Vector Autoregressions (VAR and VEC)

The structural approach to simultaneous equations modeling uses economic theory to describe the relationships between several variables of interest. The resulting model is then estimated, and used to test the empirical relevance of the theory.

Unfortunately, economic theory is often not rich enough to provide a tight specification of the dynamic relationship among variables. Furthermore, estimation and inference are complicated by the fact that endogenous variables may appear on both the left and right sides of the equations.

These problems lead to alternative, non-structural, approaches to modeling the relationship between several variables. Here we describe the estimation and analysis of vector autoregression (VAR) and the vector error correction (VEC) models. We also describe tools for testing for the presence of cointegrating relationships among several variables.

VAR Theory

The vector autoregression (VAR) is commonly used for forecasting systems of interrelated time series and for analyzing the dynamic impact of random disturbances on the system of variables.

The VAR approach sidesteps the need for structural modeling by modeling every endogenous variable in the system as a function of the lagged values of *all* of the endogenous variables in the system.

The mathematical form of a VAR is

$$y_t = A_1 y_{t-1} + \cdots + A_p y_{t-p} + B x_t + \epsilon_t$$

where y_t is a *k* vector of endogenous variables, x_t is a *d* vector of exogenous variables, $A_1, \ldots, A_p$ and B are matrices of coefficients to be estimated, and ϵ_t is a vector of innovations that may be contemporaneously correlated with each other but are uncorrelated with their own lagged values and uncorrelated with all of the right-hand side variables.

Since only lagged values of the endogenous variables appear on the right-hand side of each equation, there is no issue of simultaneity, and OLS is the appropriate estimation technique. Note that the assumption that the disturbances are not serially correlated is not restrictive because any serial correlation could be absorbed by adding more lagged *y*'s.

As an example of a VAR, suppose that industrial production (IP) and money supply (M1) are jointly determined by a two variable VAR and let a constant be the only exogenous variable. With two lagged values of the endogenous variables, the VAR is

$$IP_t = a_{11} IP_{t-1} + a_{12} M1_{t-1} + b_{11} IP_{t-2} + b_{12} M1_{t-2} + c_1 + \epsilon_{1,t}$$
$$M1_t = a_{21} IP_{t-1} + a_{22} M1_{t-1} + b_{21} IP_{t-2} + b_{22} M1_{t-2} + c_2 + \epsilon_{2,t}$$

where *a*, *b*, *c* are the parameters to be estimated.

Estimating a VAR in EViews

To specify a vector autoregression, you must create a VAR object. Select **Quick/Estimate VAR…** or type `var` in the command window. Fill out the dialog that appears with the appropriate information:

- Enter the lag information in the first edit box. This information tells EViews which lags should be included on the right-hand side of each equation. This information is entered in pairs: each pair of numbers defines a *range* of lags. For example, the lag pair:

      1 2

tells EViews to use the first and second lags of all of the variables in the system as right-hand side variables.

You can add any number of lag intervals, all entered in pairs. The lag specification:

      2 4 6 9 11 11

uses lags two through four, lags six through nine, and lag eleven.

- Enter the endogenous and exogenous variables in the appropriate edit boxes.

- Select the specification type: **Unrestricted VAR** or **Vector Error Correction** (VEC). What we

Select the specification type: **Unrestricted VAR** or **Vector Error Correction** (VEC). What we have been calling a VAR is the *unrestricted* VAR. VECs will be explained in detail below in Vector Error Correction.

- Check or uncheck the box to include or exclude the constant from the right-hand side of the equations.

The example dialog corresponds to the VAR example above: IP and M1 are the endogenous variables, the model uses lags 1 through 2, and the constant is the only exogenous variable. You will note that we have both entered C as an exogenous variable, and checked the box instructing EViews to include a constant in the VAR. This is redundant, but harmless, since EViews will only include a single constant.

VAR Estimation Output

Once you have specified the VAR, click **OK**. EViews will display the estimation results in the VAR window. Each column in the table corresponds to the equation for one endogenous variable in the VAR. For each right-hand side variable, EViews reports a coefficient point estimate, the estimated coefficient standard error, and the *t*-statistic.

Two types of regression statistics are reported at the bottom of the VAR estimation output.

R-squared	0.999221	0.999915	0.968018
Adj. R-squared	0.999195	0.999912	0.966937
Sum sq. resids	113.8813	1232.453	98.39849
S.E. equation	0.566385	1.863249	0.526478
Log likelihood	-306.3509	-744.5662	-279.4628
Akaike AIC	-1.102274	1.279331	-1.248405
Schwarz SC	-0.964217	1.417388	-1.110348
Mean dependent	70.97919	339.7451	6.333891
S.D. dependent	19.95932	198.6301	2.895381

Determinant Residual Covariance	0.259637
Log Likelihood	-1318.390
Akaike Information Criteria	-1.136514
Schwarz Criteria	-0.722342

The first part of the output displays the standard regression statistics for each equation. The results are computed separately, using the residuals for each equation and arranged in the appropriate column.

