Title stata.com

vecrank — Estimate the cointegrating rank of a VECM

Syntax

`vecrank` *depvarlist* [*if*] [*in*] [, *options*]

options	Description
Model	
`lags(#)`	use # for the maximum lag in underlying VAR model
`trend(constant)`	include an unrestricted constant in model; the default
`trend(rconstant)`	include a restricted constant in model
`trend(trend)`	include a linear trend in the cointegrating equations and a quadratic trend in the undifferenced data
`trend(rtrend)`	include a restricted trend in model
`trend(none)`	do not include a trend or a constant
Adv. model	
`sindicators(`$varlist_{si}$`)`	include normalized seasonal indicator variables $varlist_{si}$
`noreduce`	do not perform checks and corrections for collinearity among lags of dependent variables
Reporting	
`notrace`	do not report the trace statistic
`max`	report maximum-eigenvalue statistic
`ic`	report information criteria
`level99`	report 1% critical values instead of 5% critical values
`levela`	report both 1% and 5% critical values

You must `tsset` your data before using `vecrank`; see [TS] **tsset**.
depvar may contain time-series operators; see **[U] 11.4.4 Time-series varlists**.
`by`, `rolling`, and `statsby` are allowed; see **[U] 11.1.10 Prefix commands**.
`vecrank` does not allow gaps in the data.

Menu

Statistics > Multivariate time series > Cointegrating rank of a VECM

Description

`vecrank` produces statistics used to determine the number of cointegrating equations in a vector error-correction model (VECM).

Options

Model

`lags(#)` specifies the number of lags in the VAR representation of the model. The VECM will include one fewer lag of the first differences. The number of lags must be greater than zero but small enough so that the degrees of freedom used by the model are less than the number of observations.

`trend(`*trend_spec*`)` specifies one of five trend specifications to include in the model. See [TS] **vec intro** and [TS] **vec** for descriptions. The default is `trend(constant)`.

Adv. model

`sindicators(`*varlist*$_{\mathrm{si}}$`)` specifies normalized seasonal indicator variables to be included in the model. The indicator variables specified in this option must be normalized as discussed in Johansen (1995, 84). If the indicators are not properly normalized, the likelihood-ratio–based tests for the number of cointegrating equations do not converge to the asymptotic distributions derived by Johansen. For details, see *Methods and formulas* of [TS] **vec**. `sindicators()` cannot be specified with `trend(none)` or `trend(rconstant)`

`noreduce` causes `vecrank` to skip the checks and corrections for collinearity among the lags of the dependent variables. By default, `vecrank` checks whether the current lag specification causes some of the regressions performed by `vecrank` to contain perfectly collinear variables and reduces the maximum lag until the perfect collinearity is removed. See *Collinearity* in [TS] **vec** for more information.

Reporting

`notrace` requests that the output for the trace statistic not be displayed. The default is to display the trace statistic.

`max` requests that the output for the maximum-eigenvalue statistic be displayed. The default is to not display this output.

`ic` causes the output for the information criteria to be displayed. The default is to not display this output.

`level99` causes the 1% critical values to be displayed instead of the default 5% critical values.

`levela` causes both the 1% and the 5% critical values to be displayed.

Remarks and examples

stata.com

Remarks are presented under the following headings:

Introduction

Before estimating the parameters of a VECM models, you must choose the number of lags in the underlying VAR, the trend specification, and the number of cointegrating equations. `vecrank` offers several ways of determining the number of cointegrating vectors conditional on a trend specification and lag order.

vecrank implements three types of methods for determining r, the number of cointegrating equations in a VECM. The first is Johansen's "trace" statistic method. The second is his "maximum eigenvalue" statistic method. The third method chooses r to minimize an information criterion.

