td>Roots of GARCH characteristic polynomial</td>
<td>ROOTS</td>
</tr>
<tr>
<td><strong>ODS Tables Created by the COINTEG Statement or the ECM Option in the MODEL Statement</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AlphaAndBetaPa- rameterEstimaters</td>
<td>Parameter estimates of $\alpha$, $\beta$, $\beta_0$, and $\beta_1$</td>
<td>Default</td>
</tr>
<tr>
<td>AlphaInECM</td>
<td>$\alpha$ coefficients when RANK=$r$</td>
<td>PRINT=ESTIMATES</td>
</tr>
<tr>
<td>AlphaBetaInECM</td>
<td>$\Pi = \alpha \beta'$ coefficients when RANK=$r$</td>
<td>PRINT=ESTIMATES</td>
</tr>
<tr>
<td>AlphaOnAlpha</td>
<td>$\alpha$ coefficients under the restriction of $\alpha$</td>
<td>J=</td>
</tr>
<tr>
<td>AlphaOnBeta</td>
<td>$\alpha$ coefficients under the restriction of $\beta$</td>
<td>H=</td>
</tr>
<tr>
<td>AlphaTestResults</td>
<td>Hypothesis testing of $\alpha$</td>
<td>J=</td>
</tr>
<tr>
<td>BetaInECM</td>
<td>$\beta$ coefficients when RANK=$r$</td>
<td>PRINT=ESTIMATES</td>
</tr>
<tr>
<td>BetaOnBeta</td>
<td>$\beta$ coefficients under the restriction of $\beta$</td>
<td>H=</td>
</tr>
<tr>
<td>BetaOnAlpha</td>
<td>$\beta$ coefficients under the restriction of $\alpha$</td>
<td>J=</td>
</tr>
<tr>
<td>BetaTestResults</td>
<td>Hypothesis testing of $\beta$</td>
<td>H=</td>
</tr>
<tr>
<td>GrangerRepresent</td>
<td>Coefficient of Granger representation</td>
<td>PRINT=ESTIMATES</td>
</tr>
<tr>
<td>HMatrix</td>
<td>Restriction matrix for $\beta$</td>
<td>H=</td>
</tr>
<tr>
<td>JMatrix</td>
<td>Restriction matrix for $\alpha$</td>
<td>J=</td>
</tr>
<tr>
<td>WeakExogeneity</td>
<td>Testing weak exogeneity of each dependent variable with respect to BETA</td>
<td>EXOGENEITY</td>
</tr>
<tr>
<td><strong>ODS Tables Created by the CAUSAL Statement</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CausalityTest</td>
<td>Granger causality test</td>
<td>Default</td>
</tr>
<tr>
<td>GroupVars</td>
<td>Two groups of variables</td>
<td>Default</td>
</tr>
<tr>
<td><strong>ODS Tables Created by the RESTRICT Statement</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Restrict</td>
<td>Restriction table</td>
<td>Default</td>
</tr>
<tr>
<td><strong>ODS Tables Created by the TEST Statement</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Test</td>
<td>Wald test</td>
<td>Default</td>
</tr>
<tr>
<td><strong>ODS Tables Created by the OUTPUT Statement</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Forecasts</td>
<td>Forecasts table</td>
<td>Without NOPRINT</td>
</tr>
</tbody>
</table>

Note that the ODS table names suffixed by “byVar” can be obtained with the PRINTFORM=UNIVARIATE option.
This section describes the use of ODS for creating statistical graphs with the VARMAX procedure.

When ODS GRAPHICS are in effect, the VARMAX procedure produces a variety of plots for each dependent variable.

The plots available are as follows:

- The procedure displays the following plots for each dependent variable in the MODEL statement with the PLOT= option in the VARMAX statement:
  - impulse response function
  - impulse response of the transfer function
  - time series and predicted series
  - prediction errors
  - distribution of the prediction errors
  - normal quantile of the prediction errors
  - ACF of the prediction errors
  - PACF of the prediction errors
  - IACF of the prediction errors
  - log scaled white noise test of the prediction errors

- The procedure displays forecast plots for each dependent variable in the OUTPUT statement with the PLOT= option in the VARMAX statement.

### ODS Graph Names

The VARMAX procedure assigns a name to each graph it creates by using ODS. You can use these names to reference the graphs when using ODS. The names are listed in Table 43.13.

<table>
<thead>
<tr>
<th>ODS Table Name</th>
<th>Plot Description</th>
<th>Statement</th>
</tr>
</thead>
<tbody>
<tr>
<td>ErrorACFPlot</td>
<td>Autocorrelation function of prediction</td>
<td>MODEL</td>
</tr>
<tr>
<td></td>
<td>errors</td>
<td></td>
</tr>
<tr>
<td>ErrorIACFPlot</td>
<td>Inverse autocorrelation function of</td>
<td>MODEL</td>
</tr>
<tr>
<td></td>
<td>prediction errors</td>
<td></td>
</tr>
<tr>
<td>ErrorPACFPlot</td>
<td>Partial autocorrelation function of</td>
<td>MODEL</td>
</tr>
<tr>
<td></td>
<td>prediction errors</td>
<td></td>
</tr>
<tr>
<td>ErrorDiagnosticsPanel</td>
<td>Diagnostics of prediction errors</td>
<td>MODEL</td>
</tr>
<tr>
<td>ErrorNormalityPanel</td>
<td>Histogram and Q-Q plot of prediction errors</td>
<td>MODEL</td>
</tr>
<tr>
<td>ErrorDistribution</td>
<td>Distribution of prediction errors</td>
<td>MODEL</td>
</tr>
<tr>
<td>ErrorQQPlot</td>
<td>Q-Q plot of prediction errors</td>
<td>MODEL</td>
</tr>
<tr>
<td>ErrorWhiteNoisePlot</td>
<td>White noise test of prediction errors</td>
<td>MODEL</td>
</tr>
</tbody>
</table>
### Table 43.13  continued

<table>
<thead>
<tr>
<th>ODS Table Name</th>
<th>Plot Description</th>
<th>Statement</th>
</tr>
</thead>
<tbody>
<tr>
<td>ErrorPlot</td>
<td>Prediction errors</td>
<td>MODEL</td>
</tr>
<tr>
<td>ModelPlot</td>
<td>Time series and predicted series</td>
<td>MODEL</td>
</tr>
<tr>
<td>AccumulatedIRFPanel</td>
<td>Accumulated impulse response function</td>
<td>MODEL</td>
</tr>
<tr>
<td>AccumulatedIRFXPanel</td>
<td>Accumulated impulse response of transfer function</td>
<td>MODEL</td>
</tr>
<tr>
<td>OrthogonalIRFPanel</td>
<td>Orthogonalized impulse response function</td>
<td>MODEL</td>
</tr>
<tr>
<td>SimpleIRFPanel</td>
<td>Simple impulse response function</td>
<td>MODEL</td>
</tr>
<tr>
<td>SimpleIRFXPanel</td>
<td>Simple impulse response of transfer function</td>
<td>MODEL</td>
</tr>
<tr>
<td>ModelForecastsPlot</td>
<td>Time series and forecasts</td>
<td>OUTPUT</td>
</tr>
<tr>
<td>ForecastsOnlyPlot</td>
<td>Forecasts</td>
<td>OUTPUT</td>
</tr>
</tbody>
</table>

### Computational Issues

#### Computational Method

The VARMAX procedure uses numerous linear algebra routines and frequently uses the sweep operator (Goodnight 1979) and the Cholesky root (Golub and Van Loan 1983).

In addition, the VARMAX procedure uses the nonlinear optimization (NLO) subsystem to perform nonlinear optimization tasks for the maximum likelihood estimation. The optimization requires intensive computation.

#### Convergence Problems

For some data sets, the computation algorithm can fail to converge. Nonconvergence can result from a number of causes, including flat or ridged likelihood surfaces and ill-conditioned data.

If you experience convergence problems, the following points might be helpful:

- Data that contain extreme values can affect results in PROC VARMAX. Rescaling the data can improve stability.
- Changing the TECH=, MAXITER=, and MAXFUNC= options in the NLOPTIONS statement can improve the stability of the optimization process.
- Specifying a different model that might fit the data more closely and might improve convergence.

#### Memory

Let $T$ be the length of each series, $k$ be the number of dependent variables, $p$ be the order of autoregressive terms, and $q$ be the order of moving-average terms. The number of parameters to estimate for a VARMA($p$, $q$) model is

$$k + (p + q)k^2 + k \times (k + 1)/2$$
As \( k \) increases, the number of parameters to estimate increases very quickly. Furthermore, the memory requirement for VARMA\((p,q)\) quadratically increases as \( k \) and \( T \) increase.

For a VARMAX\((p,q,s)\) model and GARCH-type multivariate conditional heteroscedasticity models, the number of parameters to estimate and the memory requirements are considerable.

**Computing Time**

PROC VARMAX is computationally intensive, and execution times can be long. Extensive CPU time is often required to compute the maximum likelihood estimates.

---

**Examples: VARMAX Procedure**

**Example 43.1: Analysis of United States Economic Variables**

Consider the following four-dimensional system of US economic variables. Quarterly data for the years 1954 to 1987 are used (Lütkepohl 1993, Table E.3.).

```sas
title 'Analysis of US Economic Variables';
data us_money;
    date=intnx( 'qtr', '01jan54'd, _n_-1 );
    format date yyq . ;
    input y1 y2 y3 y4 @@;
    y1=log(y1);
    y2=log(y2);
    label y1='log(real money stock M1)'
            y2='log(GNP in bil. of 1982 dollars)'
            y3='Discount rate on 91-day T-bills'
            y4='Yield on 20-year Treasury bonds';
datalines;
450.9 1406.8 0.010800000 0.026133333
453.0 1401.2 0.008133333 0.025233333
459.1 1418.0 0.008700000 0.024900000
... more lines ...
```

The following statements plot the series:

```sas
proc sgplot data=us_money;
    series x=date y=y1 / lineattrs=(pattern=solid);
    series x=date y=y2 / lineattrs=(pattern=dash);
yaxis label="Series";
run;
```
Output 43.1.1 shows the plot of the variables $y_1$ and $y_2$.

The following statements plot the variables $y_3$ and $y_4$:

```sas
proc sgplot data=us_money;
  series x=date y=y3 / lineattrs=(pattern=solid);
  series x=date y=y4 / lineattrs=(pattern=dash);
  yaxis label="Series";
run;
```
Output 43.1.2 shows the plot of the variables $y_3$ and $y_4$.

Output 43.1.2 Plot of Data

The following statements perform the Dickey-Fuller test for stationarity, the Johansen cointegrated test integrated order 2, and the exogeneity test. The VECM(2) is fit to the data.

```sas
proc varmax data=us_money;
  id date interval=qtr;
  model y1-y4 / p=2 lagmax=6 dftest
    print=(iarr(3) estimates diagnose)
    cointtest=(johansen=(iorder=2));
  cointeg rank=1 normalize=y1 exogeneity;
run;
```

From the outputs shown in Output 43.1.5, you can see that the series has unit roots and is cointegrated in
rank 1 with integrated order 1. The fitted VECM(2) is given as

\[
\Delta y_t = \begin{pmatrix}
0.0408 \\
0.0860 \\
0.0052 \\
-0.0144 \\
\end{pmatrix} + \begin{pmatrix}
-0.0140 & 0.0065 & -0.2026 & 0.1306 \\
-0.0281 & 0.0131 & -0.4080 & 0.2630 \\
-0.0022 & 0.0010 & -0.0312 & 0.0201 \\
0.0051 & -0.0024 & 0.0741 & -0.0477 \\
\end{pmatrix} y_{t-1} \\
+ \begin{pmatrix}
0.3460 & 0.0913 & -0.3535 & -0.9690 \\
0.0994 & 0.0379 & 0.2390 & 0.2866 \\
0.1812 & 0.0786 & 0.0223 & 0.4051 \\
0.0322 & 0.0496 & -0.0329 & 0.1857 \\
\end{pmatrix} \Delta y_{t-1} + \epsilon_t
\]

The \( \Delta \) prefixed to a variable name implies differencing.

Output 43.1.3 through Output 43.1.16 show the details. Output 43.1.3 shows the descriptive statistics.

**Output 43.1.3 Descriptive Statistics**

**Analysis of US Economic Variables**

**The VARMAX Procedure**

<table>
<thead>
<tr>
<th>Variable</th>
<th>Type</th>
<th>N</th>
<th>Mean</th>
<th>Standard Deviation</th>
<th>Min</th>
<th>Max</th>
<th>Label</th>
</tr>
</thead>
<tbody>
<tr>
<td>y1</td>
<td>Dependent</td>
<td>136</td>
<td>6.21295</td>
<td>0.07924</td>
<td>6.10278</td>
<td>6.45331</td>
<td>log(real money stock M1)</td>
</tr>
<tr>
<td>y2</td>
<td>Dependent</td>
<td>136</td>
<td>7.77890</td>
<td>0.30110</td>
<td>7.24508</td>
<td>8.27461</td>
<td>log(GNP in bil. of 1982 dollars)</td>
</tr>
<tr>
<td>y3</td>
<td>Dependent</td>
<td>136</td>
<td>0.05608</td>
<td>0.03109</td>
<td>0.00813</td>
<td>0.15087</td>
<td>Discount rate on 91-day T-bills</td>
</tr>
<tr>
<td>y4</td>
<td>Dependent</td>
<td>136</td>
<td>0.06458</td>
<td>0.02927</td>
<td>0.02490</td>
<td>0.13600</td>
<td>Yield on 20-year Treasury bonds</td>
</tr>
</tbody>
</table>

Output 43.1.4 shows the output for Dickey-Fuller tests for the nonstationarity of each series. The null hypothesis is that there exists a unit root. All series have a unit root.

**Output 43.1.4 Unit Root Tests**

<table>
<thead>
<tr>
<th>Variable</th>
<th>Type</th>
<th>Rho</th>
<th>Pr &lt; Rho</th>
<th>Tau</th>
<th>Pr &lt; Tau</th>
</tr>
</thead>
<tbody>
<tr>
<td>y1</td>
<td>Zero Mean</td>
<td>0.05</td>
<td>0.6934</td>
<td>1.14</td>
<td>0.9343</td>
</tr>
<tr>
<td></td>
<td>Single Mean</td>
<td>-2.97</td>
<td>0.6572</td>
<td>-0.76</td>
<td>0.8260</td>
</tr>
<tr>
<td></td>
<td>Trend</td>
<td>-5.91</td>
<td>0.7454</td>
<td>-1.34</td>
<td>0.8725</td>
</tr>
<tr>
<td>y2</td>
<td>Zero Mean</td>
<td>0.13</td>
<td>0.7124</td>
<td>5.14</td>
<td>0.9999</td>
</tr>
<tr>
<td></td>
<td>Single Mean</td>
<td>-0.43</td>
<td>0.9309</td>
<td>-0.79</td>
<td>0.8176</td>
</tr>
<tr>
<td></td>
<td>Trend</td>
<td>-9.21</td>
<td>0.4787</td>
<td>-2.16</td>
<td>0.5063</td>
</tr>
<tr>
<td>y3</td>
<td>Zero Mean</td>
<td>-1.28</td>
<td>0.4255</td>
<td>-0.69</td>
<td>0.4182</td>
</tr>
<tr>
<td></td>
<td>Single Mean</td>
<td>-8.86</td>
<td>0.1700</td>
<td>-2.27</td>
<td>0.1842</td>
</tr>
<tr>
<td></td>
<td>Trend</td>
<td>-18.97</td>
<td>0.0742</td>
<td>-2.86</td>
<td>0.1803</td>
</tr>
<tr>
<td>y4</td>
<td>Zero Mean</td>
<td>0.40</td>
<td>0.7803</td>
<td>0.45</td>
<td>0.8100</td>
</tr>
<tr>
<td></td>
<td>Single Mean</td>
<td>-2.79</td>
<td>0.6790</td>
<td>-1.29</td>
<td>0.6328</td>
</tr>
<tr>
<td></td>
<td>Trend</td>
<td>-12.12</td>
<td>0.2923</td>
<td>-2.33</td>
<td>0.4170</td>
</tr>
</tbody>
</table>
The Johansen cointegration rank test shows whether the series is integrated order either 1 or 2 as shown in Output 43.1.5. The last two columns in Output 43.1.5 explain the cointegration rank test with integrated order 1. The results indicate that there is a cointegrated relationship with cointegration rank 1 with respect to the 0.05 significance level because the test statistic for the null hypothesis H0: $r = 0$ is 55.9633 and its corresponding $p$-value is 0.0072, less than 0.05 (indicating that H0: $r = 0$ should be rejected), and the test statistic for the null hypothesis H0: $r = 1$ is 20.6542 and its corresponding $p$-value is 0.3775, greater than 0.05 (indicating that H0: $r = 1$ cannot be rejected). Now, look at the row associated with $r = 1$. All $p$-values of the tests for the null hypothesis that the series are integrated order 2 are zeros, less than 0.05 significance level (indicating that the null hypothesis should be rejected).