The numbers at the very bottom of the table are the regression statistics for the VAR system. The determinant of the residual covariance is computed as

$$|\hat{\Omega}| = \det\left(\sum_t \hat{\epsilon}_t \hat{\epsilon}_t' / T\right)$$

where $\hat{\epsilon}_t$ is the *k* vector of residuals. The log-likelihood value is computed assuming a multivariate normal (Gaussian) distribution as:

$$\ell = -\frac{Tk}{2}(1 + \log 2\pi) - \frac{T}{2}\log|\hat{\Omega}|$$

and the two information criteria are computed as

$$\text{AIC} = -2\ell/T + 2n/T, \quad \text{SC} = -2\ell/T + n\log T/T$$

where $n = k(d+pk)$ is the total number of estimated parameters in the VAR. These information criteria can be used for model selection such as determining the lag length of the VAR; the smaller the value of the information criteria, the "better" the model. It is worth noting that by ignoring constant terms, some reference sources define the AIC for a multivariate normal system in a slightly different fashion, as:

$$\log|\hat{\Omega}| + 2n/T$$.

See also Information Criteria for additional discussion.

Working with a VAR

Once you have estimated your VAR, EViews provides various views and procedures that allow you to

use the estimated VAR for further analyses. In this section, we discuss views that are specific to VARs; for other views and procedures, see the discussion of System Estimation Views.

In empirical applications, the main uses of the VAR are the impulse response analysis, variance decomposition, and Granger causality tests.

Impulse Response Functions

An *impulse response function* traces the effect of a one standard deviation shock to one of the innovations on current and future values of the endogenous variables.

A shock to the *i*-th variable directly affects the *i*-th variable, and is also transmitted to all of the endogenous variables through the dynamic structure of the VAR.

Consider a simple bivariate VAR(1):

$$IP_t = a_{11}IP_{t-1} + a_{12}M1_{t-1} + \epsilon_{1,t}$$
$$M1_t = a_{21}IP_{t-1} + a_{22}M1_{t-1} + \epsilon_{2,t}$$

A change in $\epsilon_{1,t}$ will immediately change the value of current IP. It will also change all future values of IP and M1 since lagged IP appears in both equations.

If the innovations, $\epsilon_{1,t}$ and $\epsilon_{2,t}$ in our example, are uncorrelated, interpretation of the impulse response is straightforward. $\epsilon_{1,t}$ is the innovation for IP and $\epsilon_{2,t}$ is the innovation for M1. The impulse response functions for $\epsilon_{2,t}$ measures the effect of a one standard deviation monetary shock on current and future industrial production and money stock.

The innovations are, however, usually correlated, so that they have a common component which cannot be associated with a specific variable. A somewhat arbitrary but common method of dealing with this issue is to attribute all of the effect of any common component to the variable that comes first in the VAR system. In our example, the common component of $\epsilon_{1,t}$ and $\epsilon_{2,t}$ is totally attributed to $\epsilon_{1,t}$, because $\epsilon_{1,t}$ precedes $\epsilon_{2,t}$. $\epsilon_{1,t}$ is then the IP innovation, and $\epsilon_{2,t}$, the M1 innovation, is transformed to remove the common component.

More technically, the errors are orthogonalized by a Cholesky decomposition so that the covariance matrix of the resulting innovations is diagonal—see the Technical Notes, Impulse Response for details. While the Cholesky decomposition is widely used, it is a rather arbitrary method of attributing common effects. You should be aware that changing the order of equations can dramatically change the impulse responses.

Generating Impulse Response Functions from VARs

To obtain the impulse response functions for your VAR, select **Impulse** on the VAR toolbar. You will see the VAR Impulse Responses dialog box. EViews will compute one impulse response function for each innovation and endogenous variable pair. For example, a four variable VAR has 16 potential impulse response functions.

- In the top two edit boxes, you should enter the variables for which you wish to generate innovations, and the variables for which you wish to observe the impulse responses. The order in which you enter these variables only affects the display of results.

- In the bottom edit box, you should specify the ordering of the variables in the VAR—as explained above, there are significant implications of this choice.

- In addition, you should specify the number of periods for which you wish to trace the response function.

- On the right, specify the form of the output, either a table of numbers, separate graphs of each impulse response function, or graphs that compare the responses of each variable to each of the innovations.

- You should also choose between impulse responses and variance decomposition (see Variance Decomposition below).

- Finally, you should make a choice about the standard errors of the response functions. You can choose not to calculate standard errors, to calculate them from the asymptotic analytical formula, or to calculate them by Monte Carlo methods; see the Technical Notes, Impulse Response for further information of these options. For the Monte Carlo option, you should also specify the number of

repetitions.