All three methods are based on Johansen's maximum likelihood (ML) estimator of the parameters of a cointegrating VECM. The basic VECM is

$$\Delta\mathbf{y}_t = \boldsymbol{\alpha}\boldsymbol{\beta}'\mathbf{y}_{t-1} + \sum_{t=1}^{p-1}\boldsymbol{\Gamma}_i\Delta\mathbf{y}_{t-i} + \epsilon_t$$

where $\mathbf{y}$ is a $(K \times 1)$ vector of I(1) variables, $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are $(K \times r)$ parameter matrices with rank $r < K$, $\boldsymbol{\Gamma}_1, \ldots, \Gamma_{p-1}$ are $(K \times K)$ matrices of parameters, and ϵ_t is a $(K \times 1)$ vector of normally distributed errors that is serially uncorrelated but has contemporaneous covariance matrix $\boldsymbol{\Omega}$.

Building on the work of Anderson (1951), Johansen (1995) derives an ML estimator for the parameters and two likelihood-ratio (LR) tests for inference on r. These LR tests are known as the trace statistic and the maximum-eigenvalue statistic because the log likelihood can be written as the log of the determinant of a matrix plus a simple function of the eigenvalues of another matrix.

Let $\lambda_1, \ldots, \lambda_K$ be the K eigenvalues used in computing the log likelihood at the optimum. Furthermore, assume that these eigenvalues are sorted from the largest λ_1 to the smallest λ_K. If there are $r < K$ cointegrating equations, $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ have rank r and the eigenvalues $\lambda_{r+1}, \ldots, \lambda_K$ are zero.

The trace statistic

The null hypothesis of the trace statistic is that there are no more than r cointegrating relations. Restricting the number of cointegrating equations to be r or less implies that the remaining $K - r$ eigenvalues are zero. Johansen (1995, chap. 11 and 12) derives the distribution of the trace statistic

$$-T \sum_{i=r+1}^{K} \ln(1 - \widehat{\lambda}_i)$$

where T is the number of observations and the $\widehat{\lambda}_i$ are the estimated eigenvalues. For any given value of r, large values of the trace statistic are evidence against the null hypothesis that there are r or fewer cointegrating relations in the VECM.

One of the problems in determining the number of cointegrating equations is that the process involves more than one statistical test. Johansen (1995, chap. 6, 11, and 12) derives a method based on the trace statistic that has nominal coverage despite evaluating multiple tests. This method can be interpreted as being an estimator $\widehat{r}$ of the true number of cointegrating equations r_0. The method starts testing at $r = 0$ and accepts as $\widehat{r}$ the first value of r for which the trace statistic fails to reject the null.

▷ Example 1

We have quarterly data on the natural logs of aggregate consumption, investment, and GDP in the United States from the first quarter of 1959 through the fourth quarter of 1982. As discussed in King et al. (1991), the balanced-growth hypothesis in economics implies that we would expect to find two cointegrating equations among these three variables. In the output below, we use vecrank to determine the number of cointegrating equations using Johansen's multiple-trace test method.

```
. use http://www.stata-press.com/data/r13/balance2
(macro data for VECM/balance study)
. vecrank y i c, lags(5)
                       Johansen tests for cointegration
Trend: constant                                         Number of obs =      91
Sample:  1960q2 - 1982q4                                         Lags =       5
-------------------------------------------------------------------------------
                                                         5%
maximum                                      trace    critical
  rank    parms       LL       eigenvalue  statistic    value
    0       39    1231.1041          .       46.1492    29.68
    1       44    1245.3882      0.26943     17.5810    15.41
    2       47    1252.5055      0.14480      3.3465*    3.76
    3       48    1254.1787      0.03611
-------------------------------------------------------------------------------
```

The header produces information about the sample, the trend specification, and the number of lags included in the model. The main table contains a separate row for each possible value of r, the number of cointegrating equations. When $r = 3$, all three variables in this model are stationary.