**Output 43.1.5** Cointegration Rank Test

<table>
<thead>
<tr>
<th>$r/k-r-s$</th>
<th>4</th>
<th>3</th>
<th>2</th>
<th>1</th>
<th>Trace of 1</th>
<th>Pr &gt; Trace of 1</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>384.6090</td>
<td>214.3790</td>
<td>107.9378</td>
<td>37.0252</td>
<td>55.9633</td>
<td>0.0072</td>
</tr>
<tr>
<td>Pr &gt; Trace of I(2)</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>219.6239</td>
<td>89.2151</td>
<td>27.3261</td>
<td>20.6542</td>
<td>0.3775</td>
<td></td>
</tr>
<tr>
<td>Pr &gt; Trace of I(2)</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>73.6178</td>
<td>22.1328</td>
<td>2.6477</td>
<td>0.9803</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pr &gt; Trace of I(2)</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>38.2943</td>
<td>0.0149</td>
<td>0.9031</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pr &gt; Trace of I(2)</td>
<td>0.0000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Output 43.1.6** shows the estimates of the long-run parameter, $\beta$, and the adjustment coefficient, $\alpha$.

**Output 43.1.6** Cointegration Rank Test, Continued

<table>
<thead>
<tr>
<th>Variable</th>
<th>Beta</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
</tr>
<tr>
<td>$y_1$</td>
<td>1.00000</td>
</tr>
<tr>
<td>$y_2$</td>
<td>-0.46458</td>
</tr>
<tr>
<td>$y_3$</td>
<td>14.51619</td>
</tr>
<tr>
<td>$y_4$</td>
<td>-9.35520</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Variable</th>
<th>Alpha</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
</tr>
<tr>
<td>$y_1$</td>
<td>-0.01396</td>
</tr>
<tr>
<td>$y_2$</td>
<td>-0.02811</td>
</tr>
<tr>
<td>$y_3$</td>
<td>-0.00215</td>
</tr>
<tr>
<td>$y_4$</td>
<td>0.00510</td>
</tr>
</tbody>
</table>
Output 43.1.7 shows the estimates $\eta$ and $\xi$.

**Output 43.1.7**  Cointegration Rank Test, Continued

<table>
<thead>
<tr>
<th>Variable</th>
<th>Eta 1</th>
<th>Eta 2</th>
<th>Eta 3</th>
<th>Eta 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>y1</td>
<td>52.74907</td>
<td>41.74502</td>
<td>-20.80403</td>
<td>55.77415</td>
</tr>
<tr>
<td>y2</td>
<td>-49.10609</td>
<td>-9.40081</td>
<td>98.87199</td>
<td>22.56416</td>
</tr>
<tr>
<td>y3</td>
<td>68.29674</td>
<td>-144.83173</td>
<td>-27.35953</td>
<td>15.51142</td>
</tr>
<tr>
<td>y4</td>
<td>121.25932</td>
<td>271.80496</td>
<td>85.85156</td>
<td>-130.11599</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Variable</th>
<th>Xi 1</th>
<th>Xi 2</th>
<th>Xi 3</th>
<th>Xi 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>y1</td>
<td>-0.00842</td>
<td>-0.00052</td>
<td>-0.00208</td>
<td>-0.00250</td>
</tr>
<tr>
<td>y2</td>
<td>0.00141</td>
<td>0.00213</td>
<td>-0.00736</td>
<td>-0.00058</td>
</tr>
<tr>
<td>y3</td>
<td>-0.00445</td>
<td>0.00541</td>
<td>-0.00150</td>
<td>0.00310</td>
</tr>
<tr>
<td>y4</td>
<td>-0.00211</td>
<td>-0.00064</td>
<td>-0.00130</td>
<td>0.00197</td>
</tr>
</tbody>
</table>

Output 43.1.8 shows that the VECM(2) is fit to the data. The RANK=1 option in the COINTEG statement produces the estimates of the long-run parameter, $\beta$, and the adjustment coefficient, $\alpha$.

**Output 43.1.8**  Parameter Estimates

**Analysis of US Economic Variables**

**The VARMAX Procedure**

<table>
<thead>
<tr>
<th>Type of Model</th>
<th>VECM(2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Estimation Method</td>
<td>Maximum Likelihood Estimation</td>
</tr>
<tr>
<td>Cointegrated Rank</td>
<td>1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Beta 1</th>
<th>Variable</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.00000</td>
<td>y1</td>
</tr>
<tr>
<td>-0.46458</td>
<td>y2</td>
</tr>
<tr>
<td>14.51619</td>
<td>y3</td>
</tr>
<tr>
<td>-9.35520</td>
<td>y4</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Alpha 1</th>
<th>Variable</th>
</tr>
</thead>
<tbody>
<tr>
<td>-0.01396</td>
<td>y1</td>
</tr>
<tr>
<td>-0.02811</td>
<td>y2</td>
</tr>
<tr>
<td>-0.00215</td>
<td>y3</td>
</tr>
<tr>
<td>0.00510</td>
<td>y4</td>
</tr>
</tbody>
</table>
Output 43.1.9 shows the parameter estimates in terms of the constant, the lag 1 coefficients \((y_{t-1})\) that are contained in the \(\alpha\beta'\) estimates, and the coefficients that are associated with the lag 1 first differences \((\Delta y_{t-1})\).

**Output 43.1.9** Parameter Estimates, Continued

<table>
<thead>
<tr>
<th>Parameter Alpha * Beta' Estimates</th>
</tr>
</thead>
<tbody>
<tr>
<td>Variable</td>
</tr>
<tr>
<td>----------</td>
</tr>
<tr>
<td>(y_1)</td>
</tr>
<tr>
<td>(y_2)</td>
</tr>
<tr>
<td>(y_3)</td>
</tr>
<tr>
<td>(y_4)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>AR Coefficients of Differenced Lag</th>
</tr>
</thead>
<tbody>
<tr>
<td>DIF Lag</td>
</tr>
<tr>
<td>---------</td>
</tr>
<tr>
<td>1</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>
Output 43.1.10 through Output 43.1.12 show the parameter estimates and their significance.

### Output 43.1.10 Parameter Estimates, Continued

<table>
<thead>
<tr>
<th>Equation</th>
<th>Parameter</th>
<th>Estimate</th>
<th>Standard Error</th>
<th>t Value</th>
<th>Pr &gt;</th>
<th>Variable</th>
</tr>
</thead>
<tbody>
<tr>
<td>D_y1</td>
<td>CONST1</td>
<td>0.04076</td>
<td>0.01418</td>
<td>2.87</td>
<td>0.0048</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>AR1_1_1</td>
<td>-0.01396</td>
<td>0.00495</td>
<td>-2.82</td>
<td>0.0056</td>
<td>y1(t-1)</td>
</tr>
<tr>
<td></td>
<td>AR1_1_2</td>
<td>0.00648</td>
<td>0.00230</td>
<td>2.82</td>
<td>0.0056</td>
<td>y2(t-1)</td>
</tr>
<tr>
<td></td>
<td>AR1_1_3</td>
<td>-0.20263</td>
<td>0.07191</td>
<td>-2.82</td>
<td>0.0056</td>
<td>y3(t-1)</td>
</tr>
<tr>
<td></td>
<td>AR1_1_4</td>
<td>0.13059</td>
<td>0.04634</td>
<td>2.82</td>
<td>0.0056</td>
<td>y4(t-1)</td>
</tr>
<tr>
<td></td>
<td>AR2_1_1</td>
<td>0.34603</td>
<td>0.06414</td>
<td>5.39</td>
<td>&lt;0.001</td>
<td>D_y1(t-1)</td>
</tr>
<tr>
<td></td>
<td>AR2_1_2</td>
<td>0.09131</td>
<td>0.07334</td>
<td>1.25</td>
<td>0.2154</td>
<td>D_y2(t-1)</td>
</tr>
<tr>
<td></td>
<td>AR2_1_3</td>
<td>-0.35351</td>
<td>0.11024</td>
<td>-3.21</td>
<td>0.0017</td>
<td>D_y3(t-1)</td>
</tr>
<tr>
<td></td>
<td>AR2_1_4</td>
<td>-0.96895</td>
<td>0.20737</td>
<td>-4.67</td>
<td>&lt;0.001</td>
<td>D_y4(t-1)</td>
</tr>
<tr>
<td>D_y2</td>
<td>CONST2</td>
<td>0.08595</td>
<td>0.01679</td>
<td>5.12</td>
<td>&lt;0.001</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>AR1_2_1</td>
<td>-0.02811</td>
<td>0.00586</td>
<td>-4.79</td>
<td>&lt;0.001</td>
<td>y1(t-1)</td>
</tr>
<tr>
<td></td>
<td>AR1_2_2</td>
<td>0.01306</td>
<td>0.00272</td>
<td>4.79</td>
<td>&lt;0.001</td>
<td>y2(t-1)</td>
</tr>
<tr>
<td></td>
<td>AR1_2_3</td>
<td>-0.40799</td>
<td>0.08514</td>
<td>-4.79</td>
<td>&lt;0.001</td>
<td>y3(t-1)</td>
</tr>
<tr>
<td></td>
<td>AR1_2_4</td>
<td>0.26294</td>
<td>0.05487</td>
<td>4.79</td>
<td>&lt;0.001</td>
<td>y4(t-1)</td>
</tr>
<tr>
<td></td>
<td>AR2_2_1</td>
<td>0.09936</td>
<td>0.07594</td>
<td>1.31</td>
<td>0.1932</td>
<td>D_y1(t-1)</td>
</tr>
<tr>
<td></td>
<td>AR2_2_2</td>
<td>0.03791</td>
<td>0.08683</td>
<td>0.44</td>
<td>0.6632</td>
<td>D_y2(t-1)</td>
</tr>
<tr>
<td></td>
<td>AR2_2_3</td>
<td>0.23900</td>
<td>0.13052</td>
<td>1.83</td>
<td>0.0695</td>
<td>D_y3(t-1)</td>
</tr>
<tr>
<td></td>
<td>AR2_2_4</td>
<td>0.28661</td>
<td>0.24552</td>
<td>1.17</td>
<td>0.2453</td>
<td>D_y4(t-1)</td>
</tr>
<tr>
<td>D_y3</td>
<td>CONST3</td>
<td>0.00518</td>
<td>0.01608</td>
<td>0.32</td>
<td>0.7476</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>AR1_3_1</td>
<td>-0.00215</td>
<td>0.00562</td>
<td>-0.38</td>
<td>0.7024</td>
<td>y1(t-1)</td>
</tr>
<tr>
<td></td>
<td>AR1_3_2</td>
<td>0.00100</td>
<td>0.00261</td>
<td>0.38</td>
<td>0.7024</td>
<td>y2(t-1)</td>
</tr>
<tr>
<td></td>
<td>AR1_3_3</td>
<td>-0.03121</td>
<td>0.08151</td>
<td>-0.38</td>
<td>0.7024</td>
<td>y3(t-1)</td>
</tr>
<tr>
<td></td>
<td>AR1_3_4</td>
<td>0.02011</td>
<td>0.05253</td>
<td>0.38</td>
<td>0.7024</td>
<td>y4(t-1)</td>
</tr>
<tr>
<td></td>
<td>AR2_3_1</td>
<td>0.18118</td>
<td>0.07271</td>
<td>2.49</td>
<td>0.0140</td>
<td>D_y1(t-1)</td>
</tr>
<tr>
<td></td>
<td>AR2_3_2</td>
<td>0.07859</td>
<td>0.08313</td>
<td>0.95</td>
<td>0.3463</td>
<td>D_y2(t-1)</td>
</tr>
<tr>
<td></td>
<td>AR2_3_3</td>
<td>0.02234</td>
<td>0.12496</td>
<td>0.18</td>
<td>0.8584</td>
<td>D_y3(t-1)</td>
</tr>
<tr>
<td></td>
<td>AR2_3_4</td>
<td>0.40508</td>
<td>0.23506</td>
<td>1.72</td>
<td>0.0873</td>
<td>D_y4(t-1)</td>
</tr>
<tr>
<td>D_y4</td>
<td>CONST4</td>
<td>-0.01438</td>
<td>0.00803</td>
<td>-1.79</td>
<td>0.0758</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>AR1_4_1</td>
<td>0.00510</td>
<td>0.00281</td>
<td>1.82</td>
<td>0.0713</td>
<td>y1(t-1)</td>
</tr>
<tr>
<td></td>
<td>AR1_4_2</td>
<td>-0.00237</td>
<td>0.00130</td>
<td>-1.82</td>
<td>0.0713</td>
<td>y2(t-1)</td>
</tr>
<tr>
<td></td>
<td>AR1_4_3</td>
<td>0.07407</td>
<td>0.04072</td>
<td>1.82</td>
<td>0.0713</td>
<td>y3(t-1)</td>
</tr>
<tr>
<td></td>
<td>AR1_4_4</td>
<td>-0.04774</td>
<td>0.02624</td>
<td>-1.82</td>
<td>0.0713</td>
<td>y4(t-1)</td>
</tr>
<tr>
<td></td>
<td>AR2_4_1</td>
<td>0.03222</td>
<td>0.03632</td>
<td>0.89</td>
<td>0.3768</td>
<td>D_y1(t-1)</td>
</tr>
<tr>
<td></td>
<td>AR2_4_2</td>
<td>0.04961</td>
<td>0.04153</td>
<td>1.19</td>
<td>0.2345</td>
<td>D_y2(t-1)</td>
</tr>
<tr>
<td></td>
<td>AR2_4_3</td>
<td>-0.03292</td>
<td>0.06243</td>
<td>-0.53</td>
<td>0.5990</td>
<td>D_y3(t-1)</td>
</tr>
<tr>
<td></td>
<td>AR2_4_4</td>
<td>0.18568</td>
<td>0.11744</td>
<td>1.58</td>
<td>0.1164</td>
<td>D_y4(t-1)</td>
</tr>
</tbody>
</table>
**Example 43.1: Analysis of United States Economic Variables**

**Output 43.1.11** Parameter Estimates, Continued

<table>
<thead>
<tr>
<th>Equation</th>
<th>Parameter</th>
<th>Estimate</th>
<th>Standard Error</th>
<th>t Value</th>
<th>Pr &gt;</th>
<th>Variable</th>
</tr>
</thead>
<tbody>
<tr>
<td>D_y1</td>
<td>ALPHA1_1</td>
<td>-0.01396</td>
<td>0.00495</td>
<td>-2.82</td>
<td>0.0056</td>
<td>beta1^* dep(t-1)</td>
</tr>
<tr>
<td></td>
<td>BETA1_1</td>
<td>1.00000</td>
<td></td>
<td></td>
<td></td>
<td>y1(t-1)</td>
</tr>
<tr>
<td>D_y2</td>
<td>ALPHA2_1</td>
<td>-0.02811</td>
<td>0.00586</td>
<td>-4.79</td>
<td>&lt;.0001</td>
<td>beta1^* dep(t-1)</td>
</tr>
<tr>
<td></td>
<td>BETA2_1</td>
<td>-0.46458</td>
<td></td>
<td></td>
<td></td>
<td>y2(t-1)</td>
</tr>
<tr>
<td>D_y3</td>
<td>ALPHA3_1</td>
<td>-0.00215</td>
<td>0.00562</td>
<td>-0.38</td>
<td>0.7024</td>
<td>beta1^* dep(t-1)</td>
</tr>
<tr>
<td></td>
<td>BETA3_1</td>
<td>14.51619</td>
<td></td>
<td></td>
<td></td>
<td>y3(t-1)</td>
</tr>
<tr>
<td>D_y4</td>
<td>ALPHA4_1</td>
<td>0.00510</td>
<td>0.00281</td>
<td>1.82</td>
<td>0.0713</td>
<td>beta1^* dep(t-1)</td>
</tr>
<tr>
<td></td>
<td>BETA4_1</td>
<td>-9.35520</td>
<td></td>
<td></td>
<td></td>
<td>y4(t-1)</td>
</tr>
</tbody>
</table>

**Output 43.1.12** Parameter Estimates, Continued

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Estimate</th>
<th>Standard Error</th>
<th>t Value</th>
<th>Pr &gt;</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>COV1_1</td>
<td>0.00005</td>
<td>0.00001</td>
<td>8.19</td>
<td>&lt;.0001</td>
<td></td>
</tr>
<tr>
<td>COV1_2</td>
<td>0.00001</td>
<td>0.00001</td>
<td>2.78</td>
<td>0.0062</td>
<td></td>
</tr>
<tr>
<td>COV2_2</td>
<td>0.00007</td>
<td>0.00001</td>
<td>8.19</td>
<td>&lt;.0001</td>
<td></td>
</tr>
<tr>
<td>COV1_3</td>
<td>-0.00001</td>
<td>0.00001</td>
<td>-1.60</td>
<td>0.1118</td>
<td></td>
</tr>
<tr>
<td>COV2_3</td>
<td>0.00002</td>
<td>0.00001</td>
<td>2.71</td>
<td>0.0077</td>
<td></td>
</tr>
<tr>
<td>COV3_3</td>
<td>0.00007</td>
<td>0.00001</td>
<td>8.19</td>
<td>&lt;.0001</td>
<td></td>
</tr>
<tr>
<td>COV1_4</td>
<td>-0.00000</td>
<td>0.00000</td>
<td>-1.31</td>
<td>0.1936</td>
<td></td>
</tr>
<tr>
<td>COV2_4</td>
<td>0.00001</td>
<td>0.00000</td>
<td>3.29</td>
<td>0.0013</td>
<td></td>
</tr>
<tr>
<td>COV3_4</td>
<td>0.00002</td>
<td>0.00000</td>
<td>6.67</td>
<td>&lt;.0001</td>
<td></td>
</tr>
<tr>
<td>COV4_4</td>
<td>0.00002</td>
<td>0.00000</td>
<td>8.19</td>
<td>&lt;.0001</td>
<td></td>
</tr>
</tbody>
</table>
Output 43.1.13 shows the innovation covariance matrix estimates, the log-likelihood, the various information criteria results, and the tests for white noise residuals. According to the portmanteau test results, the residuals have significant correlations at lag 2 and 3, indicating that a VECM(3) model might be a better fit than the VECM(2) model.