The **Table** option tabulates each value of the impulse responses. There is a separate table for each endogenous variable which tabulates the responses to all the innovations that you specify. The numbers in parentheses are the standard errors, computed according to your choice of the **Response standard errors** option.

The **Combined response graphs** option displays these tabular results in a set of graphs, with one graph corresponding to each endogenous variable table.

The **Multiple graphs** option displays the impulse response functions with a separate graph for each endogenous variable and response pair:

If you choose to compute the standard errors, EViews displays the plus/minus two standard deviation bands, alongside the impulse responses. The standard error bands will not be displayed if you select the **Combined response graph** option.

Variance Decomposition

Variance decomposition provides a different method of depicting the system dynamics. Impulse response functions trace the effects of a shock to an endogenous variable on the variables in the VAR. By contrast, variance decomposition decomposes variation in an endogenous variable into the component shocks to the endogenous variables in the VAR. The variance decomposition gives information about the relative importance of each random innovation to the variables in the VAR.

See the Technical Notes, Variance Decomposition for additional details.

How to compute variance decomposition from VARs

To obtain the variance decomposition of a VAR, click **Impulse** in the VAR toolbar and choose the **Variance decomposition** option. You should provide the same information as for impulse responses above, except that the choice of innovations is not needed. Note that there are no standard errors for variance decomposition.

The **Table** option displays the variance decomposition in tabular form. EViews displays a separate variance decomposition for each endogenous variable. The column S.E. is the forecast error of the variable for each forecast horizon. The source of this forecast error is the variation in current and future values of the innovations to each endogenous variable in the VAR.

The remaining columns give the percentage of the variance due to each innovation; each row adds up to 100. For the variable that comes first in the VAR ordering, the only source of the one period ahead variation is its own innovation and the first number is always 100 percent. We reiterate that this decomposition of variance depends critically on the ordering of equations.

The **Combined response graphs** option plots the decomposition of each forecast variance as line graphs. The variance decomposition is displayed as separate line graphs with the *y*-axis height measuring the relative importance of each innovation. You can also change the line graphs to stacked lines; double click in the white area of the graph and change the **Graph Type** to **Stacked Lines**. The stacked line graph shows the relative importance of each innovation as the height between each successive line. Note that the last line, corresponding to the innovation that comes last in the VAR ordering, is always flat at 100 because of the adding up constraint.

Granger Causality Tests

You can test Granger causality by running a VAR on the system of equations and testing for zero restrictions on the VAR coefficients (see Group Views, Granger Causality for a discussion of Granger's definition of causality). We recommend, however, that you *not* use a VAR to test for Granger Causality, since EViews provides alternative tools that are either easier to use, or more generally applicable than the standard VAR.

You should perform your Granger Causality test in EViews using a group or a system object. Your choice of approach should depend upon your particular testing situation:

- Displaying the Granger test view of a group object is the easiest way to perform the test (see Group Views, Granger Causality). The primary limitation of this approach is that you are restricted, in constructing the test statistic, to using VARs with a constant as the only exogenous variable.

- The system object allows you to include additional exogenous variables in the test, and to impose cross-equation restrictions on the coefficients of the VAR. It is, however, more difficult to set up the

test using this approach than using the group Granger test view.

Note that if you wish, you could perform the Granger test using single equation methods. For example, suppose you have a VAR(4) of M1, IP, TB3 with a constant and linear trend and you want to test whether the lagged M1's are jointly significant in the IP equation. One way to test this hypothesis is to estimate the IP equation using single equation least squares:

```
ls ip m1(-1 to -4) ip(-1 to -4) tb3(-1 to -4) c @trend(50.1)
```

and from the equation window choose **View/Coefficient Tests/Wald** and type the restrictions in the Wald Test dialog as

```
c(1)=0, c(2)=0, c(3)=0, c(4)=0
```

Note that the estimated OLS equation should be identical to the equation estimated by estimating the VAR.

Obtaining the Coefficients of a VAR

Estimated coefficients of VARs can be accessed by reference to a *two dimensional* array C. To see the correspondence between each element of C and the estimated coefficients, select **View/Representations** from the VAR toolbar.

The first number in C refers to the equation number of the VAR, while the second number in C refers to the variable number in each equation. For example, C(2,3) is the coefficient of the third regressor in the second equation of the VAR. The C(2,3) coefficient of a VAR named VAR01 can then be accessed by the command

```
var01.c(2,3)
```

Vector Error Correction and Cointegration Theory

The finding that many macro time series may contain a unit root has spurred the development of the theory of non-stationary time series analysis. Engle and Granger (1987) pointed out that a linear combination of two or more non-stationary series may be stationary. If such a stationary, or *I(0)*, linear combination exists, the non-stationary (with a unit root), time series are said to be *cointegrated*. The stationary linear combination is called the *cointegrating equation* and may be interpreted as a long-run equilibrium relationship between the variables. For example, consumption and income are likely to be cointegrated. If they were not, then in the long-run consumption might drift above or below income, so that consumers were irrationally spending or piling up savings.