### Output 43.1.13 Diagnostic Checks

**Covariances of Innovations**

<table>
<thead>
<tr>
<th>Variable</th>
<th>y1</th>
<th>y2</th>
<th>y3</th>
<th>y4</th>
</tr>
</thead>
<tbody>
<tr>
<td>y1</td>
<td>0.00005</td>
<td>0.00001</td>
<td>-0.00001</td>
<td>-0.00000</td>
</tr>
<tr>
<td>y2</td>
<td>0.00001</td>
<td>0.00007</td>
<td>0.00002</td>
<td>0.00001</td>
</tr>
<tr>
<td>y3</td>
<td>-0.00001</td>
<td>0.00002</td>
<td>0.00007</td>
<td>0.00002</td>
</tr>
<tr>
<td>y4</td>
<td>-0.00000</td>
<td>0.00001</td>
<td>0.00002</td>
<td>0.00002</td>
</tr>
</tbody>
</table>

**Log-likelihood** 2479.23

**Information Criteria**

<table>
<thead>
<tr>
<th>Criteria</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>AICC</td>
<td>-4859</td>
</tr>
<tr>
<td>HQC</td>
<td>-4844.07</td>
</tr>
<tr>
<td>AIC</td>
<td>-4886.46</td>
</tr>
<tr>
<td>SBC</td>
<td>-4782.14</td>
</tr>
<tr>
<td>FPEC</td>
<td>2.23E-18</td>
</tr>
</tbody>
</table>

**Schematic Representation of Cross Correlations of Residuals**

<table>
<thead>
<tr>
<th>Variable/Lag</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>y1</td>
<td>++..</td>
<td>. . . . . .</td>
<td>++..</td>
<td>. . . . . .</td>
<td>. .</td>
<td>. . . . . .</td>
<td>. .</td>
</tr>
<tr>
<td>y2</td>
<td>++++</td>
<td>. . . . . .</td>
<td>. .</td>
<td>. . . . . .</td>
<td>. .</td>
<td>. . . . . .</td>
<td>. .</td>
</tr>
<tr>
<td>y3</td>
<td>.+++</td>
<td>. . . . . .</td>
<td>. .</td>
<td>. . . . . .</td>
<td>. .</td>
<td>. . . . . .</td>
<td>. .</td>
</tr>
<tr>
<td>y4</td>
<td>.+++</td>
<td>. . . . . .</td>
<td>. .</td>
<td>. . . . . .</td>
<td>. .</td>
<td>. . . . . .</td>
<td>. .</td>
</tr>
</tbody>
</table>

* + is > 2*std error, - is < -2*std error, . is between*

**Portmanteau Test for Cross Correlations of Residuals**

<table>
<thead>
<tr>
<th>Up To Lag</th>
<th>DF</th>
<th>Chi-Square</th>
<th>Pr &gt; ChiSq</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>16</td>
<td>53.90</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>4</td>
<td>32</td>
<td>74.03</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>5</td>
<td>48</td>
<td>103.08</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>6</td>
<td>64</td>
<td>116.94</td>
<td>&lt;.0001</td>
</tr>
</tbody>
</table>
Example 43.1: Analysis of United States Economic Variables

Output 43.1.14 describes how well each univariate equation fits the data. The residuals for $y_3$ and $y_4$ differ from normality. Except for the residuals for $y_3$, there are no AR effects on other residuals. Except for the residuals for $y_4$, there are no ARCH effects on other residuals.

### Output 43.1.14  Diagnostic Checks, Continued

<table>
<thead>
<tr>
<th>Variable</th>
<th>R-Square</th>
<th>Standard Deviation</th>
<th>F Value</th>
<th>Pr &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>$y_1$</td>
<td>0.6754</td>
<td>0.00712</td>
<td>32.51</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>$y_2$</td>
<td>0.3070</td>
<td>0.00843</td>
<td>6.92</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>$y_3$</td>
<td>0.1328</td>
<td>0.00807</td>
<td>2.39</td>
<td>0.0196</td>
</tr>
<tr>
<td>$y_4$</td>
<td>0.0831</td>
<td>0.00403</td>
<td>1.42</td>
<td>0.1963</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Variable</th>
<th>Durbin Watson</th>
<th>Chi-Square Pr &gt; ChiSq</th>
<th>F Value</th>
<th>Pr &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>$y_1$</td>
<td>2.13418</td>
<td>7.19</td>
<td>0.0275</td>
<td>1.62</td>
</tr>
<tr>
<td>$y_2$</td>
<td>2.04003</td>
<td>1.20</td>
<td>0.5483</td>
<td>1.23</td>
</tr>
<tr>
<td>$y_3$</td>
<td>1.86892</td>
<td>253.76</td>
<td>&lt;.0001</td>
<td>1.78</td>
</tr>
<tr>
<td>$y_4$</td>
<td>1.98440</td>
<td>105.21</td>
<td>&lt;.0001</td>
<td>21.01</td>
</tr>
</tbody>
</table>

### Output 43.1.15  Infinite Order AR Representation

<table>
<thead>
<tr>
<th>Lag</th>
<th>Variable</th>
<th>$y_1$</th>
<th>$y_2$</th>
<th>$y_3$</th>
<th>$y_4$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>$y_1$</td>
<td>1.33208</td>
<td>0.09780</td>
<td>-0.55614</td>
<td>-0.83836</td>
</tr>
<tr>
<td></td>
<td>$y_2$</td>
<td>0.07125</td>
<td>1.05096</td>
<td>-0.16899</td>
<td>0.54955</td>
</tr>
<tr>
<td></td>
<td>$y_3$</td>
<td>0.17903</td>
<td>0.07859</td>
<td>0.99113</td>
<td>0.42520</td>
</tr>
<tr>
<td></td>
<td>$y_4$</td>
<td>0.03732</td>
<td>0.04724</td>
<td>0.04116</td>
<td>1.13795</td>
</tr>
<tr>
<td>2</td>
<td>$y_1$</td>
<td>-0.34603</td>
<td>-0.09131</td>
<td>0.35351</td>
<td>0.96895</td>
</tr>
<tr>
<td></td>
<td>$y_2$</td>
<td>-0.09936</td>
<td>-0.03791</td>
<td>-0.23900</td>
<td>-0.28661</td>
</tr>
<tr>
<td></td>
<td>$y_3$</td>
<td>-0.18118</td>
<td>-0.07859</td>
<td>-0.02234</td>
<td>-0.40508</td>
</tr>
<tr>
<td></td>
<td>$y_4$</td>
<td>-0.03222</td>
<td>-0.04961</td>
<td>0.03292</td>
<td>-0.18568</td>
</tr>
<tr>
<td>3</td>
<td>$y_1$</td>
<td>0.00000</td>
<td>0.00000</td>
<td>0.00000</td>
<td>0.00000</td>
</tr>
<tr>
<td></td>
<td>$y_2$</td>
<td>0.00000</td>
<td>0.00000</td>
<td>0.00000</td>
<td>0.00000</td>
</tr>
<tr>
<td></td>
<td>$y_3$</td>
<td>0.00000</td>
<td>0.00000</td>
<td>0.00000</td>
<td>0.00000</td>
</tr>
<tr>
<td></td>
<td>$y_4$</td>
<td>0.00000</td>
<td>0.00000</td>
<td>0.00000</td>
<td>0.00000</td>
</tr>
</tbody>
</table>

The PRINT=(IARR) option provides the VAR(2) representation in Output 43.1.15.
Output 43.1.16 shows whether each variable is the weak exogeneity of other variables. The variable \( y_1 \) is not the weak exogeneity of other variables, \( y_2, y_3, \) and \( y_4 \); the variable \( y_2 \) is not the weak exogeneity of other variables, \( y_1, y_3, \) and \( y_4 \); the variables \( y_3 \) and \( y_4 \) are the weak exogeneity of other variables.

### Output 43.1.16 Weak Exogeneity Test

<table>
<thead>
<tr>
<th>Variable</th>
<th>DF</th>
<th>Chi-Square</th>
<th>Pr &gt; ChiSq</th>
</tr>
</thead>
<tbody>
<tr>
<td>( y_1 )</td>
<td>1</td>
<td>6.55</td>
<td>0.0105</td>
</tr>
<tr>
<td>( y_2 )</td>
<td>1</td>
<td>12.54</td>
<td>0.0004</td>
</tr>
<tr>
<td>( y_3 )</td>
<td>1</td>
<td>0.09</td>
<td>0.7695</td>
</tr>
<tr>
<td>( y_4 )</td>
<td>1</td>
<td>1.81</td>
<td>0.1786</td>
</tr>
</tbody>
</table>

### Example 43.2: Analysis of German Economic Variables

This example considers a three-dimensional VAR(2) model. The model contains the logarithms of a quarterly, seasonally adjusted West German fixed investment, disposable income, and consumption expenditures. The data used are in Lütkepohl (1993, Table E.1).

```sas
proc varmax data=west; id date interval=qtr; model y1-y3 / p=2 dify=(1) print=(decompose(6) impulse=(stderr) estimates diagnose) printform=both lagmax=3; causal group1=(y1) group2=(y2 y3); output lead=5; run;
```

```sas
title 'Analysis of German Economic Variables';
data west;
  date = intnx('qtr', '01jan60'd, _n_-1);
  format date yyq. ;
  input y1 y2 y3 @@;
  y1 = log(y1);
  y2 = log(y2);
  y3 = log(y3);
  label y1 = 'logarithm of investment'
        y2 = 'logarithm of income'
        y3 = 'logarithm of consumption';
datalines;
180 451 415 179 179 465 421 185 485 434 192 493 448
211 509 459 202 520 458 207 521 479 214 540 487
... more lines ...
data use;
  set west;
  where date < '01jan79'd;
  keep date y1 y2 y3;
run;
```

First, the differenced data are modeled as a VAR(2) with the following result:

\[
\Delta y_t = \begin{pmatrix} -0.01672 \\ 0.01577 \\ 0.01293 \end{pmatrix} + \begin{pmatrix} -0.31963 & 0.14599 & 0.96122 \\ 0.04393 & -0.15273 & 0.28850 \\ -0.00242 & 0.22481 & -0.26397 \end{pmatrix} \Delta y_{t-1} \\
+ \begin{pmatrix} -0.16055 & 0.11460 & 0.93439 \\ 0.05003 & 0.01917 & -0.01020 \\ 0.03388 & 0.35491 & -0.02223 \end{pmatrix} \Delta y_{t-2} + \epsilon_t
\]

The parameter estimates AR1_{1_2}, AR1_{1_3}, AR2_{1_2}, and AR2_{1_3} are insignificant, and the VARX model is fitted in the next step.

The detailed output is shown in Output 43.2.1 through Output 43.2.8.

Output 43.2.1 shows the descriptive statistics.

### Output 43.2.1 Descriptive Statistics

#### Analysis of German Economic Variables

**The VARMAX Procedure**

<table>
<thead>
<tr>
<th>Number of Observations</th>
<th>75</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of Pairwise Missing</td>
<td>0</td>
</tr>
<tr>
<td>Observation(s) eliminated by differencing</td>
<td>1</td>
</tr>
</tbody>
</table>

#### Simple Summary Statistics

<table>
<thead>
<tr>
<th>Variable</th>
<th>Type</th>
<th>N</th>
<th>Mean</th>
<th>Standard Deviation</th>
<th>Min</th>
<th>Max</th>
<th>Difference</th>
<th>Label</th>
</tr>
</thead>
<tbody>
<tr>
<td>y1</td>
<td>Dependent</td>
<td>75</td>
<td>0.01811</td>
<td>0.04680</td>
<td>-0.14018</td>
<td>0.19358</td>
<td>1</td>
<td>logarithm of investment</td>
</tr>
<tr>
<td>y2</td>
<td>Dependent</td>
<td>75</td>
<td>0.02071</td>
<td>0.01208</td>
<td>-0.02886</td>
<td>0.05023</td>
<td>1</td>
<td>logarithm of income</td>
</tr>
<tr>
<td>y3</td>
<td>Dependent</td>
<td>75</td>
<td>0.01987</td>
<td>0.01040</td>
<td>-0.01300</td>
<td>0.04483</td>
<td>1</td>
<td>logarithm of consumption</td>
</tr>
</tbody>
</table>
Output 43.2.2 shows that a VAR(2) model is fit to the data.

Output 43.2.2 Parameter Estimates

Analysis of German Economic Variables

The VARMAX Procedure

<table>
<thead>
<tr>
<th>Type of Model</th>
<th>VAR(2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Estimation Method</td>
<td>Least Squares Estimation</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Variable</th>
<th>Constant</th>
</tr>
</thead>
<tbody>
<tr>
<td>y1</td>
<td>-0.01672</td>
</tr>
<tr>
<td>y2</td>
<td>0.01577</td>
</tr>
<tr>
<td>y3</td>
<td>0.01293</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Lag</th>
<th>Variable</th>
<th>y1</th>
<th>y2</th>
<th>y3</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>y1</td>
<td>-0.31963</td>
<td>0.14599</td>
<td>0.96122</td>
</tr>
<tr>
<td></td>
<td>y2</td>
<td>0.04393</td>
<td>-0.15273</td>
<td>0.28850</td>
</tr>
<tr>
<td></td>
<td>y3</td>
<td>-0.00242</td>
<td>0.22481</td>
<td>-0.26397</td>
</tr>
<tr>
<td>2</td>
<td>y1</td>
<td>-0.16055</td>
<td>0.11460</td>
<td>0.93439</td>
</tr>
<tr>
<td></td>
<td>y2</td>
<td>0.05003</td>
<td>0.01917</td>
<td>-0.01020</td>
</tr>
<tr>
<td></td>
<td>y3</td>
<td>0.03388</td>
<td>0.35491</td>
<td>-0.02223</td>
</tr>
</tbody>
</table>
Output 43.2.3 shows the parameter estimates and their significance.

**Output 43.2.3** Parameter Estimates, Continued

<table>
<thead>
<tr>
<th>Variable/Lag</th>
<th>Schematic Representation</th>
<th>C</th>
<th>AR1</th>
<th>AR2</th>
</tr>
</thead>
<tbody>
<tr>
<td>y1</td>
<td></td>
<td></td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>y2</td>
<td>+</td>
<td></td>
<td>...</td>
<td></td>
</tr>
<tr>
<td>y3</td>
<td>+</td>
<td></td>
<td>+.</td>
<td>+.</td>
</tr>
</tbody>
</table>

* + is > 2*std error, - is < -2*std error, . is between, * is N/A

**Model Parameter Estimates**

<table>
<thead>
<tr>
<th>Equation</th>
<th>Parameter</th>
<th>Estimate</th>
<th>Standard Error</th>
<th>t Value</th>
<th>Pr &gt;</th>
<th>Variable</th>
</tr>
</thead>
<tbody>
<tr>
<td>y1</td>
<td>CONST1</td>
<td>-0.01672</td>
<td>0.01723</td>
<td>-0.97</td>
<td>0.3352</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>AR1_1_1</td>
<td>-0.31963</td>
<td>0.12546</td>
<td>-2.55</td>
<td>0.0132</td>
<td>y1(t-1)</td>
</tr>
<tr>
<td></td>
<td>AR1_1_2</td>
<td>0.14599</td>
<td>0.54567</td>
<td>0.27</td>
<td>0.7899</td>
<td>y2(t-1)</td>
</tr>
<tr>
<td></td>
<td>AR1_1_3</td>
<td>0.96122</td>
<td>0.66431</td>
<td>1.45</td>
<td>0.1526</td>
<td>y3(t-1)</td>
</tr>
<tr>
<td></td>
<td>AR2_1_1</td>
<td>-0.16055</td>
<td>0.12491</td>
<td>-1.29</td>
<td>0.2032</td>
<td>y1(t-2)</td>
</tr>
<tr>
<td></td>
<td>AR2_1_2</td>
<td>0.11460</td>
<td>0.53457</td>
<td>0.21</td>
<td>0.8309</td>
<td>y2(t-2)</td>
</tr>
<tr>
<td></td>
<td>AR2_1_3</td>
<td>0.93439</td>
<td>0.66510</td>
<td>1.40</td>
<td>0.1647</td>
<td>y3(t-2)</td>
</tr>
<tr>
<td>y2</td>
<td>CONST2</td>
<td>0.01577</td>
<td>0.00437</td>
<td>3.60</td>
<td>0.0006</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>AR1_2_1</td>
<td>0.04393</td>
<td>0.03186</td>
<td>1.38</td>
<td>0.1726</td>
<td>y1(t-1)</td>
</tr>
<tr>
<td></td>
<td>AR1_2_2</td>
<td>-0.15273</td>
<td>0.13857</td>
<td>-1.10</td>
<td>0.2744</td>
<td>y2(t-1)</td>
</tr>
<tr>
<td></td>
<td>AR1_2_3</td>
<td>0.28850</td>
<td>0.16870</td>
<td>1.71</td>
<td>0.0919</td>
<td>y3(t-1)</td>
</tr>
<tr>
<td></td>
<td>AR2_2_1</td>
<td>0.05003</td>
<td>0.03172</td>
<td>1.58</td>
<td>0.1195</td>
<td>y1(t-2)</td>
</tr>
<tr>
<td></td>
<td>AR2_2_2</td>
<td>0.01917</td>
<td>0.13575</td>
<td>0.14</td>
<td>0.8882</td>
<td>y2(t-2)</td>
</tr>
<tr>
<td></td>
<td>AR2_2_3</td>
<td>-0.01020</td>
<td>0.16890</td>
<td>-0.06</td>
<td>0.9520</td>
<td>y3(t-2)</td>
</tr>
<tr>
<td>y3</td>
<td>CONST3</td>
<td>0.01293</td>
<td>0.00353</td>
<td>3.67</td>
<td>0.0005</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>AR1_3_1</td>
<td>-0.00242</td>
<td>0.02568</td>
<td>-0.09</td>
<td>0.9251</td>
<td>y1(t-1)</td>
</tr>
<tr>
<td></td>
<td>AR1_3_2</td>
<td>0.22481</td>
<td>0.11168</td>
<td>2.01</td>
<td>0.0482</td>
<td>y2(t-1)</td>
</tr>
<tr>
<td></td>
<td>AR1_3_3</td>
<td>-0.26397</td>
<td>0.13596</td>
<td>-1.94</td>
<td>0.0565</td>
<td>y3(t-1)</td>
</tr>
<tr>
<td></td>
<td>AR2_3_1</td>
<td>0.03388</td>
<td>0.02556</td>
<td>1.33</td>
<td>0.1896</td>
<td>y1(t-2)</td>
</tr>
<tr>
<td></td>
<td>AR2_3_2</td>
<td>0.35491</td>
<td>0.10941</td>
<td>3.24</td>
<td>0.0019</td>
<td>y2(t-2)</td>
</tr>
<tr>
<td></td>
<td>AR2_3_3</td>
<td>-0.02223</td>
<td>0.13612</td>
<td>-0.16</td>
<td>0.8708</td>
<td>y3(t-2)</td>
</tr>
</tbody>
</table>
Output 43.2.4 shows the innovation covariance matrix estimates, the various information criteria results, and the tests for white noise residuals. The residuals are uncorrelated except at lag 3 for $y_2$ variable.

### Output 43.2.4 Diagnostic Checks

#### Covariances of Innovations

<table>
<thead>
<tr>
<th>Variable</th>
<th>$y_1$</th>
<th>$y_2$</th>
<th>$y_3$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$y_1$</td>
<td>0.00213</td>
<td>0.00007</td>
<td>0.00012</td>
</tr>
<tr>
<td>$y_2$</td>
<td>0.00007</td>
<td>0.00014</td>
<td>0.00006</td>
</tr>
<tr>
<td>$y_3$</td>
<td>0.00012</td>
<td>0.00006</td>
<td>0.00009</td>
</tr>
</tbody>
</table>

#### Information Criteria

<table>
<thead>
<tr>
<th>Criteria</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>AICC</td>
<td>-1527.51</td>
</tr>
<tr>
<td>HQC</td>
<td>-1536.46</td>
</tr>
<tr>
<td>AIC</td>
<td>-1561.11</td>
</tr>
<tr>
<td>SBC</td>
<td>-1499.27</td>
</tr>
<tr>
<td>FPEC</td>
<td>2.18E-11</td>
</tr>
</tbody>
</table>

#### Cross Correlations of Residuals

<table>
<thead>
<tr>
<th>Lag</th>
<th>Variable</th>
<th>$y_1$</th>
<th>$y_2$</th>
<th>$y_3$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>$y_1$</td>
<td>1.00000</td>
<td>0.13242</td>
<td>0.28275</td>
</tr>
<tr>
<td></td>
<td>$y_2$</td>
<td>0.13242</td>
<td>1.00000</td>
<td>0.55526</td>
</tr>
<tr>
<td></td>
<td>$y_3$</td>
<td>0.28275</td>
<td>0.55526</td>
<td>1.00000</td>
</tr>
<tr>
<td>1</td>
<td>$y_1$</td>
<td>0.01461</td>
<td>-0.00666</td>
<td>-0.02394</td>
</tr>
<tr>
<td></td>
<td>$y_2$</td>
<td>-0.01125</td>
<td>-0.00167</td>
<td>-0.04515</td>
</tr>
<tr>
<td></td>
<td>$y_3$</td>
<td>-0.00993</td>
<td>-0.06780</td>
<td>-0.09593</td>
</tr>
<tr>
<td>2</td>
<td>$y_1$</td>
<td>0.07253</td>
<td>-0.00226</td>
<td>-0.01621</td>
</tr>
<tr>
<td></td>
<td>$y_2$</td>
<td>-0.08096</td>
<td>-0.01066</td>
<td>-0.02047</td>
</tr>
<tr>
<td></td>
<td>$y_3$</td>
<td>-0.02660</td>
<td>-0.01392</td>
<td>-0.02263</td>
</tr>
<tr>
<td>3</td>
<td>$y_1$</td>
<td>0.09915</td>
<td>0.04484</td>
<td>0.05243</td>
</tr>
<tr>
<td></td>
<td>$y_2$</td>
<td>-0.00289</td>
<td>0.14059</td>
<td>0.25984</td>
</tr>
<tr>
<td></td>
<td>$y_3$</td>
<td>-0.03364</td>
<td>0.05374</td>
<td>0.05644</td>
</tr>
</tbody>
</table>

#### Schematic Representation of Cross Correlations of Residuals

<table>
<thead>
<tr>
<th>Variable/Lag</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>$y_1$</td>
<td>+ . +</td>
<td>.</td>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td>$y_2$</td>
<td>-- . +</td>
<td>.</td>
<td>.</td>
<td>. +</td>
</tr>
<tr>
<td>$y_3$</td>
<td>++ . .</td>
<td>.</td>
<td>.</td>
<td>.</td>
</tr>
</tbody>
</table>

+ is $> 2 \times$ std error, - is $< -2 \times$ std error, . is between

#### Portmanteau Test for Cross Correlations of Residuals

<table>
<thead>
<tr>
<th>Up To Lag</th>
<th>DF</th>
<th>Chi-Square</th>
<th>Pr &gt; ChiSq</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>9</td>
<td>9.69</td>
<td>0.3766</td>
</tr>
</tbody>
</table>
Output 43.2.5 describes how well each univariate equation fits the data. The residuals are off from the normality, but have no AR effects. The residuals for y1 variable have the ARCH effect.

Output 43.2.5 Diagnostic Checks Continued

<table>
<thead>
<tr>
<th>Univariate Model ANOVA Diagnostics</th>
</tr>
</thead>
<tbody>
<tr>
<td>Variable</td>
</tr>
<tr>
<td>----------</td>
</tr>
<tr>
<td>y1</td>
</tr>
<tr>
<td>y2</td>
</tr>
<tr>
<td>y3</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Univariate Model White Noise Diagnostics</th>
</tr>
</thead>
<tbody>
<tr>
<td>Variable</td>
</tr>
<tr>
<td>----------</td>
</tr>
<tr>
<td>y1</td>
</tr>
<tr>
<td>y2</td>
</tr>
<tr>
<td>y3</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Univariate Model AR Diagnostics</th>
</tr>
</thead>
<tbody>
<tr>
<td>Variable</td>
</tr>
<tr>
<td>----------</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>y1</td>
</tr>
<tr>
<td>y2</td>
</tr>
<tr>
<td>y3</td>
</tr>
</tbody>
</table>
Output 43.2.6 is the output in a matrix format associated with the PRINT=(IMPULSE=) option for the impulse response function and standard errors. The $y_3$ variable in the first row is an impulse variable. The $y_1$ variable in the first column is a response variable. The numbers, 0.96122, 0.41555, -0.40789 at lag 1 to 3 are decreasing.

**Output 43.2.6 Impulse Response Function**

<table>
<thead>
<tr>
<th>Variable Resposne</th>
<th>Lag</th>
<th>y1</th>
<th>y2</th>
<th>y3</th>
</tr>
</thead>
<tbody>
<tr>
<td>y1</td>
<td>1</td>
<td>-0.31963</td>
<td>0.14599</td>
<td>0.96122</td>
</tr>
<tr>
<td>STD</td>
<td>0.12546</td>
<td>0.54567</td>
<td>0.66431</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>-0.05430</td>
<td>0.26174</td>
<td>0.41555</td>
<td></td>
</tr>
<tr>
<td>STD</td>
<td>0.12919</td>
<td>0.54728</td>
<td>0.66311</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>0.11904</td>
<td>0.35283</td>
<td>-0.40789</td>
<td></td>
</tr>
<tr>
<td>STD</td>
<td>0.08362</td>
<td>0.38489</td>
<td>0.47867</td>
<td></td>
</tr>
<tr>
<td>y2</td>
<td>1</td>
<td>0.04393</td>
<td>-0.15273</td>
<td>0.28850</td>
</tr>
<tr>
<td>STD</td>
<td>0.03186</td>
<td>0.13857</td>
<td>0.16870</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>0.02858</td>
<td>0.11377</td>
<td>-0.08820</td>
<td></td>
</tr>
<tr>
<td>STD</td>
<td>0.03184</td>
<td>0.13425</td>
<td>0.16250</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>-0.00884</td>
<td>0.07147</td>
<td>0.11977</td>
<td></td>
</tr>
<tr>
<td>STD</td>
<td>0.01583</td>
<td>0.07914</td>
<td>0.09462</td>
<td></td>
</tr>
<tr>
<td>y3</td>
<td>1</td>
<td>-0.00242</td>
<td>0.22481</td>
<td>-0.26397</td>
</tr>
<tr>
<td>STD</td>
<td>0.02568</td>
<td>0.11168</td>
<td>0.13596</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>0.04517</td>
<td>0.26088</td>
<td>0.10998</td>
<td></td>
</tr>
<tr>
<td>STD</td>
<td>0.02563</td>
<td>0.10820</td>
<td>0.13101</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>-0.00055</td>
<td>-0.09818</td>
<td>0.09096</td>
<td></td>
</tr>
<tr>
<td>STD</td>
<td>0.01646</td>
<td>0.07823</td>
<td>0.10280</td>
<td></td>
</tr>
</tbody>
</table>

The proportions of decomposition of the prediction error covariances of three variables are given in Output 43.2.7. If you see the $y_3$ variable in the first column, then the output explains that about 64.713% of the one-step-ahead prediction error covariances of the variable $y_{3t}$ is accounted for by its own innovations, about 7.995% is accounted for by $y_{1t}$ innovations, and about 27.292% is accounted for by $y_{2t}$ innovations.
The table in Output 43.2.8 gives forecasts and their prediction error covariances.

### Output 43.2.7 Proportions of Prediction Error Covariance Decomposition

<table>
<thead>
<tr>
<th>Proportions of Prediction Error Covariances by Variable</th>
</tr>
</thead>
<tbody>
<tr>
<td>Variable</td>
</tr>
<tr>
<td>y1</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>y2</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>y3</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>

### Output 43.2.8 Forecasts

<table>
<thead>
<tr>
<th>Forecasts</th>
</tr>
</thead>
<tbody>
<tr>
<td>Variable</td>
</tr>
<tr>
<td>y1</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>y2</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>y3</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>
Output 43.2.9 shows that you cannot reject Granger noncausality from \((y_2, y_3)\) to \(y_1\) using the 0.05 significance level.

**Output 43.2.9** Granger Causality Tests

<table>
<thead>
<tr>
<th>Granger-Causality Wald Test</th>
</tr>
</thead>
<tbody>
<tr>
<td>Test</td>
</tr>
<tr>
<td>------</td>
</tr>
<tr>
<td>1</td>
</tr>
</tbody>
</table>

Test 1: Group 1 Variables: \(y_1\)
Group 2 Variables: \(y_2 y_3\)

The following SAS statements show that the variable \(y_1\) is the exogenous variable and fit the VARX(2,1) model to the data:

```sas
proc varmax data=use;
   id date interval=qtr;
   model y2 y3 = y1 / p=2 dify=(1) difx=(1) xlag=1 lagmax=3
            print=(estimates diagnose);
run;
```

The fitted VARX(2,1) model is written as

\[
\begin{pmatrix}
\Delta y_{2t} \\
\Delta y_{3t}
\end{pmatrix}
= \begin{pmatrix}
0.01542 & 0.002520 \\
0.01319 & 0.05130
\end{pmatrix} \Delta y_{1t} + \begin{pmatrix}
0.03870 & 0.00363 \\
0.12258 & 0.25811
\end{pmatrix} \Delta y_{1,t-1}
+ \begin{pmatrix}
-0.12258 & 0.25811 \\
0.24367 & -0.31809
\end{pmatrix} \begin{pmatrix}
\Delta y_{2,t-1} \\
\Delta y_{3,t-1}
\end{pmatrix}
+ \begin{pmatrix}
0.01651 & 0.03498 \\
0.34921 & -0.01664
\end{pmatrix} \begin{pmatrix}
\Delta y_{2,t-2} \\
\Delta y_{3,t-2}
\end{pmatrix}
+ \begin{pmatrix}
\epsilon_{1t} \\
\epsilon_{2t}
\end{pmatrix}
\]

The detailed output is shown in **Output 43.2.10** through **Output 43.2.13**.

**Output 43.2.10** shows the parameter estimates in terms of the constant, the current and the lag one coefficients of the exogenous variable, and the lag two coefficients of the dependent variables.
Output 43.2.10 Parameter Estimates

Analysis of German Economic Variables

The VARMAX Procedure

<table>
<thead>
<tr>
<th>Type of Model</th>
<th>VARX(2,1)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Estimation Method</td>
<td>Least Squares Estimation</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Variable</th>
<th>Constant</th>
</tr>
</thead>
<tbody>
<tr>
<td>y2</td>
<td>0.01542</td>
</tr>
<tr>
<td>y3</td>
<td>0.01319</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Lag</th>
<th>Variable</th>
<th>y1</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>y2</td>
<td>0.02520</td>
</tr>
<tr>
<td></td>
<td>y3</td>
<td>0.05130</td>
</tr>
<tr>
<td>1</td>
<td>y2</td>
<td>0.03870</td>
</tr>
<tr>
<td></td>
<td>y3</td>
<td>0.00363</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Lag</th>
<th>Variable</th>
<th>y2</th>
<th>y3</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>y2</td>
<td>-0.12258</td>
<td>0.25811</td>
</tr>
<tr>
<td></td>
<td>y3</td>
<td>0.24367</td>
<td>-0.31809</td>
</tr>
<tr>
<td>2</td>
<td>y2</td>
<td>0.01651</td>
<td>0.03498</td>
</tr>
<tr>
<td></td>
<td>y3</td>
<td>0.34921</td>
<td>-0.01664</td>
</tr>
</tbody>
</table>

Output 43.2.11 shows the parameter estimates and their significance.

Output 43.2.11 Parameter Estimates, Continued

<table>
<thead>
<tr>
<th>Equation</th>
<th>Parameter</th>
<th>Estimate</th>
<th>Standard Error</th>
<th>t Value</th>
<th>Pr &gt;</th>
<th>Variable</th>
</tr>
</thead>
<tbody>
<tr>
<td>y2</td>
<td>CONST1</td>
<td>0.01542</td>
<td>0.00443</td>
<td>3.48</td>
<td>0.0009</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>XL0_1_1</td>
<td>0.02520</td>
<td>0.03130</td>
<td>0.81</td>
<td>0.4237</td>
<td>y1(t)</td>
</tr>
<tr>
<td></td>
<td>XL1_1_1</td>
<td>0.03870</td>
<td>0.03252</td>
<td>1.19</td>
<td>0.2383</td>
<td>y1(t-1)</td>
</tr>
<tr>
<td></td>
<td>AR1_1_1</td>
<td>-0.12258</td>
<td>0.13903</td>
<td>-0.88</td>
<td>0.3811</td>
<td>y2(t-1)</td>
</tr>
<tr>
<td></td>
<td>AR1_1_2</td>
<td>0.25811</td>
<td>0.17370</td>
<td>1.49</td>
<td>0.1421</td>
<td>y3(t-1)</td>
</tr>
<tr>
<td></td>
<td>AR2_1_1</td>
<td>0.01651</td>
<td>0.13766</td>
<td>0.12</td>
<td>0.9049</td>
<td>y2(t-2)</td>
</tr>
<tr>
<td></td>
<td>AR2_1_2</td>
<td>0.03498</td>
<td>0.16783</td>
<td>0.21</td>
<td>0.8356</td>
<td>y3(t-2)</td>
</tr>
<tr>
<td>y3</td>
<td>CONST2</td>
<td>0.01319</td>
<td>0.00346</td>
<td>3.81</td>
<td>0.0003</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>XL0_2_1</td>
<td>0.05130</td>
<td>0.02441</td>
<td>2.10</td>
<td>0.0394</td>
<td>y1(t)</td>
</tr>
<tr>
<td></td>
<td>XL1_2_1</td>
<td>0.00363</td>
<td>0.02536</td>
<td>0.14</td>
<td>0.8668</td>
<td>y1(t-1)</td>
</tr>
<tr>
<td></td>
<td>AR1_2_1</td>
<td>0.24367</td>
<td>0.10842</td>
<td>2.25</td>
<td>0.0280</td>
<td>y2(t-1)</td>
</tr>
<tr>
<td></td>
<td>AR1_2_2</td>
<td>-0.31809</td>
<td>0.13546</td>
<td>-2.35</td>
<td>0.0219</td>
<td>y3(t-1)</td>
</tr>
<tr>
<td></td>
<td>AR2_2_1</td>
<td>0.34921</td>
<td>0.10736</td>
<td>3.25</td>
<td>0.0018</td>
<td>y2(t-2)</td>
</tr>
<tr>
<td></td>
<td>AR2_2_2</td>
<td>-0.01664</td>
<td>0.13088</td>
<td>-0.13</td>
<td>0.8992</td>
<td>y3(t-2)</td>
</tr>
</tbody>
</table>
Output 43.2.12 shows the innovation covariance matrix estimates, the various information criteria results, and the tests for white noise residuals. The residuals are uncorrelated except at lag 3 for $y_2$ variable.