A vector error correction (VEC) model is a restricted VAR that has cointegration restrictions built into the specification, so that it is designed for use with nonstationary series that are known to be cointegrated. The VEC specification restricts the long-run behavior of the endogenous variables to converge to their cointegrating relationships while allowing a wide range of short-run dynamics. The cointegration term is known as the *error correction* term since the deviation from long-run equilibrium is corrected gradually through a series of partial short-run adjustments.

As a simple example, consider a two variable system with one cointegrating equation and no lagged difference terms. The cointegrating equation is

$$y_{2,t} = \beta y_{1,t}$$

and the VEC is

$$\Delta y_{1,t} = \gamma_1 (y_{2,t-1} - \beta y_{1,t-1}) + \epsilon_{1,t}$$
$$\Delta y_{2,t} = \gamma_2 (y_{2,t-1} - \beta y_{1,t-1}) + \epsilon_{2,t}$$

In this simple model, the only right-hand side variable is the error correction term. In long run equilibrium, this term is zero. However, if y_1 and y_2 deviated from long run equilibrium last period, the error correction term is nonzero and each variable adjusts to partially restore the equilibrium relation. The coefficients γ_1 and γ_2 measure the speed of adjustment.

In this model, the two endogenous variables $y_{1,t}$ and $y_{2,t}$ will have nonzero means but the cointegrating equation will have a zero intercept. To keep the example simple, despite the fact that the use of lagged differences is common, we have included no lagged differences on the right-hand side.

If the two endogenous variables $y_{1,t}$ and $y_{2,t}$ have no trend and the cointegrating equations have an intercept, the VEC has the form

$$\Delta y_{1,t} = \gamma_1(y_{2,t-1} - \mu - \beta y_{1,t-1}) + \epsilon_{1,t}$$
$$\Delta y_{2,t} = \gamma_2(y_{2,t-1} - \mu - \beta y_{1,t-1}) + \epsilon_{2,t}$$

Another VEC specification assumes that there are linear trends in the series and a constant in the cointegrating equations, so that it has the form

$$\Delta y_{1,t} = \delta_1 + \gamma_1(y_{2,t-1} - \mu - \beta y_{1,t-1}) + \epsilon_{1,t}$$
$$\Delta y_{2,t} = \delta_2 + \gamma_2(y_{2,t-1} - \mu - \beta y_{1,t-1}) + \epsilon_{2,t}$$

Similarly, there may be a trend in the cointegrating equation, but no separate trends in the two VEC equations. Lastly, if there is a separate linear trend outside the parentheses in each VEC equation, then there is an implicit quadratic trend in the series.

For additional discussion of VAR and VEC models, see Davidson and MacKinnon (1993, pp. 715–730), and Hamilton (1994a, Chapter 19, pp. 571–629).

Testing for Cointegration

Given a group of non-stationary series, we may be interested in determining whether the series are cointegrated, and if they are, in identifying the cointegrating (long-run equilibrium) relationships. EViews implements VAR-based cointegration tests using the methodology developed by Johansen (1991, 1995). Johansen's method is to test the restrictions imposed by cointegration on the unrestricted VAR involving the series.

Johansen's Cointegration Test

Consider a VAR of order p:

$$y_t = A_1 y_{t-1} + \cdots + A_p y_{t-p} + B x_t + \epsilon_t,$$

where y_t is a k-vector of non-stationary I(1) variables, x_t is a d vector of deterministic variables, and ϵ_t is a vector of innovations. We can rewrite the VAR as:

$$\Delta y_t = \Pi y_{t-1} + \sum_{i=1}^{p-1} \Gamma_i \Delta y_{t-i} + B x_t + \epsilon_t,$$

where

$$\Pi = \sum_{i=1}^{p} A_i - I, \quad \Gamma_i = - \sum_{j=i+1}^{p} A_j.$$

Granger's representation theorem asserts that if the coefficient matrix Π has reduced rank $r<k$, then there exist $k \times r$ matrices α and β each with rank r such that $\Pi = \alpha\beta'$ and $\beta'y_t$ is stationary. r is the number of cointegrating relations (the *cointegrating rank*) and each column of β is the cointegrating vector. The elements of α are known as the *adjustment parameters* in the vector error correction model. Johansen's method is to estimate the Π matrix in an unrestricted form, then test whether we can reject the restrictions implied by the reduced rank of Π.