**Output 43.2.12** Diagnostic Checks

| Covariances of Innovations |
|-----------------------------|---------------------|
| Variable       | $y_2$   | $y_3$   |
| $y_2$          | 0.00014 | 0.00006 |
| $y_3$          | 0.00006 | 0.00009 |

<table>
<thead>
<tr>
<th>Information Criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td>AICC</td>
</tr>
<tr>
<td>HQC</td>
</tr>
<tr>
<td>AIC</td>
</tr>
<tr>
<td>SBC</td>
</tr>
<tr>
<td>FPEC</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Cross Correlations of Residuals</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lag</td>
</tr>
<tr>
<td>-----</td>
</tr>
<tr>
<td>0</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>1</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>2</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>3</td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Schematic Representation of Cross Correlations of Residuals</th>
</tr>
</thead>
<tbody>
<tr>
<td>Variable/Lag</td>
</tr>
<tr>
<td>$y_2$</td>
</tr>
<tr>
<td>$y_3$</td>
</tr>
</tbody>
</table>

* + is > 2*std error, - is < -2*std error, . is between

<table>
<thead>
<tr>
<th>Portmanteau Test for Cross Correlations of Residuals</th>
</tr>
</thead>
<tbody>
<tr>
<td>Up To Lag</td>
</tr>
<tr>
<td>-----------</td>
</tr>
<tr>
<td>3</td>
</tr>
</tbody>
</table>

Output 43.2.13 describes how well each univariate equation fits the data. The residuals are off from the normality, but have no ARCH and AR effects.
Output 43.2.13 Diagnostic Checks Continued

<table>
<thead>
<tr>
<th>Variable</th>
<th>R-Square</th>
<th>Deviation</th>
<th>F Value</th>
<th>Pr &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>y2</td>
<td>0.0897</td>
<td>0.01188</td>
<td>1.08</td>
<td>0.3809</td>
</tr>
<tr>
<td>y3</td>
<td>0.2796</td>
<td>0.00926</td>
<td>4.27</td>
<td>0.0011</td>
</tr>
</tbody>
</table>

Univariate Model White Noise Diagnostics

<table>
<thead>
<tr>
<th>Durbin Watson</th>
<th>Chi-Square</th>
<th>Pr &gt; ChiSq</th>
<th>F Value</th>
<th>Pr &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>y2</td>
<td>2.02413</td>
<td>14.54</td>
<td>0.0007</td>
<td>0.49</td>
</tr>
<tr>
<td>y3</td>
<td>2.13414</td>
<td>32.27</td>
<td>&lt;.0001</td>
<td>0.08</td>
</tr>
</tbody>
</table>

Univariate Model AR Diagnostics

<table>
<thead>
<tr>
<th>AR1</th>
<th>AR2</th>
<th>AR3</th>
<th>AR4</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>F Value</td>
<td>Pr &gt; F</td>
<td>F Value</td>
<td>Pr &gt; F</td>
</tr>
<tr>
<td>y2</td>
<td>0.04</td>
<td>0.8448</td>
<td>0.04</td>
</tr>
<tr>
<td>y3</td>
<td>0.62</td>
<td>0.4343</td>
<td>0.62</td>
</tr>
</tbody>
</table>

Example 43.3: Analysis of Restricted Cointegrated Systems

The structural relationships between economic time series have been of interest for decades. Because of the cointegration, the vector error correction model (VECM), introduced by Engle and Granger (1987), is one of the most important tools for performing such analysis. Although there exist analytical solutions for a nonrestricted VECM and some restricted VECMs in special forms, the estimation of a generally restricted VECM relies on numerical methods. This section illustrates how to use the RESTRICT (or BOUND) and TEST statements, together with the COINTEG statement, to estimate the restricted VECM and perform the statistical tests. For more information about this topic, see Boswijk and Doornik (2004) and references therein.

The data are simulated based on the VECM,

\[
\Delta y_t = \alpha \beta' y_{t-1} + \Phi_1^{*} \Delta y_{t-1} + \Theta_0^{*} x_t + \epsilon_t
\]

\[
= \begin{bmatrix}
0.01 & -0.02 \\
-0.03 & 0.04 \\
0.05 & -0.06 \\
0 & 0
\end{bmatrix}
\begin{bmatrix}
1 & 0 & -1 & 0 \\
0 & 1 & 0 & -1
\end{bmatrix}
y_{t-1}
\]

\[
+ \begin{bmatrix}
-0.01 & 0.03 & 0.05 & -0.02 \\
0.02 & -0.04 & 0.06 & 0.03 \\
0 & 0 & 0.10 & 0 \\
0 & 0 & 0 & 0.04
\end{bmatrix}
\Delta y_{t-1} + \begin{bmatrix}
0 \\
0 \\
0 \\
0
\end{bmatrix}
x_t + \epsilon_t,
\]

\[\epsilon_t \sim iid N(0, \Sigma), \Sigma = I_4\]

where \(I_4\) is the \(4 \times 4\) identity matrix.
The following statements implement the simulation:

```r
title 'Analysis of Restricted Cointegrated Systems';
proc iml;
alpha = {0.01 -0.02, -0.03 0.04, 0.05 -0.06, 0 0};
beta = {1 0, 0 1, -1 0, 0 -1};
phiStar = {-0.01 0.03 0.05 -0.02,
          0.02 -0.04 0.06 0.03,
          0 0 0.10 0,
          0 0 0 0.04};
Pi = alpha * beta`;
A1 = I(4) + Pi + phiStar;
A2 = -phiStar;
phi = A1 // A2;
sig = I(4);

/* to simulate the vector time series */
T = 600;
myseed = 2;
call varmasim(y, phi) sigma=sig n=T seed=myseed;
x = J(T,1,0);
do i = 1 to T;
   x[i] = normal(myseed);
end;
y = y || x;

cn = {'y1' 'y2' 'y3' 'y4' 'x'};
create simul5 from y[colname=cn];
append from y;
close;
quit;
```

**Weak Exogeneity Tests**

This example shows different methods for checking weak exogeneity.

The first method uses the EXOGENEITY option in the following statements, and the test results are shown in Output 43.3.1:

```r
/* Method 1 -- To use the EXOGENEITY option */
ods output LogLikelihood = tbl_ll_g;
proc varmax data=simul5;
   model y1 y2 y3 y4 = x / noint p=2;
   cointeg rank=2 exogeneity;
run;
```
Output 43.3.1 Test Weak Exogeneity with the EXOGENEITY Option

Analysis of Restricted Cointegrated Systems

The VARMAX Procedure

<table>
<thead>
<tr>
<th>Testing Weak Exogeneity of Each Variable</th>
<th>DF</th>
<th>Chi-Square</th>
<th>Pr &gt; ChiSq</th>
</tr>
</thead>
<tbody>
<tr>
<td>y1</td>
<td>2</td>
<td>102.96</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>y2</td>
<td>2</td>
<td>116.12</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>y3</td>
<td>2</td>
<td>200.80</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>y4</td>
<td>2</td>
<td>3.99</td>
<td>0.1357</td>
</tr>
</tbody>
</table>

The second method uses the RESTRICT statement and then the likelihood ratio (LR) test in the following statements. The results are shown in Output 43.3.2. In theory, the first and second methods should have exactly same statistics and $p$-values because they implement the same LR tests. However, because of the difference between the analytical solution and the numerical solution for the restricted VECM, the statistics are a little different, although for the 0.05 significance level they lead to the same correct conclusion: the variable $y_1$ is not the weak exogeneity of variables $y_2$, $y_3$, and $y_4$; the variable $y_2$ is not the weak exogeneity of variables $y_1$, $y_3$, and $y_4$; the variable $y_3$ is not the weak exogeneity of variables $y_1$, $y_2$, and $y_4$; the variable $y_4$ is the weak exogeneity of variables $y_1$, $y_2$, and $y_3$.
Chapter 43: The VARMAX Procedure

Output 43.3.2 Test Weak Exogeneity with the RESTRICT Statement and LR Tests

Analysis of Restricted Cointegrated Systems

<table>
<thead>
<tr>
<th>Test</th>
<th>DF</th>
<th>Stat</th>
<th>pValue</th>
</tr>
</thead>
<tbody>
<tr>
<td>H0: Alpha(1)=0</td>
<td>2</td>
<td>109.05157</td>
<td>0</td>
</tr>
<tr>
<td>H0: Alpha(2)=0</td>
<td>2</td>
<td>124.56535</td>
<td>0</td>
</tr>
<tr>
<td>H0: Alpha(3)=0</td>
<td>2</td>
<td>238.35505</td>
<td>0</td>
</tr>
<tr>
<td>H0: Alpha(4)=0</td>
<td>2</td>
<td>5.0877699</td>
<td>0.0785606</td>
</tr>
</tbody>
</table>

The third method uses the TEST statement, which implements the Wald tests. Asymptotically, the Wald test has the same distribution as the LR test.

/* Method 3 -- To use the TEST statement and the Wald test */
proc varmax data=simul5;
  model y1 y2 y3 y4 = x / noint p=2;
  cointeg rank=2;
  test alpha(1,1:2) = 0;
  test alpha(2,1:2) = 0;
  test alpha(3,1:2) = 0;
  test alpha(4,1:2) = 0;
run;

Based on the test results shown in Output 43.3.3, the same correct conclusion can be obtained at the 0.05 significance level: the variable \( y_1 \) is not the weak exogeneity of variables \( y_2, y_3, \) and \( y_4 \); the variable \( y_2 \) is not the weak exogeneity of variables \( y_1, y_3, \) and \( y_4 \); the variable \( y_3 \) is not the weak exogeneity of variables \( y_1, y_2, \) and \( y_4 \); the variable \( y_4 \) is the weak exogeneity of variables \( y_1, y_2, \) and \( y_3 \).

Output 43.3.3 Test Weak Exogeneity with the TEST Statement, Wald Tests

Analysis of Restricted Cointegrated Systems

The VARMAX Procedure

<table>
<thead>
<tr>
<th>Testing of the Parameters</th>
<th>Test</th>
<th>DF</th>
<th>Chi-Square</th>
<th>Pr &gt; ChiSq</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
<td>2</td>
<td>113.27</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>2</td>
<td>129.15</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>2</td>
<td>245.21</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>2</td>
<td>4.81</td>
<td>0.0903</td>
</tr>
</tbody>
</table>

Identification

This example shows how important it is to identify \( \alpha \) and \( \beta \) when applying the Wald test on \( \alpha \). First, in the following statements, there are no constraints on \( \beta \):

```plaintext
proc varmax data=simul5;
  model y1 y2 y3 y4 = x / noint p=2;
  cointeg rank=2;
  test alpha(1,2) = alpha(2,2) + alpha(3,2);
run;
```
As shown in Output 43.3.4, based on the test results, the null hypothesis H0: $\alpha[1,2] = \alpha[2,2] + \alpha[3,2]$ should be rejected at the 0.05 significance level, although the true parameter values for the data generating process indicate that H0 is correct.

**Output 43.3.4** Importance of Identifying $\alpha$ and $\beta$ in the Wald Test

**Analysis of Restricted Cointegrated Systems**

The VARMAX Procedure

<table>
<thead>
<tr>
<th>Testing of the Parameters</th>
<th>Test</th>
<th>DF</th>
<th>Chi-Square</th>
<th>Pr &gt; ChiSq</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
<td>21.44</td>
<td>&lt;.0001</td>
<td></td>
</tr>
</tbody>
</table>

In the following statements, $r^2$ constraints are now imposed on $\beta$, where $r$ is the cointegration rank:

```plaintext
proc varmax data=simul5;
  model y1 y2 y3 y4 = x / noint p=2;
  cointeg rank=2;
  restrict beta(3:4,1:2) = -I(2);
  test alpha(1,2) = alpha(2,2) + alpha(3,2);
run;
```

As shown in Output 43.3.5, the null hypothesis cannot be rejected at 0.05 significance level; that is to say, the correct conclusion is achieved.

**Output 43.3.5** Importance of Identifying $\alpha$ and $\beta$ in the Wald Test, Continued

**Analysis of Restricted Cointegrated Systems**

The VARMAX Procedure

<table>
<thead>
<tr>
<th>Testing of the Parameters</th>
<th>Test</th>
<th>DF</th>
<th>Chi-Square</th>
<th>Pr &gt; ChiSq</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
<td>0.18</td>
<td>0.6750</td>
<td></td>
</tr>
</tbody>
</table>

Besides $\alpha$, other short-run parameters in a VECM can also be tested by using the TEST statement. Because short-run parameters other than $\alpha$ are identified in a VECM, it is not necessary to impose additional constraints on $\alpha$ and $\beta$. The following statements test the null hypothesis H0: $\Phi_1^* = 0$:

```plaintext
proc varmax data=simul5;
  model y1 y2 y3 y4 = x / noint p=2;
  cointeg rank=2;
  test ar(2);
run;
```

According to the results shown in Output 43.3.6, the null hypothesis should be rejected at the 0.05 significance level.
Output 43.3.6  Wald Tests for Short-Run Parameters

Analysis of Restricted Cointegrated Systems

The VARMAX Procedure

<table>
<thead>
<tr>
<th>Test DF</th>
<th>Chi-Square</th>
<th>Pr &gt; ChiSq</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>32.79</td>
<td>0.0079</td>
</tr>
</tbody>
</table>

The following statements test the null hypothesis H0: $\Theta_0^* = 0$:

```plaintext
proc varmax data=simul5;
  model y1 y2 y3 y4 = x / noint p=2;
  cointeg rank=2;
  test xl;
run;
```

According to the results shown in Output 43.3.7, the null hypothesis cannot be rejected at the 0.05 significance level.

Output 43.3.7  Wald Tests for Short-Run Parameters, Continued

Analysis of Restricted Cointegrated Systems

The VARMAX Procedure

<table>
<thead>
<tr>
<th>Test DF</th>
<th>Chi-Square</th>
<th>Pr &gt; ChiSq</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>6.01</td>
<td>0.1982</td>
</tr>
</tbody>
</table>

Besides the parameters that are estimated in a VECM, you can also use the TEST statement on $\Pi(=\alpha\beta')$, and $\delta_0$ or $\delta_1$ for Case 2 or 4 when the constant or linear trend, respectively, is restricted in the error correction term. However, keep in mind that the covariance matrix for these parameter estimates is singular when the cointegration rank is less than the number of dependent variables; hence, you might not get the results for some tests.

```plaintext
proc varmax data=simul5;
  model y1 y2 y3 y4 = x / noint p=2;
  cointeg rank=2;
  test ar(1,4,1:4);
  test ar(1,4,{1 3});
run;
```

As shown in Output 43.3.8, the first test on H0: $\Pi[4,\cdot] = 0$ cannot be calculated, whereas the second test on H0: $\Pi[4, 1] = \Pi[4, 3] = 0$ can be.
**Tests for Long-Run Parameter**

This example focuses on testing the relationships on the long-run parameter $\beta$. Here, only the following specific form of hypothesis is discussed,

$$H_0: \beta = (H, \phi)$$

where $H$ is a known $k \times r_1$ matrix, $\phi$ is a freely varying $k \times (r - r_1)$ parameter matrix, $k$ is the number of dependent variables, $r$ is the cointegration rank, and $0 \leq r_1 \leq r$. Other forms of hypothesis—for example, $H_0: \beta = (H_1\phi_1, \ldots, H_r\phi_r)$ or $H_0: H\text{vec}(\beta) = h$—are omitted, although they can also be implemented in the same logic. The following statements test the null hypothesis that $(1 \ 0 -1 0)'$ is in the cointegrating space that is spanned by $\beta$:

```sas
/* Use the RESTRICT statement and LR test for restrictions on Beta. */
/* H0: Beta = [ H, phi ] where H is known and phi is free */
ods output LogLikelihood = tbl_ll_r2;
proc varmax data=simul5;
   model y1 y2 y3 y4 = x / noint p=2;
   cointeg rank=2;
   restrict beta(,1) = {1, 0, -1, 0};
   nloptions tech=qn maxit=5000;
run;

proc iml;
   use tbl_ll_g;
   read all var {nValue1} into ll_g;
   close;
   use tbl_ll_r2;
   read all var {nValue1} into ll_r;
   close;
   DF = (4-2)*1; /* DF = (k-r)*r_1 */
   Stat = -2*ll_r - ll_g;
   pValue = 1-cdf("CHISQUARE", stat, df);
   Test = "H0: Beta[1,1:4] = (1 0 -1 0)'";
   print Test DF Stat pValue;
quit;
```

According to the result shown in Output 43.3.9, the null hypothesis cannot be rejected at the 0.05 significance level.
Output 43.3.9 LR Tests on Long-Run Parameter $\beta$

Analysis of Restricted Cointegrated Systems

<table>
<thead>
<tr>
<th>Test</th>
<th>DF</th>
<th>Stat</th>
<th>pValue</th>
</tr>
</thead>
<tbody>
<tr>
<td>H0: Beta$[1:1:4] = {1 0 -1 0}'</td>
<td>2</td>
<td>2.5835027</td>
<td>0.2747891</td>
</tr>
</tbody>
</table>

The following statements test the null hypothesis that the cointegrating space is spanned by $(1 0 -1 0 1 0 -1)'$:

```{}
/* H0: Beta = H, where H is the true Beta for DGP */
ods output LogLikelihood = tbl_ll_r3;
proc varmax data=simul5;
  model y1 y2 y3 y4 = x / noint p=2;
  cointeg rank=2;
  restrict beta = I(2) // (-I(2));
  nloptions tech=qn maxit=5000;
run;

proc iml;
  use tbl_ll_g;
  read all var {nValue1} into ll_g;
  close;
  use tbl_ll_r3;
  read all var {nValue1} into ll_r;
  close;
  DF = (4-2)*2; /* DF = (k-r)*r_1 */
  Stat = -2*(ll_r - ll_g);
  pValue = 1-cdf("CHISQUARE", stat, df);
  Test = "H0: Beta = \{1 0, 0 1, -1 0, 0 -1\}"
  print Test DF Stat pValue;
quit;
```

According to the result shown in Output 43.3.10, the null hypothesis cannot be rejected at the 0.05 significance level.