The number of cointegrating relations

If you have k endogenous variables, each of which has one unit root, there can be from zero to k-1 linearly independent, cointegrating relations. If there are no cointegrating relations, standard time series analyses such as the (unrestricted) VAR may be applied to the first-differences of the data. Since there are k separate integrated elements driving the series, levels of the series do not appear in the VAR in this case.

Conversely, if there is one cointegrating equation in the system, then a single linear combination of the levels of the endogenous series $\beta'y_{t-1}$, should be added to each equation in the VAR. When multiplied by a coefficient for an equation, the resulting term $\alpha\beta'y_{t-1}$, is referred to as an error correction term. If there are additional cointegrating equations, each will contribute an additional error correction term involving a different linear combination of the levels of the series.

If there are exactly *k* cointegrating relations, none of the series has a unit root, and the VAR may be specified in terms of the levels of all of the series. Note that in some cases, the individual unit root tests will show that some of the series are integrated, but the Johansen tests show that the cointegrating rank is *k*. This contradiction may be the result of specification error.

The Cointegrating Relations (Vector)

Each column of the β matrix gives an estimate of a cointegrating vector. The cointegrating vector is not identified unless we impose some arbitrary normalization. EViews adopts the normalization so that the *r* cointegrating relations are solved for the first *r* variables in the y_t vector as a function of the remaining *k − r* variables.

Note that one consequence of this normalization is that the normalized vectors which EViews provides will not, in general, be orthogonal, despite the orthogonality of the unnormalized coefficients.

Deterministic Trend Assumptions

Your series may have nonzero means and deterministic trends as well as stochastic trends. Similarly, the cointegrating equations may have intercepts and deterministic trends. The asymptotic distribution of the LR test statistic for the reduced rank test does not have the usual χ^2 distribution and depends on the assumptions made with respect to deterministic trends. EViews provides tests for the following five possibilities considered by Johansen (see Johansen, 1995, pp. 80–84 for details):

1. Series *y* have no deterministic trends and the cointegrating equations do not have intercepts:

$$H_2(r): \ \Pi y_{t-1} + B x_t = \alpha \beta' y_{t-1}$$

2. Series *y* have no deterministic trends and the cointegrating equations have intercepts:

$$H_1^*(r): \ \Pi y_{t-1} + B x_t = \alpha(\beta' y_{t-1} + \rho_0)$$

3. Series *y* have linear trends but the cointegrating equations have only intercepts:

$$H_1(r): \ \Pi y_{t-1} + B x_t = \alpha(\beta' y_{t-1} + \rho_0) + \alpha_\perp \gamma_0$$

4. Both series *y* and the cointegrating equations have linear trends:

$$H^*(r): \ \Pi y_{t-1} + B x_t = \alpha(\beta' y_{t-1} + \rho_0 + \rho_1 t) + \alpha_\perp \gamma_0$$

5. Series *y* have quadratic trends and the cointegrating equations have linear trends:

$$H(r): \ \Pi y_{t-1} + B x_t = \alpha(\beta' y_{t-1} + \rho_0 + \rho_1 t) + \alpha_\perp(\gamma_0 + \gamma_1 t)$$

where $\alpha_\perp$ is the (non-unique) $k \times (k-r)$ matrix such that $\alpha' \alpha_\perp = 0$ and $\text{rank}([\alpha \,|\, \alpha_\perp]) = k$.

These five cases are nested from the most restrictive to the least restrictive, given any particular cointegrating rank *r*:

$$H_2(r) \subset H_1^*(r) \subset H_1(r) \subset H^*(r) \subset H(r)$$

For each case, EViews tabulates the critical values for the reduced rank test as given by Osterwald-Lenum (1992), not those tabulated in Johansen and Juselius (1990). Note that the critical values are available for up to 10 series and may not be appropriate for models that contain other deterministic regressors. For example, a shift dummy variable in the VAR implies a broken linear trend in the *y* series.

If you include 0-1 seasonal dummy variables in the VAR, this will affect both the mean and the trend of the *y* series. Johansen (1995, page 84) suggests using *centered* (orthogonalized) seasonal dummy variables, which only shift the mean without contributing to the trend. Centered seasonal dummy variables for quarterly and monthly series can be generated by the commands

```
series d_q=@seas(q)-1/4
series d_m=@seas(m)-1/12
```

for quarter *q* and month *m*, respectively.

There are two dimensions you can vary while performing tests within this framework. You can assume one of the five cases listed above, and carry out tests for the cointegrating rank. Alternatively,

you can fix the rank, and test which of the five cases describes the data best. These tests are standard χ^2 tests and are described in the Technical Notes, [Nested Tests of Deterministic Trends under Cointegration](). EViews provides you with the option of summarizing all five cases, so you can look at all possible combinations of rank and intercept-trend.