Output 43.3.10 LR Tests on Long-Run Parameter $\beta$, Continued

Analysis of Restricted Cointegrated Systems

<table>
<thead>
<tr>
<th>Test</th>
<th>DF</th>
<th>Stat</th>
<th>pValue</th>
</tr>
</thead>
<tbody>
<tr>
<td>H0: Beta = ${1 0, 0 1, -1 0, 0 -1}$</td>
<td>4</td>
<td>1.5854995</td>
<td>0.8113959</td>
</tr>
</tbody>
</table>

The following statements test the null hypothesis that the cointegrating space is spanned by $(1 0 1 0, 0 1 0 1)'$, the orthogonal matrix to the true $\beta$ for the data generating process:

```{}
/* H0: Beta = H, where H is the matrix orthogonal to the true Beta for DGP */
ods output LogLikelihood = tbl_ll_r4;
proc varmax data=simul5;
  model y1 y2 y3 y4 = x / noint p=2;
  cointeg rank=2;
```
Example 43.4: Analysis of Euro Foreign Exchange Reference Rates

```
restrict beta = {1 0, 0 1, 1 0, 0 1};
nloptions tech=qn maxit=5000;
run;

proc iml;
use tbl_ll_g;
read all var {nValue1} into ll_g;
close;
use tbl_ll_r4;
read all var {nValue1} into ll_r;
close;
DF = (4-2)*2; /* DF = (k-r)*r_1 */
Stat = -2*(ll_r - ll_g);
pValue = 1-cdf("CHISQUARE", stat, df);
Test = "H0: Beta = {1 0, 0 1, 1 0, 0 1}";
print Test DF Stat pValue;
quit;
```

According to the result shown in Output 43.3.11, the null hypothesis should be rejected at the 0.05 significance level.

Output 43.3.11  LR Tests on Long-Run Parameter $\beta$, Continued

<table>
<thead>
<tr>
<th>Test</th>
<th>DF</th>
<th>Stat</th>
<th>pValue</th>
</tr>
</thead>
<tbody>
<tr>
<td>H0: Beta = {1 0, 0 1, 1 0, 0 1}</td>
<td>4</td>
<td>227.69117</td>
<td>0</td>
</tr>
</tbody>
</table>

For the VECM, the BOUND statement can be regarded as an alias of the RESTRICT statement; that is, you can directly replace any RESTRICT statement with a BOUND statement and get the same result. The linear inequality constraints in the restricted cointegrated systems are not discussed in this section, although they are also supported in the BOUND and RESTRICT statements. For more information, see the sections “BOUND Statement” on page 3024 and “RESTRICT Statement” on page 3050.

Obtaining the numerical solution for the restricted VECM is not an easy task in most cases. You might need to use the INITIAL and NLOPTIONS statements to tune the process. For more information, see the sections “INITIAL Statement” on page 3032 and “NLOPTIONS Statement” on page 3049.

Example 43.4: Analysis of Euro Foreign Exchange Reference Rates

This example illustrates how to use and select the VARMA-GARCH model for exchange rates, a general type of financial data. As shown in much of the literature, the financial variables are cross-correlated and autocorrelated not only on first moments, but also on second moments. The VARMA-GARCH model and the vector error correction GARCH model are often used to catch the stylized fact.

The data, downloaded from European Central Bank website (https://www.ecb.europa.eu), consist of four pairs of daily foreign exchange reference rates: the euro and the Australian dollar (AUD), the euro and the British pound sterling (GBP), the euro and the Japanese yen (JPY), and the euro and the US dollar (USD). The full sample covers the period from January 4, 1999, to February 12, 2015 (4,127 days). In the following statements, the series are logarithmically transformed, and the returns (in percentage) are calculated:
title 'Analysis of Euro Foreign Exchange Reference Rates';
data eurofxrr;
  input date : MMDDYY10. aud gbp jpy usd;
  label aud='The euro and the Australian dollar'
    usd='The euro and the U.S. dollar'
    jpy='The euro and the Japanese yen'
    gbp='The euro and the British pound sterling';
  logAUD = log(AUD); logGBP = log(GBP);
  logJPY = log(JPY); logUSD = log(USD);
  rAUD = (logAUD - lag(logAUD))*100;
  rGBP = (logGBP - lag(logGBP))*100;
  rJPY = (logJPY - lag(logJPY))*100;
  rUSD = (logUSD - lag(logUSD))*100;
datalines;
01/04/1999  1.9100  0.71110  133.73  1.1789
01/05/1999  1.8944  0.71220  130.96  1.1790
01/06/1999  1.8820  0.70760  131.42  1.1743
01/07/1999  1.8474  0.70585  129.43  1.1632
... more lines ...
02/10/2015  1.4522  0.74200  134.67  1.1297
02/11/2015  1.4606  0.73960  135.50  1.1314
02/12/2015  1.4761  0.73760  135.72  1.1328;

Although it is well known that unit roots exist in the exchange rate series and they are not cointegrated, you
 can use the following statements to verify:

    /*--- Unit Roots and Cointegration in Log Exchange Rates ---*/
    proc varmax data=eurofxrr;
      model logAUD logGBP logJPY logUSD / p=2 dftest cointtest;
    run;

According to the results of the Dickey-Fuller unit root tests shown in Output 43.4.1, the null hypothesis that
there is a unit root in each series cannot be rejected at the 5% significance level. The results of the Johansen
cointegration rank trace tests shown in Output 43.4.2 confirm that there is no cointegration between series
because the null hypothesis that the cointegration rank is 0, in both unrestricted and restricted cases, cannot
be rejected at the 5% significance level. Because there is no cointegration, you do not need to consider vector
error correction models; otherwise, the final selected model might be a vector error correction GARCH
model, instead of a VARMA-GARCH model.
Output 43.4.1 Dickey-Fuller Unit Root Tests

Analysis of Euro Foreign Exchange Reference Rates

The VARMAX Procedure

<table>
<thead>
<tr>
<th>Variable</th>
<th>Type</th>
<th>Unit Root Test</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Rho</td>
<td>Pr &lt; Rho</td>
</tr>
<tr>
<td>logAUD</td>
<td>Zero Mean</td>
<td>-1.05</td>
</tr>
<tr>
<td></td>
<td>Single Mean</td>
<td>-9.44</td>
</tr>
<tr>
<td></td>
<td>Trend</td>
<td>-13.85</td>
</tr>
<tr>
<td>logGBP</td>
<td>Zero Mean</td>
<td>-0.57</td>
</tr>
<tr>
<td></td>
<td>Single Mean</td>
<td>-3.23</td>
</tr>
<tr>
<td></td>
<td>Trend</td>
<td>-11.11</td>
</tr>
<tr>
<td>logJPY</td>
<td>Zero Mean</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td>Single Mean</td>
<td>-6.11</td>
</tr>
<tr>
<td></td>
<td>Trend</td>
<td>-6.56</td>
</tr>
<tr>
<td>logUSD</td>
<td>Zero Mean</td>
<td>-1.46</td>
</tr>
<tr>
<td></td>
<td>Single Mean</td>
<td>-3.29</td>
</tr>
<tr>
<td></td>
<td>Trend</td>
<td>-5.76</td>
</tr>
</tbody>
</table>

Output 43.4.2 Johansen Cointegration Rank Trace Tests

<table>
<thead>
<tr>
<th>Rank</th>
<th>H0: Rank=0</th>
<th>H1: Rank=1</th>
<th>Eigenv:</th>
<th>Trace Pr &gt; Trace</th>
<th>Drift in ECM</th>
<th>Drift in Process</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.0059</td>
<td>36.6836</td>
<td>0.3601</td>
<td>Constant Linear</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0.0018</td>
<td>12.1427</td>
<td>0.9269</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>2</td>
<td>0.0008</td>
<td>4.7724</td>
<td>0.8319</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>3</td>
<td>3</td>
<td>0.0003</td>
<td>1.3036</td>
<td>0.2532</td>
<td></td>
</tr>
</tbody>
</table>

Before modeling returns, you can test whether unit roots still exist in the differenced data with the following statement:

```sas
/*--- Unit Roots in Returns and Model Specification ---*/
proc varmax data=eurofxrr;
   model rAUD rGBP rJPY rUSD / p=2 dftest;
      test const; test ar(1); test ar(2);
run;
```
Output 43.4.3 shows that there is no unit root in each differenced series.

**Output 43.4.3**  Dickey-Fuller Unit Root Tests

**Analysis of Euro Foreign Exchange Reference Rates**

*The VARMAX Procedure*

<table>
<thead>
<tr>
<th>Variable</th>
<th>Type</th>
<th>Rho</th>
<th>Pr &lt; Rho</th>
<th>Tau</th>
<th>Pr &lt; Tau</th>
</tr>
</thead>
<tbody>
<tr>
<td>rAUD</td>
<td>Zero Mean</td>
<td>-4242.7</td>
<td>0.0001</td>
<td>-46.04</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td></td>
<td>Single Mean</td>
<td>-4243.7</td>
<td>0.0001</td>
<td>-46.04</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td></td>
<td>Trend</td>
<td>-4244.2</td>
<td>0.0001</td>
<td>-46.04</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>rGBP</td>
<td>Zero Mean</td>
<td>-4358.4</td>
<td>0.0001</td>
<td>-46.67</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td></td>
<td>Single Mean</td>
<td>-4358.4</td>
<td>0.0001</td>
<td>-46.67</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td></td>
<td>Trend</td>
<td>-4358.5</td>
<td>0.0001</td>
<td>-46.66</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>rJPY</td>
<td>Zero Mean</td>
<td>-4181.4</td>
<td>0.0001</td>
<td>-45.72</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td></td>
<td>Single Mean</td>
<td>-4181.4</td>
<td>0.0001</td>
<td>-45.72</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td></td>
<td>Trend</td>
<td>-4181.9</td>
<td>0.0001</td>
<td>-45.72</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>rUSD</td>
<td>Zero Mean</td>
<td>-4306.8</td>
<td>0.0001</td>
<td>-46.40</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td></td>
<td>Single Mean</td>
<td>-4306.8</td>
<td>0.0001</td>
<td>-46.39</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td></td>
<td>Trend</td>
<td>-4307.4</td>
<td>0.0001</td>
<td>-46.39</td>
<td>&lt;.0001</td>
</tr>
</tbody>
</table>

The preceding statements also test whether the constant and each of two lags of AR terms are 0. The test results are shown in Output 43.4.4.

**Output 43.4.4**  Tests on Constant and AR Terms

<table>
<thead>
<tr>
<th>Testing of the Parameters</th>
<th>Test DF</th>
<th>Chi-Square</th>
<th>Pr &gt; ChiSq</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
<td>4.46</td>
<td>0.7976</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>59.42</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>15.67</td>
<td>0.4759</td>
</tr>
</tbody>
</table>

The null hypothesis that the constant term is 0 and the null hypothesis that the second lag AR term is 0 are both accepted at the 5% significance level. However, the null hypothesis that the first lag AR term is 0 is rejected at the 5% significance level. In the remaining model selection process, only the first lag AR term is considered.

The following statements estimate a zero-mean VAR(1) model and also print some diagnostic results:

```plaintext
/*--- VAR Model ---*/

proc varmax data=eurofxrr;
   model rAUD rGBP rJPY rUSD / noint p=1 print=(diagnose);
run;
```
Output 43.4.5 shows the information criteria for the estimated zero-mean VAR(1) model. In this example, AICC is used as the criterion for model selection: the smaller the AICC, the better the model.

**Output 43.4.5** Information Criteria for the VAR Model

Analysis of Euro Foreign Exchange Reference Rates

<table>
<thead>
<tr>
<th>Information Criteria</th>
<th>AIC</th>
<th>HQC</th>
<th>AICC</th>
<th>SBC</th>
<th>FPEC</th>
</tr>
</thead>
<tbody>
<tr>
<td>AICC</td>
<td>-1745.29</td>
<td>-1687.44</td>
<td>-1745.64</td>
<td>-1581.19</td>
<td>0.011938</td>
</tr>
</tbody>
</table>

Diagnostics are printed because the PRINT=(DIAGNOSE) option is specified. As shown in Output 43.4.6, the null hypotheses that there is no ARCH effect in each series are all rejected at the 5% significance level.

**Output 43.4.6** Tests on ARCH Effects

<table>
<thead>
<tr>
<th>Univariate Model White Noise Diagnostics</th>
<th>Normality</th>
<th>ARCH</th>
</tr>
</thead>
<tbody>
<tr>
<td>Variable: rAUD</td>
<td>Durbin</td>
<td>Pr &gt; ChiSq</td>
</tr>
<tr>
<td>Watson</td>
<td>1.99811</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Chi-Square</td>
<td>8277.31</td>
<td></td>
</tr>
<tr>
<td>Pr &gt; ChiSq</td>
<td>217.35</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>F Value</td>
<td>&lt;.0001</td>
<td></td>
</tr>
<tr>
<td>Pr &gt; F</td>
<td>217.35</td>
<td>&lt;.0001</td>
</tr>
</tbody>
</table>

To find the right GARCH model, you can start with the VAR(1)-CCC-GARCH(1,1) model (which is usually the fastest one to be estimated) as in the following statement:

```plaintext
/*---- VAR CCC GARCH Model ----*/

proc varmax data=eurofxrr;
  model rAUD rGBP rJPY rUSD / noint p=1;
  garch p=1 q=1 form=ccc;
run;
```

Compared to the AICC for the zero-mean VAR(1) model (shown in **Output 43.4.5**), the AICC for VAR(1)-CCC-GARCH(1,1) model, as shown in **Output 43.4.7**, dramatically decreases, which means that the ARCH effects do play an important role and should be modeled.
Chapter 43: The VARMAX Procedure

Output 43.4.7 Information Criteria for VAR CCC GARCH Model

Analysis of Euro Foreign Exchange Reference Rates

The VARMAX Procedure

<table>
<thead>
<tr>
<th>Information Criteria</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>AICC</td>
<td>-4646.77</td>
</tr>
<tr>
<td>HQC</td>
<td>-4571.24</td>
</tr>
<tr>
<td>AIC</td>
<td>-4647.35</td>
</tr>
<tr>
<td>SBC</td>
<td>-4432.31</td>
</tr>
<tr>
<td>FPEC</td>
<td>0.011966</td>
</tr>
</tbody>
</table>

As indicated by its name, a basic assumption of the CCC GARCH model is that the conditional correlation is time-invariant, which might not be true. The following statements estimate a BEKK GARCH model to see whether modeling the conditional correlation could improve the model performance:

```plaintext
/*--- VAR BEKK GARCH Model ---*/

proc varmax data=eurofxrr outest=oediagbekk;
  model rAUD rGBP rJPY rUSD / noint p=1;
  garch p=1 q=1 form=bekk;
run;
```

As shown in Output 43.4.8, the AICC for the VAR BEKK GARCH model does get smaller than the AICC for the CCC GARCH model (shown in Output 43.4.7). The smaller AICC implies that the assumption of the CCC GARCH model might be inaccurate.