Among the five intercept-trend cases, EViews uses (3) as the default. This choice is based on the assumption that long-run equilibrium conditions (such as the relation between income and consumption) probably do not have trends. In choosing your final model of the data, you should be guided by economic interpretations as well as the statistical criteria. You should look at the normalized cointegrating equations to see if they correspond to your beliefs about long-run relations among the variables. Sometimes it is helpful to rearrange the order of the variables; a reordering does not affect the substantive results.

How to Perform a Johansen Test

To carry out the Johansen test, select **View/Cointegration Test…** on the group or VAR toolbar. Note that since this is a test for cointegration, this test is only valid when you are working with series that are known to be nonstationary. You may wish first to apply unit root tests to each series in the VAR (see [Unit Root Tests]()).

The Cointegration Test dialog asks you to provide information about the test.

Specification

The first step is to choose one of the options specifying the type of deterministic trends that are present in the data. The first five options provide particular alternatives for whether an intercept or trend term should be included in the specification of the cointegrating equations.

The sixth option runs the Johansen procedure under all five sets of assumptions. Since this selection estimates a large number of specifications, only a summary table of results will be displayed. You must select one of the other options to see the full output of the test.

Exogenous Variables

The dialog also allows you to specify exogenous variables, such as seasonal dummies, to include in the test VARs. The constant and linear trend should *not* be listed in this dialog; these should be chosen from the specification options. If you do include exogenous variables, you should be aware that the critical values reported by EViews do not account for their inclusion.

Lag Intervals

You should specify the lags of the test VAR as pairs of intervals. *Note that in contrast to some other statistical packages, the lags are specified as lags of the first differenced terms, not in terms of the levels.* For example, if you type `1 4` in the field, the test VAR regresses Δy_t on Δy_{t-1}, Δy_{t-2}, Δy_{t-3}, Δy_{t-4} and other exogenous variables that you have specified. Some other programs specify the lags in levels; if you type `1 4` in EViews, the highest lag in level is 5. To run a cointegration test with one lag in the *levels*, type `0 0` in the field.

Interpreting the Results of a Johansen Test

As an example, the output for a four-variable system used in [Johansen and Juselius (1990)]() for the Danish data is shown below. The four variables are

LRM:	log real money
LRY:	log real income
IBO:	long term interest rate (bond rate)
IDE:	short term interest rate (deposit rate)

The test assumes no trend in the series with an intercept in the cointegration relation (second specification in the dialog), includes three orthogonalized seasonal dummy variables, and uses two lags in *levels* which is specified as `1 1` in the field box.

The number of cointegrating relations

The first part of the table performs the trace test for the number of cointegrating relations:

Date: 10/16/97   Time: 22:33
Sample: 1974:1 1987:3
Included observations: 53

Test assumption: No deterministic trend in the data
Series: LRM LRY IBO IDE
Exogenous series: D1 D2 D3
Warning: Critical values were derived assuming no exogenous series
Lags interval: 1 to 1

Eigenvalue	Likelihood Ratio	5 Percent Critical Value	1 Percent Critical Value	Hypothesized No. of CE(s)
0.433165	49.14436	53.12	60.16	None
0.177584	19.05691	34.91	41.07	At most 1
0.112791	8.694964	19.96	24.60	At most 2
0.043411	2.352233	9.24	12.97	At most 3

*(**) denotes rejection of the hypothesis at 5%(1%) significance level
L.R. rejects any cointegration at 5% significance level

The eigenvalues are presented in the first column, while the second column (Likelihood Ratio) gives the LR test statistic:

$$Q_r = -T \sum_{i=r+1}^{k} \log(1 - \lambda_i)$$

for $r = 0,1,...,k-1$ where λ_i is the i-th largest eigenvalue. Q_r is the so-called *trace statistic* and is the test of $H_1(r)$ against $H_1(k)$.

To determine the number of cointegrating relations r, subject to the assumptions made about the trends in the series, we can proceed sequentially from $r = 0$ to $r = k-1$ until we fail to reject. The first row in the upper table tests the hypothesis of no cointegration, the second row tests the hypothesis of one cointegrating relation, the third row tests the hypothesis of two cointegrating relations, and so on, all against the alternative hypothesis of full rank, i.e. all series in the VAR are stationary.

The trace statistic does not reject any of the hypotheses at the 5% level. Note that EViews displays the critical values for the trace statistic reported by Osterwald-Lenum (1992), not those tabulated in Johansen and Juselius (1990). Johansen also proposes an alternative LR test statistic, known as the *maximum eigenvalue statistic*, which tests $H_1(r)$ against $H_1(r+1)$. The maximum eigenvalue statistic can be computed from the trace statistic as

$$Q_{max} = -T \log(1 - \lambda_{r+1}) = Q_r - Q_{r+1}$$

EViews does not provide critical values for the maximum eigenvalue statistic; critical values for this statistic are tabulated in Osterwald-Lenum (1992).

The Cointegrating Equations

Below the results of the cointegration rank tests, EViews provides the estimates of the cointegrating vector or relations. Although EViews displays all possible $k-1$ cointegrating relations, you are most likely to be interested in the first r estimates, where r is determined by the LR tests.

The cointegrating vector is not identified unless we impose some arbitrary normalization. EViews adopts a normalization such that the first r series in the y_t vector are normalized to an identity matrix. The normalized cointegrating relation assuming one cointegrating relation $r = 1$ is given by

Normalized Cointegrating Coefficients: 1 Cointegrating Equation(s)

LRM	LRY	IBO	IDE	C
1.000000	-1.032949	5.206919	-4.215880	-6.059932
	(0.13366)	(0.66493)	(1.17916)	(0.83100)

Log likelihood   669.1154

which can be written as

$$LRM - 1.03\,LRY + 5.21\,IBO - 4.22\,IDE - 6.06$$

The number in parentheses under the estimated coefficients are the asymptotic standard errors. Some of the normalized coefficients will be shown without standard errors. This will be the case for

coefficients that are normalized to 1.0 and for coefficients that are not identified. In the latter case, the coefficient (usually 0) is the result of an arbitrary identifying assumption.

The appearance of the normalized cointegrating relations depends on how you order the series in the VAR. For example, if you want a relation with a unit coefficient on LRY, specify the LRY series as the first series in your VAR. Of course, nothing of substance is affected since any linear combination of the cointegrating relations is also a cointegrating relation.

Summary of Five Tests

If you are not sure about the deterministic trends in the data, you may choose the summary of the cointegration tests under all five models. The output displays the log likelihood and the two information criteria under each model as a model selection guide. See Information Criteria for a discussion of the use of information criteria in model selection.

The results of the LR tests shown at the bottom of the table are at the 5% significance level.

Estimating a VEC in EViews

An unrestricted VAR does not assume the presence of cointegration. If you wish to impose cointegrating restrictions among the variables in the VAR, you should use a vector error correction (VEC) model.

As the VEC specification only applies to cointegrated series, you should run the Johansen cointegration test prior to VEC specification as described above. This allows you to confirm that the variables are cointegrated and to determine the number of cointegrating equations.

To set up a VEC, click the **Estimate** button in the VAR toolbar and choose the **Vector Error Correction** specification. You should provide the same information as for an unrestricted VAR and you may optionally specify exogenous variables. However, the specification of the exogenous intercepts and trends should be chosen from the five models discussed above. This choice should be the same as in the cointegration test.

It is important to note that the lag specification that EViews prompts you to enter refers to lags of the first difference terms in the VEC. For example, `1  1` specifies a model involving a regression of the first differences on one lag of the *first difference*.

You must also specify the number of cointegrating equations in the VEC model. This number should be determined from the cointegration test. The maximum number of cointegrating equations is one less than the number of endogenous variables in the VAR.

To estimate the VEC, click **OK**. Estimation of a VEC model proceeds by first determining one or more cointegrating equations using the Johansen procedure. The first difference of each endogenous variable is then regressed on a one period lag of the cointegrating equation(s) and lagged first differences of all of the endogenous variables in the system.

Working with a VEC

Working with a VEC is analogous to working with a VAR, so we refer you to the discussion in Working with a VAR. Note that standard errors for the impulse response functions are not available for the VEC.

Obtaining the Coefficients of a VEC

As explained in the VAR section, estimated coefficients of VARs can be accessed by reference to a two dimensional array C. Similarly, estimated coefficients of the cointegrating vector and the adjustment parameters can be accessed by reference to the two dimensional arrays B and A, respectively.

To see the correspondence between each element of A, B, C and the estimated coefficients, select **View/Representations** from the VAR toolbar.

The first index to A is the equation number of the VEC, while the second index to A is the number of the cointegrating equation. For example, A(2,1) is the adjustment coefficient of the first cointegrating equation in the second equation of the VEC.

The first index to B is the number of the cointegrating equation, while the second index to B is the variable number in the cointegrating equation. For example, B(2,1) is the coefficient of the first variable in the second cointegrating equation.

The first index to C is the equation number of the VEC, while the second index to C is the variable number of the first-differenced regressor of the VEC. For example, C(2,1) is the coefficient of the first first-differenced regressor in the second equation of the VEC.

You can access each element of the coefficient by referring to the name of the VEC followed by a dot and coefficient element:

```
var01.a(2,1)
var01.b(2,1)
var01.c(2,1)
```

Forecasting from a VAR or VEC

To calculate a forecast from a VAR or Vector Error Correction model, click on **Procs/Make Model** on the VAR toolbar. You will see a model window with the VAR or VEC in it. You can make any changes you want, including modifying the ASSIGN statement, and then push the **Solve** button on the model's toolbar to calculate a forecast. See Model Solve for further discussion on how to forecast from models.