Output 43.4.8 Information Criteria for VAR BEKK GARCH Model

Analysis of Euro Foreign Exchange Reference Rates

The VARMAX Procedure

<table>
<thead>
<tr>
<th>Information Criteria</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>AICC</td>
<td>-5667.7</td>
</tr>
<tr>
<td>HQC</td>
<td>-5539.55</td>
</tr>
<tr>
<td>AIC</td>
<td>-5669.38</td>
</tr>
<tr>
<td>SBC</td>
<td>-5302.54</td>
</tr>
<tr>
<td>FPEC</td>
<td>0.011979</td>
</tr>
</tbody>
</table>

One shortcoming of the BEKK GARCH model is that it has too many parameters. In practice, especially for a large number of dependent variables, the scalar BEKK GARCH model and the diagonal BEKK GARCH model are often applied, as shown in the following statements. In the RESTRICT statement, matrix operations are used; using matrix operations is much more concise than restricting tens of ARCH and GARCH parameters one by one.
Example 43.4: Analysis of Euro Foreign Exchange Reference Rates

/**** VAR Scalar BEKK GARCH Model ****/

proc varmax data=eurofxrr outest=odediagbekk;
   model rAUD rGBP rJPY rUSD / noint p=1;
   garch p=1 q=1 form=bekk;
   restrict ach(1)=ach(1,1,1)*I(4), gch(1)=gch(1,1,1)*I(4);
run;

/**** VAR Diagonal BEKK GARCH Model ****/

proc varmax data=eurofxrr outest=odediagbekk;
   model rAUD rGBP rJPY rUSD / noint p=1;
   garch p=1 q=1 form=bekk;
   restrict ach(1)=ach(1)#I(4), gch(1)=gch(1)#I(4);
run;

The AICCs for the scalar and diagonal BEKK GARCH models are shown in Output 43.4.9 and Output 43.4.10, respectively, and both of them are larger than the AICC for the BEKK GARCH model (shown in Output 43.4.8). Hence, so far, the VAR BEKK GARCH model is the best.

**Output 43.4.9** Information Criteria for VAR Scalar BEKK GARCH Model

**Output 43.4.10** Information Criteria for VAR Diagonal BEKK GARCH Model

Another type of multivariate GARCH model that is suitable for modeling the time-varying conditional correlation is the dynamic conditional correlation (DCC) GARCH model, as indicated by its name. The following statements estimate the DCC GARCH model:
Chapter 43: The VARMAX Procedure

/*--- VAR DCC GARCH Model ---*/
proc varmax data=eurofxrr;
   model rAUD rGBP rJPY rUSD / noint p=1;
   garch p=1 q=1 form=dcc;
run;

As shown in Output 43.4.11, the AICC for the VAR DCC GARCH model is smaller than the AICC for the VAR BEKK GARCH model (shown in Output 43.4.8), implying that the best model should be in the class of DCC GARCH models.

Output 43.4.11 Information Criteria for VAR DCC GARCH Model

Analysis of Euro Foreign Exchange Reference Rates

The VARMAX Procedure

<table>
<thead>
<tr>
<th>Information Criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td>AICC     -5689.43</td>
</tr>
<tr>
<td>HQC       -5609.5</td>
</tr>
<tr>
<td>AIC       -5690.08</td>
</tr>
<tr>
<td>SBC       -5462.39</td>
</tr>
<tr>
<td>FPEC      0.011973</td>
</tr>
</tbody>
</table>

Could the DCC GARCH model be more parsimonious? The following statements use the sample correlation matrix of the standardized residuals (saving six parameters) to calculate the unconditional correlation matrix in the DCC GARCH model:

/*--- Parsimonious VAR DCC GARCH Model ---*/
proc varmax data=eurofxrr;
   model rAUD rGBP rJPY rUSD / noint p=1;
   garch p=1 q=1 form=dcc corrconst=expect;
run;

The AICC of the parsimonious VAR DCC GARCH model, as shown in Output 43.4.12, becomes a little smaller. Hence, the best model so far is the parsimonious VAR DCC GARCH model.

Output 43.4.12 Information Criteria for the Parsimonious VAR DCC GARCH Model

Analysis of Euro Foreign Exchange Reference Rates

The VARMAX Procedure

<table>
<thead>
<tr>
<th>Information Criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td>AICC     -5694.89</td>
</tr>
<tr>
<td>HQC       -5628.19</td>
</tr>
<tr>
<td>AIC       -5695.35</td>
</tr>
<tr>
<td>SBC       -5505.6</td>
</tr>
<tr>
<td>FPEC      0.011973</td>
</tr>
</tbody>
</table>
Another way to refine the model is to try different subforms of GARCH models for each series. The following statements estimate the VAR DCC EGARCH model and produce Output 43.4.13:

```latex
/**-- VAR DCC EGARCH Model --*/

proc varmax data=eurofxrr;
   model rAUD rGBP rJPY rUSD / noint p=1;
   garch p=1 q=1 form=dcc subform=egarch corrconst=expect;
   nloptions maxit=5000 pall;
run;
```

The following statements estimate the VAR DCC PGARCH model and produce Output 43.4.14:

```latex
/**-- VAR DCC PGARCH Model --*/

proc varmax data=eurofxrr;
   model rAUD rGBP rJPY rUSD / noint p=1;
   garch p=1 q=1 form=dcc subform=pgarch corrconst=expect;
   nloptions maxit=5000 pall;
run;
```

The following statements estimate the VAR DCC QGARCH model and produce Output 43.4.15:

```latex
/**-- VAR DCC QGARCH Model --*/

proc varmax data=eurofxrr;
   model rAUD rGBP rJPY rUSD / noint p=1;
   garch p=1 q=1 form=dcc subform=qgarch corrconst=expect;
   nloptions maxit=5000 pall;
run;
```

The following statements estimate the VAR DCC TGARCH model and produce Output 43.4.16:

```latex
/**-- VAR DCC TGARCH Model --*/

proc varmax data=eurofxrr;
   model rAUD rGBP rJPY rUSD / noint p=1;
   garch p=1 q=1 form=dcc subform=tgarch corrconst=expect;
   nloptions maxit=5000 pall;
run;
```

Comparing the AICCs shown in Output 43.4.13 through Output 43.4.16, you find that the AICC for the VAR DCC PGARCH model is the smallest. Hence, the best model becomes the zero-mean VAR(1)-DCC-PGARCH(1,1) model, whose unconditional correlation matrix is estimated by the sample correlation matrix of the standardized residuals.
Output 43.4.13  Information Criteria for the Parsimonious VAR DCC EGARCH Model

Analysis of Euro Foreign Exchange Reference Rates

The VARMAX Procedure

<table>
<thead>
<tr>
<th>Information Criteria</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>AICC</td>
<td>-5704.33</td>
</tr>
<tr>
<td>HQC</td>
<td>-5628.81</td>
</tr>
<tr>
<td>AIC</td>
<td>-5704.92</td>
</tr>
<tr>
<td>SBC</td>
<td>-5489.87</td>
</tr>
<tr>
<td>FPEC</td>
<td>0.011982</td>
</tr>
</tbody>
</table>

Output 43.4.14  Information Criteria for the Parsimonious VAR DCC PGARCH Model

Analysis of Euro Foreign Exchange Reference Rates

The VARMAX Procedure

<table>
<thead>
<tr>
<th>Information Criteria</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>AICC</td>
<td>-5724.44</td>
</tr>
<tr>
<td>HQC</td>
<td>-5640.1</td>
</tr>
<tr>
<td>AIC</td>
<td>-5725.16</td>
</tr>
<tr>
<td>SBC</td>
<td>-5484.82</td>
</tr>
<tr>
<td>FPEC</td>
<td>0.011974</td>
</tr>
</tbody>
</table>

Output 43.4.15  Information Criteria for the Parsimonious VAR DCC QGARCH Model

Analysis of Euro Foreign Exchange Reference Rates

The VARMAX Procedure

<table>
<thead>
<tr>
<th>Information Criteria</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>AICC</td>
<td>-5696.97</td>
</tr>
<tr>
<td>HQC</td>
<td>-5621.44</td>
</tr>
<tr>
<td>AIC</td>
<td>-5697.55</td>
</tr>
<tr>
<td>SBC</td>
<td>-5482.51</td>
</tr>
<tr>
<td>FPEC</td>
<td>0.011972</td>
</tr>
</tbody>
</table>
Output 43.4.16 Information Criteria for the Parsimonious VAR DCC TGARCH Model

Analysis of Euro Foreign Exchange Reference Rates

The VARMAX Procedure

<table>
<thead>
<tr>
<th>Information</th>
<th>Criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td>AICC</td>
<td>-5705.59</td>
</tr>
<tr>
<td>HQC</td>
<td>-5630.06</td>
</tr>
<tr>
<td>AIC</td>
<td>-5706.17</td>
</tr>
<tr>
<td>SBC</td>
<td>-5491.13</td>
</tr>
<tr>
<td>FPEC</td>
<td>0.011973</td>
</tr>
</tbody>
</table>

Output 43.4.17 shows that most of the AR parameter estimates in the VAR DCC PGARCH model are not significant.

Output 43.4.17 AR Parameter Estimates for the Parsimonious VAR DCC PGARCH Model

<table>
<thead>
<tr>
<th>Equation</th>
<th>Parameter</th>
<th>Estimate</th>
<th>Standard Error</th>
<th>t Value</th>
<th>Pr &gt;</th>
<th>Variable</th>
</tr>
</thead>
<tbody>
<tr>
<td>rAUD</td>
<td>AR1_1_1</td>
<td>0.05719</td>
<td>0.01790</td>
<td>3.19</td>
<td>0.0014</td>
<td>rAUD(t-1)</td>
</tr>
<tr>
<td></td>
<td>AR1_1_2</td>
<td>0.00040</td>
<td>0.02396</td>
<td>0.02</td>
<td>0.9868</td>
<td>rGBP(t-1)</td>
</tr>
<tr>
<td></td>
<td>AR1_1_3</td>
<td>-0.02307</td>
<td>0.01619</td>
<td>-1.42</td>
<td>0.1543</td>
<td>rJPY(t-1)</td>
</tr>
<tr>
<td></td>
<td>AR1_1_4</td>
<td>0.02005</td>
<td>0.02020</td>
<td>0.99</td>
<td>0.3210</td>
<td>rUSD(t-1)</td>
</tr>
<tr>
<td>rGBP</td>
<td>AR1_2_1</td>
<td>0.02683</td>
<td>0.01146</td>
<td>2.34</td>
<td>0.0193</td>
<td>rAUD(t-1)</td>
</tr>
<tr>
<td></td>
<td>AR1_2_2</td>
<td>0.04515</td>
<td>0.01880</td>
<td>2.40</td>
<td>0.0163</td>
<td>rGBP(t-1)</td>
</tr>
<tr>
<td></td>
<td>AR1_2_3</td>
<td>-0.00467</td>
<td>0.01137</td>
<td>-0.41</td>
<td>0.6815</td>
<td>rJPY(t-1)</td>
</tr>
<tr>
<td></td>
<td>AR1_2_4</td>
<td>-0.04648</td>
<td>0.01474</td>
<td>-3.15</td>
<td>0.0016</td>
<td>rUSD(t-1)</td>
</tr>
<tr>
<td>rJPY</td>
<td>AR1_3_1</td>
<td>0.05599</td>
<td>0.01846</td>
<td>3.03</td>
<td>0.0024</td>
<td>rAUD(t-1)</td>
</tr>
<tr>
<td></td>
<td>AR1_3_2</td>
<td>-0.05011</td>
<td>0.02696</td>
<td>-1.86</td>
<td>0.0632</td>
<td>rGBP(t-1)</td>
</tr>
<tr>
<td></td>
<td>AR1_3_3</td>
<td>-0.00186</td>
<td>0.01895</td>
<td>-0.10</td>
<td>0.9220</td>
<td>rJPY(t-1)</td>
</tr>
<tr>
<td></td>
<td>AR1_3_4</td>
<td>-0.00839</td>
<td>0.02226</td>
<td>-0.38</td>
<td>0.7063</td>
<td>rUSD(t-1)</td>
</tr>
<tr>
<td>rUSD</td>
<td>AR1_4_1</td>
<td>0.03855</td>
<td>0.01512</td>
<td>2.55</td>
<td>0.0108</td>
<td>rAUD(t-1)</td>
</tr>
<tr>
<td></td>
<td>AR1_4_2</td>
<td>0.00551</td>
<td>0.02289</td>
<td>0.24</td>
<td>0.8099</td>
<td>rGBP(t-1)</td>
</tr>
<tr>
<td></td>
<td>AR1_4_3</td>
<td>0.00086</td>
<td>0.01477</td>
<td>0.06</td>
<td>0.9536</td>
<td>rJPY(t-1)</td>
</tr>
<tr>
<td></td>
<td>AR1_4_4</td>
<td>-0.03202</td>
<td>0.02011</td>
<td>-1.59</td>
<td>0.1113</td>
<td>rUSD(t-1)</td>
</tr>
</tbody>
</table>

The following statements test the significance of some parameter estimates:

```
/**** Significance Of Some Parameter Estimates *****/

proc varmax data=eurofxrr;
   model rAUD rGBP rJPY rUSD / noint p=1;
   garch p=1 q=1 form=dcc subform=pgarch corrconst=expect;
   nloptions maxit=5000 pall;
   test ar(1, 1, 2:4), ar(1, 2, 3), ar(1, 3, 3:4), ar(1, 4, 2:4);
run;
```

As shown in Output 43.4.18, the null hypothesis that all nine of the parameters in the TEST statement are 0 cannot be rejected at the 5% significance level.
The following statements estimate the VAR DCC PGARCH model without those insignificant parameters:

```sas
/*--- VAR DCC PGARCH Model w/o Insignificant Parameters ---*/
proc varmax data=eurofxrr;
  model rAUD rGBP rJPY rUSD / noint p=1;
  garch p=1 q=1 form=dcc subform=pgarch corrconst=expect;
  nloptions maxit=5000 pall;
  restrict ar(1, 1, 2:4), ar(1, 2, 3), ar(1, 3, 3:4), ar(1, 4, 2:4);
run;
```

As shown in Output 43.4.19, the AICC does improve and decrease. Further refining the model is possible but beyond the scope of this example. Hence, the best model, according to the AICC, is the zero-mean VAR(1)-DCC-PGARCH(1,1) model without insignificant AR parameters, and its unconditional correlation matrix is estimated by the sample correlation matrix of the standardized residuals.

**Output 43.4.19** Information Criteria for the VAR DCC PGARCH Model without Insignificant Parameters

**Analysis of Euro Foreign Exchange Reference Rates**

The VARMAX Procedure

<table>
<thead>
<tr>
<th>Information Criteria</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>AICC</td>
<td>-5735.05</td>
</tr>
<tr>
<td>HQC</td>
<td>-5670.56</td>
</tr>
<tr>
<td>AIC</td>
<td>-5735.48</td>
</tr>
<tr>
<td>SBC</td>
<td>-5552.06</td>
</tr>
<tr>
<td>FPEC</td>
<td>0.011996</td>
</tr>
</tbody>
</table>

This example focuses only on using the information criterion to distinguish models. In practice, the forecast performance of the model might be more important. The VARMAX procedure supports multistep forecasting in both VARMAX-GARCH models and vector error correction GARCH models. Hence, although it is not covered in this example, you can also use the VARMAX procedure and a criterion based on out-of-sample forecast to perform model selection.
Example 43.5: Numerous Examples

The following are examples of syntax for model fitting:

```sas
/* Data 'a' Generated Process */
proc iml;
  sig = {1.0 0.5, 0.5 1.25};
  phi = {1.2 -0.5, 0.6 0.3};
  call varmasim(y,phi) sigma = sig n = 100 seed = 46859;
  cn = {'y1' 'y2'};
  create a from y[colname=cn];
  append from y;
run;

/* when the series has a linear trend */
proc varmax data=a;
  model y1 y2 / p=1 trend=linear;
run;

/* Fit subset of AR order 1 and 3 */
proc varmax data=a;
  model y1 y2 / p=(1,3);
run;

/* Check if the series is nonstationary */
proc varmax data=a;
  model y1 y2 / p=1 dftest print=(roots);
run;

/* Fit VAR(1) in differencing */
proc varmax data=a;
  model y1 y2 / p=1 print=(roots) dify=(1);
run;

/* Fit VAR(1) in seasonal differencing */
proc varmax data=a;
  model y1 y2 / p=1 dify=(4) lagmax=5;
run;

/* Fit VAR(1) in both regular and seasonal differencing */
proc varmax data=a;
  model y1 y2 / p=1 dify=(1,4) lagmax=5;
run;

/* Fit VAR(1) in different differencing */
proc varmax data=a;
  model y1 y2 / p=1 dif=(y1(1,4) y2(1)) lagmax=5;
run;

/* Options related to prediction */
proc varmax data=a;
  model y1 y2 / p=1 lagmax=3
    print=(impulse covpe(5) decompose(5));
```

run;

/* Options related to tentative order selection */
proc varmax data=a;
  model y1 y2 / p=1 lagmax=5 minic
       print=(parcoef pcancorr pcorr);
run;

/* Automatic selection of the AR order */
proc varmax data=a;
  model y1 y2 / minic=(type=aic p=5);
run;

/* Compare results of LS and Yule-Walker Estimators */
proc varmax data=a;
  model y1 y2 / p=1 print=(yw);
run;

/* BVAR(1) of the nonstationary series y1 and y2 */
proc varmax data=a;
  model y1 y2 / p=1
       prior=(lambda=1 theta=0.2 ivar);
run;

/* BVAR(1) of the nonstationary series y1 */
proc varmax data=a;
  model y1 y2 / p=1
       prior=(lambda=0.1 theta=0.15 ivar=(y1));
run;

/* Data 'b' Generated Process */
proc iml;
  sig = { 0.5 0.14 -0.08 -0.03, 0.14 0.71 0.16 0.1,
          -0.08 0.16 0.65 0.23, -0.03 0.1 0.23 0.16};
  sig = sig * 0.0001;
  phi = {1.2 -0.5 0. 0.1, 0.6 0.3 -0.2 0.5,
         0.4 0. -0.2 0.1, -1.0 0.2 0.7 -0.2};
  call varmasim(y,phi) sigma = sig n = 100 seed = 32567;
  cn = {'y1' 'y2' 'y3' 'y4'};
  create b from y[colname=cn];
  append from y;
quit;