Illustration

As an illustration, we estimate a bivariate VAR of the log of personal consumption expenditure (CS) and the log of personal income (INC) using quarterly data over the period 60.1–96.4. The plot of the two series shows that the two series are drifting together, suggesting cointegration. To establish cointegration, we must first check whether each series is integrated and contains a unit root. The results of the augmented Dickey-Fuller test with four lags and a constant and linear trend in the test equation are presented below:

	log(CS)	Dlog(CS)	log(INC)	Dlog(INC)
ADF *t*-stat	–0.52	–3.67	–0.92	–3.78

The 5% critical value is –3.44 and the tests indicate that both series contain a unit root.

Choosing the Lag Order of a VAR

To estimate a VAR of log(CS) and log(INC), click **Quick/Estimate VAR…** or type `var` in the command window. Fill in the dailog to fit a fourth order VAR with a constant as the only exogenous regressor. Click **OK** to estimate the VAR. After estimation, you can select **View/Residual Graphs** to plot the residuals of each equation in the VAR.

The lag order of the VAR is often selected somewhat arbitrarily, with standard recommendations suggesting that you set it long enough to ensure that the residuals are white noise. However, if you choose the lag length too large, the estimates become imprecise. You can use a likelihood ratio (LR) test to test the appropriate lag length. To carry out the LR test, estimate the VAR twice, each with different lags. The estimation results for lag orders 4 (top) and 2 (bottom) are displayed below:

```
Lags interval: 1 to 4
  Determinant Residual Covariance          1.62E-09
  Log Likelihood                           1077.958
  Akaike Information Criteria             -19.99951
  Schwarz Criteria                       -19.63499


Lags interval: 1 to 2
  Determinant Residual Covariance          1.90E-09
  Log Likelihood                           1065.907
  Akaike Information Criteria             -19.94476
  Schwarz Criteria                       -19.74225
```

The LR test statistic for the hypothesis of lag 2 against lag 4 can be computed as

$$LR = -2(\ell_2 - \ell_4) = -2(1065.907 - 1077.958) = 24.102$$

where ℓ is the log likelihood reported at the very bottom of the table. The LR test statistic is

asymptotically distributed x^2 with degrees of freedom equal to the number of restrictions under test. In this case, there are 8 zero restrictions to move from VAR(4) to VAR(2) so the LR test statistic is asymptotically distributed x^2(8). To carry out the test, you can compare the test statistic with the critical values from the x^2 table. Alternatively, you can compute the *p*-value in EViews by entering the commands:

```
scalar pval = 1-@cchisq(24.102,8)
show pval
```

The result 0.002 is displayed on the status line at the bottom of the EViews window and we reject the null hypothesis of 2 lags. The Akaike information criterion also selects VAR(4) in favor of VAR(2) but the Schwarz information criterion selects the parsimonious VAR(2).

Note: to compute the LR test appropriately, you must estimate the two VARs using the same sample period. The VAR with the longer lag will have a shorter sample, so you should set the sample period to the sample estimated by the longer VAR.

Testing for Cointegration

Given that each series has a unit root, we now test whether CS and INC are cointegrated over the sample period and if so, what the cointegrating relation is. The estimated cointegrating relation will give us an estimate of the long run propensity to consume.

To test cointegration, click **View/Cointegration Test…** and fill in the dialog. We use the VAR with lag 4 and assume that the series contain linear trends but the cointegrating relation only includes a constant (the default third model). The results are:

Date: 10/16/97   Time: 23:19
Sample: 1960:1 1996:4
Included observations: 147
Test assumption: Linear deterministic trend in the data
Series: LOG(CS) LOG(INC)
Lags interval: 1 to 4

Eigenvalue	Likelihood Ratio	5 Percent Critical Value	1 Percent Critical Value	Hypothesized No. Of CE(s)
0.145518	25.20440	15.41	20.04	None **
0.014098	2.087119	3.76	6.65	At most 1

*(**) denotes rejection of the hypothesis at 5%(1%) significance level
L.R. test indicates 1 cointegrating equation(s) at 5% significance level

The LR test rejects the hypothesis of no cointegration but not the hypothesis of at most one cointegration relation. The estimated cointegrating relation is:

Normalized Cointegrating Coefficients: 1 Cointegrating Equation(s)

LOG(CS)	LOG(INC)	C
1.000000	-1.002850	0.257338
	(0.00320)	

Log likelihood   1076.042

As expected, the estimated long run marginal propensity to consume is not significantly different from one. You can plot the cointegrating relation by the command

```
plot log(cs)-1.002850*log(inc)+0.257338
```

You can draw the zero line by double clicking in the white area of the graph and choosing **Graph Attributes: Zero line**.