/* Cointegration Rank Test using Trace statistics */
proc varmax data=b;
  model y1-y4 / p=2 lagmax=4 cointtest;
run;

/* Cointegration Rank Test using Max statistics */
proc varmax data=b;
  model y1-y4 / p=2 lagmax=4 cointtest=(johansen=(type=max));
run;

/* Common Trends Test using Filter(Differencing) statistics */
Example 43.5: Numerous Examples

proc varmax data=b;
   model y1-y4 / p=2 lagmax=4 cointtest=(sw);
run;

/* Common Trends Test using Filter(Residual) statistics */
proc varmax data=b;
   model y1-y4 / p=2 lagmax=4 cointtest=(sw=(type=filtres lag=1));
run;

/* Common Trends Test using Kernel statistics */
proc varmax data=b;
   model y1-y4 / p=2 lagmax=4 cointtest=(sw=(type=kernel lag=1));
run;

/* Cointegration Rank Test for I(2) */
proc varmax data=b;
   model y1-y4 / p=2 lagmax=4 cointtest=(johansen=(iorder=2));
run;

/* Fit VECM(2) with rank=3 */
proc varmax data=b;
   model y1-y4 / p=2 lagmax=4 print=(roots iarr);
   cointeg rank=3 normalize=y1;
run;

/* Weak Exogenous Testing for each variable */
proc varmax data=b outstat=bbb;
   model y1-y4 / p=2 lagmax=4;
   cointeg rank=3 exogeneity normalize=y1;
run;

/* Hypotheses Testing for long-run and adjustment parameter */
proc varmax data=b outstat=bbb;
   model y1-y4 / p=2 lagmax=4;
   cointeg rank=3 normalize=y1
       h=(1 0 0, 0 1 0, -1 0 0, 0 0 1)
       j=(1 0 0, 0 1 0, 0 0 1, 0 0 0);
run;

/* ordinary regression model */
proc varmax data=grunfeld;
   model y1 y2 = x1-x3;
run;

/* Ordinary regression model with subset lagged terms */
proc varmax data=grunfeld;
   model y1 y2 = x1 / xlag=(1,3);
run;

/* VARX(1,1) with no current time Exogenous Variables */
proc varmax data=grunfeld;
   model y1 y2 = x1 / p=1 xlag=1 nocurrentx;
run;
Example 43.6: Illustration of ODS Graphics

This example illustrates the use of ODS Graphics. For information about the graphics available in the VARMAX procedure, see the section “ODS Graphics” on page 3165.

The following statements use the SASHELP.WORKERS data set to study the time series of electrical workers and its interaction with the series of masonry workers. The series and predict plots, the residual plot, and the forecast plot are created in Output 43.6.1 through Output 43.6.3. These are a selection of the plots created by the VARMAX procedure.

```sas
title "Illustration of ODS Graphics";
proc varmax data=sashelp.workers plot(unpack)=(residual model forecasts);
   id date interval=month;
   model electric masonry / dify=(1,12) noint p=1;
   output lead=12;
run;
```
Output 43.6.1  Series and Predicted Series Plots
Output 43.6.2 Residual Plot

Prediction Errors for ELECTRIC

DATE


Error

-7.5, -5.0, -2.5, 0.0, 2.5, 5.0, 7.5

Prediction Errors, One Standard Error, Two Standard Errors


### Subject Index

- **Akaike’s information criterion**
  - VARMAX procedure, 3110
- **asymptotic distribution of impulse response functions**
  - VARMAX procedure, 3094, 3104
- **asymptotic distribution of the parameter estimation**
  - VARMAX procedure, 3104
- **Bayesian vector autoregressive models**
  - VARMAX procedure, 3046, 3099
- **cointegration**
  - VARMAX procedure, 3112
- **cointegration testing**
  - VARMAX procedure, 3043, 3117
- **common trends**
  - VARMAX procedure, 3113
- **common trends testing**
  - VARMAX procedure, 3045, 3114
- **computational details**
  - VARMAX procedure, 3166
- **confidence limits**
  - VARMAX procedure, 3151
- **convergence problems**
  - VARMAX procedure, 3166
- **covariance stationarity**
  - VARMAX procedure, 3142
- **CPU requirements**
  - VARMAX procedure, 3167
- **decomposition of prediction error covariance**
  - VARMAX procedure, 3039, 3085
- **Dickey-Fuller test**
  - VARMAX procedure, 3043
- **differencing**
  - VARMAX procedure, 3035
- **dynamic simultaneous equation models**
  - VARMAX procedure, 3068
- **example of Bayesian VAR modeling**
  - VARMAX procedure, 2989
- **example of Bayesian VECM modeling**
  - VARMAX procedure, 2996
- **example of causality testing**
  - VARMAX procedure, 3006
- **example of cointegration testing**
  - VARMAX procedure, 2992
- **example of multivariate GARCH modeling**
  - VARMAX procedure, 3143
- **example of restricted parameter estimation and testing**
  - VARMAX procedure, 3004
- **example of VAR modeling**
  - VARMAX procedure, 2982
- **example of VARMA modeling**
  - VARMAX procedure, 3105
- **example of vector autoregressive modeling with exogenous variables**
  - VARMAX procedure, 3001
- **example of vector error correction modeling**
  - VARMAX procedure, 2990
- **examples of multivariate GARCH modeling**
  - VARMAX procedure, 3007
- **forecasting**
  - VARMAX procedure, 3082
- **forecasting of Bayesian vector autoregressive models**
  - VARMAX procedure, 3100
- **Granger causality test**
  - VARMAX procedure, 3095
- **impulse response function**
  - VARMAX procedure, 3040, 3071
- **infinite order AR representation**
  - VARMAX procedure, 3039
- **infinite order MA representation**
  - VARMAX procedure, 3040, 3071
- **invertibility**
  - VARMAX procedure, 3101
- **J function**
  - creating a matrix of identical values, 3062
- **long-run relations testing**
  - VARMAX procedure, 3125
- **memory requirements**
  - VARMAX procedure, 3166
- **minimum information criteria method**
  - VARMAX procedure, 3091
- **missing values**
  - VARMAX procedure, 3064
- **multivariate GARCH modeling**
  - VARMAX procedure, 3030
- **multivariate model diagnostic checks**
  - VARMAX procedure, 3110
- **ODS graph names**
  - VARMAX procedure, 3165
output data sets
  VARMAX procedure, 3151

partial autoregression coefficient
  VARMAX procedure, 3040, 3088
partial canonical correlation
  VARMAX procedure, 3040, 3090
partial correlation
  VARMAX procedure, 3089
prediction error covariance
  VARMAX procedure, 3038, 3082, 3084

sample cross covariances
  VARMAX procedure, 3039, 3087
sample cross-correlations
  VARMAX procedure, 3038, 3087
seasonal dummies and time trends
  VARMAX procedure, 3098
state space representation
  VARMAX procedure, 3065
stationarity
  VARMAX procedure, 3093, 3101

tentative order selection
  VARMAX procedure, 3087
time intervals
  VARMAX procedure, 3031

univariate model diagnostic checks
  VARMAX procedure, 3112

VARMAX procedure
  Akaike's information criterion, 3110
  asymptotic distribution of impulse response functions, 3094, 3104
  asymptotic distribution of the parameter estimation, 3104
  Bayesian vector autoregressive models, 3046, 3099
cointegration, 3112
cointegration testing, 3043, 3117
common trends, 3113
common trends testing, 3045, 3114
computational details, 3166
confidence limits, 3151
convergence problems, 3166
covariance stationarity, 3142
CPU requirements, 3167
decomposition of prediction error covariance, 3039, 3085
Dickey-Fuller test, 3043
differencing, 3035
dynamic simultaneous equation models, 3068
example of Bayesian VAR modeling, 2989
example of Bayesian VECM modeling, 2996
example of causality testing, 3006
example of cointegration testing, 2992
example of multivariate GARCH modeling, 3143
example of restricted parameter estimation and testing, 3004
example of VAR modeling, 2982
example of VARMA modeling, 3105
example of vector autoregressive modeling with exogenous variables, 3001
example of vector error correction modeling, 2990
examples of multivariate GARCH modeling, 3007
forecasting, 3082
forecasting of Bayesian vector autoregressive models, 3100
Granger causality test, 3095
impulse response function, 3040, 3071
infinite order AR representation, 3039
infinite order MA representation, 3040, 3071
invertibility, 3101
long-run relations testing, 3125
memory requirements, 3166
minimum information criteria method, 3091
missing values, 3064
multivariate GARCH modeling, 3030
multivariate model diagnostic checks, 3110
ODS graph names, 3165
output data sets, 3151
partial autoregression coefficient, 3040, 3088
partial canonical correlation, 3040, 3090
partial correlation, 3089
prediction error covariance, 3038, 3082, 3084
sample cross covariances, 3039, 3087
sample cross-correlations, 3038, 3087
seasonal dummies and time trends, 3098
state space representation, 3065
stationarity, 3093, 3101
tentative order selection, 3087
time intervals, 3031
univariate model diagnostic checks, 3112
vector autoregressive models, 3092
vector autoregressive models with exogenous variables, 3096
vector autoregressive moving-average models, 3064, 3101
vector error correction models, 3048, 3115
weak exogeneity testing, 3127
Yule-Walker estimates, 3041

VARMAX procedure, 3092
vector autoregressive models with exogenous variables
VARMAX procedure, 3096
vector autoregressive moving-average models
VARMAX procedure, 3064, 3101
vector error correction models
    VARMAX procedure, 3048, 3115

weak exogeneity testing
    VARMAX procedure, 3127

Yule-Walker estimates
    VARMAX procedure, 3041
Syntax Index

ALIGN= option
   ID statement (VARMAX), 3031
ALPHA= option
   OUTPUT statement (VARMAX), 3050
BACK= option
   OUTPUT statement (VARMAX), 3050
BOUND statement
   VARMAX procedure, 3024, 3189
BY statement
   VARMAX procedure, 3025
CAUSAL statement
   VARMAX procedure, 3025
CENTER option
   MODEL statement (VARMAX), 3034
COINTEG statement
   VARMAX procedure, 3026, 3126
COINTTEST= option
   MODEL statement (VARMAX), 3043
COINTTEST=(JOHANSEN) option
   MODEL statement (VARMAX), 3044
COINTTEST=(JOHANSEN=(IORDER=)) option
   MODEL statement (VARMAX), 3044, 3133
COINTTEST=(JOHANSEN=(NORMALIZE=)) option
   MODEL statement (VARMAX), 3044, 3120
COINTTEST=(JOHANSEN=(TYPE=)) option
   MODEL statement (VARMAX), 3044, 3120
COINTTEST=(SIGLEVEL=) option
   MODEL statement (VARMAX), 3045
COINTTEST=(SW) option
   MODEL statement (VARMAX), 3045, 3114
COINTTEST=(SW=(LAG=)) option
   MODEL statement (VARMAX), 3045
COINTTEST=(SW=(TYPE=)) option
   MODEL statement (VARMAX), 3045
CORRCONSTANT= option
   GARCH statement, 3030
DATA= option
   PROC VARMAX statement, 3021
DFTEST option
   MODEL statement (VARMAX), 3043, 3167
DFTEST=(DLAG=) option
   MODEL statement (VARMAX), 3043
DIF= option
   MODEL statement (VARMAX), 3034
DIFX= option
   MODEL statement (VARMAX), 3035
DIFY= option
   MODEL statement (VARMAX), 3035, 3178
ECM= option
   MODEL statement (VARMAX), 3048
ECM=(ECTRENД) option
   MODEL statement (VARMAX), 3048
ECM=(NORMALIZE=) option
   MODEL statement (VARMAX), 3049
ECM=(RANK=) option
   MODEL statement (VARMAX), 3048
ECTRENД option
   COINTEG statement (VARMAX), 3027, 3123
EXOGENEITY option
   COINTEG statement (VARMAX), 3027, 3129
FORM= option
   GARCH statement, 3030
GARCH statement
   VARMAX procedure, 3030
GROUP1 option
   CAUSAL statement (VARMAX), 3025
GROUP2 option
   CAUSAL statement (VARMAX), 3025
H= option
   COINTEG statement (VARMAX), 3027, 3126
ID statement
   VARMAX procedure, 3031
INITIAL statement
   VARMAX procedure, 3032
INTERV AL= option
   ID statement (VARMAX), 3031
J function, 3062
J= option
   COINTEG statement (VARMAX), 3029
LAGMAX= option
   MODEL statement (VARMAX), 3037
LEAD= option
   OUTPUT statement (VARMAX), 3050
METHOD= option
   MODEL statement (VARMAX), 3035
MINIC= option
   MODEL statement (VARMAX), 3042
MINIC=(P=) option  
  MODEL statement (VARMAX), 3042, 3092

MINIC=(PERERROR=) option  
  MODEL statement (VARMAX), 3042

MINIC=(Q=) option  
  MODEL statement (VARMAX), 3042, 3092

MINIC=(TYPE=) option  
  MODEL statement (VARMAX), 3042

MODEL statement  
  VARMAX procedure, 3034

NLC option  
  COINTEG statement (VARMAX), 3029

NLOPTIONS statement  
  VARMAX procedure, 3049, 3105

NOCURRENTX option  
  MODEL statement (VARMAX), 3036

NOINT option  
  MODEL statement (VARMAX), 3036

NOPRINT option  
  MODEL statement (VARMAX), 3036
  OUTPUT statement (VARMAX), 3050
  PROC VARMAX statement (VARMAX), 3152

NORMALIZE= option  
  COINTEG statement (VARMAX), 2993, 3029, 3167

NSEASON= option  
  MODEL statement (VARMAX), 3036

OUT= option  
  OUTPUT statement (VARMAX), 3050, 3151

OUTCOV option  
  PROC VARMAX statement, 3021, 3152

OUTEST= option  
  PROC VARMAX statement, 3021, 3152

OUTHT= option  
  GARCH statement, 3030
  PROC VARMAX statement, 3154

OUTPUT statement  
  VARMAX procedure, 3049

OUTSTAT= option  
  PROC VARMAX statement, 3022, 3157

P= option  
  GARCH statement, 3030
  MODEL statement (VARMAX), 3041

PRINT= option  
  MODEL statement (VARMAX), 3038

PRINT=(COVPE) option  
  MODEL statement (VARMAX), 3038, 3083

PRINT=(COVX) option  
  MODEL statement (VARMAX), 3039

PRINT=(COVY) option  
  MODEL statement (VARMAX), 3039

PRINT=(DECOMPOSE) option  
  MODEL statement (VARMAX), 3039, 3085

PRINT=(DIAGNOSE) option  
  MODEL statement (VARMAX), 3039

PRINT=(DYNAMIC) option  
  MODEL statement (VARMAX), 3039, 3069

PRINT=(ESTIMATES) option  
  MODEL statement (VARMAX), 3039

PRINT=(IARR) option  
  MODEL statement (VARMAX), 2993, 3039

PRINT=(IMPULSE) option  
  MODEL statement (VARMAX), 3076

PRINT=(IMPULSE=) option  
  MODEL statement (VARMAX), 3040

PRINT=(IMPULSX) option  
  MODEL statement (VARMAX), 3072

PRINT=(IMPULSX=) option  
  MODEL statement (VARMAX), 3040

PRINT=(PARCOEF) option  
  MODEL statement (VARMAX), 3040, 3088

PRINT=(PCANCORR) option  
  MODEL statement (VARMAX), 3040, 3091

PRINT=(PCORR) option  
  MODEL statement (VARMAX), 3040, 3090

PRINT=(ROOTS) option  
  MODEL statement (VARMAX), 3041, 3093

PRINT=(YW) option  
  MODEL statement (VARMAX), 3041

PRINTALL option  
  MODEL statement (VARMAX), 3038

PRINTFORM= option  
  MODEL statement (VARMAX), 3038, 3072

PRIOR option  
  MODEL statement (VARMAX), 3046

PRIOR=(IVAR) option  
  MODEL statement (VARMAX), 3046

PRIOR=(LAMBDA=) option  
  MODEL statement (VARMAX), 3046

PRIOR=(MEAN=) option  
  MODEL statement (VARMAX), 3046

PRIOR=(NREP=) option  
  MODEL statement (VARMAX), 3046

PRIOR=(THETA=) option  
  MODEL statement (VARMAX), 3047

PROC VARMAX statement (VARMAX), 3021

Q= option
GARCH statement, 3030
MODEL statement (VARMAX), 3042, 3105

RANK= option
   COINTEG statement (VARMAX), 2993, 3026, 3126

RESTRICT statement
   VARMAX procedure, 3050, 3189

SCENTER option
   MODEL statement (VARMAX), 3036

SUBFORM= option
   GARCH statement, 3030

TEST statement
   VARMAX procedure, 3006, 3062, 3189

TREND= option
   MODEL statement (VARMAX), 3037

TREND=LINEAR option
   MODEL statement (VARMAX), 3123

VARDEF= option
   MODEL statement (VARMAX), 3037
   VARMAX procedure, 3018
   syntax, 3018

XLAG= option
   MODEL statement (VARMAX), 3042