In this example, because the trace statistic at $r = 0$ of 46.1492 exceeds its critical value of 29.68, we reject the null hypothesis of no cointegrating equations. Similarly, because the trace statistic at $r = 1$ of 17.581 exceeds its critical value of 15.41, we reject the null hypothesis that there is one or fewer cointegrating equation. In contrast, because the trace statistic at $r = 2$ of 3.3465 is less than its critical value of 3.76, we cannot reject the null hypothesis that there are two or fewer cointegrating equations. Because Johansen's method for estimating r is to accept as $\widehat{r}$ the first r for which the null hypothesis is not rejected, we accept $r = 2$ as our estimate of the number of cointegrating equations between these three variables. The "*" by the trace statistic at $r = 2$ indicates that this is the value of r selected by Johansen's multiple-trace test procedure. The eigenvalue shown in the last line of output computes the trace statistic in the preceding line.

◁

▷ Example 2

In the previous example, we used the default 5% critical values. We can estimate r with 1% critical values instead by specifying the `level99` option.

```
. vecrank y i c, lags(5) level99
                       Johansen tests for cointegration
Trend: constant                                         Number of obs =      91
Sample:  1960q2 - 1982q4                                         Lags =       5
-------------------------------------------------------------------------------
                                                         1%
maximum                                      trace    critical
  rank    parms       LL       eigenvalue  statistic    value
    0       39    1231.1041          .       46.1492    35.65
    1       44    1245.3882      0.26943     17.5810*   20.04
    2       47    1252.5055      0.14480      3.3465     6.65
    3       48    1254.1787      0.03611
-------------------------------------------------------------------------------
```

The output indicates that switching from the 5% to the 1% level changes the resulting estimate from $r = 2$ to $r = 1$.

◁

The maximum-eigenvalue statistic

The alternative hypothesis of the trace statistic is that the number of cointegrating equations is strictly larger than the number r assumed under the null hypothesis. Instead, we could assume a given r under the null hypothesis and test this against the alternative that there are $r+1$ cointegrating equations. Johansen (1995, chap. 6, 11, and 12) derives an LR test of the null of r cointegrating relations against the alternative of $r+1$ cointegrating relations. Because the part of the log likelihood that changes with r is a simple function of the eigenvalues of a $(K \times K)$ matrix, this test is known as the maximum-eigenvalue statistic. This method is used less often than the trace statistic method because no solution to the multiple-testing problem has yet been found.

▷ Example 3

In the output below, we reexamine the balanced-growth hypothesis. We use the `levela` option to obtain both the 5% and 1% critical values, and we use the `notrace` option to suppress the table of trace statistics.

```
. vecrank y i c, lags(5) max levela notrace

                       Johansen tests for cointegration
Trend: constant                                         Number of obs =      91
Sample:  1960q2 - 1982q4                                         Lags =       5
-------------------------------------------------------------------------------

maximum                                           max      5% critical  1% critical
  rank    parms       LL          eigenvalue  statistic      value        value
    0       39     1231.1041                   28.5682       20.97        25.52
    1       44     1245.3882      0.26943      14.2346       14.07        18.63
    2       47     1252.5055      0.14480       3.3465        3.76         6.65
    3       48     1254.1787      0.03611
-------------------------------------------------------------------------------
```

We can reject $r=1$ in favor of $r=2$ at the 5% level but not at the 1% level. As with the trace statistic method, whether we choose to specify one or two cointegrating equations in our VECM will depend on the significance level we use here.

◁

Minimizing an information criterion

Many multiple-testing problems in the time-series literature have been solved by defining an estimator that minimizes an information criterion with known asymptotic properties. Selecting the lag length in an autoregressive model is probably the best-known example. Gonzalo and Pitarakis (1998) and Aznar and Salvador (2002) have shown that this approach can be applied to determining the number of cointegrating equations in a VECM. As in the lag-length selection problem, choosing the number of cointegrating equations that minimizes either the Schwarz Bayesian information criterion (SBIC) or the Hannan and Quinn information criterion (HQIC) provides a consistent estimator of the number of cointegrating equations.

▷ Example 4

We use these information-criteria methods to estimate the number of cointegrating equations in our balanced-growth data.

```
. vecrank y i c, lags(5) ic notrace
                        Johansen tests for cointegration
Trend: constant                                        Number of obs =      91
Sample:  1960q2 - 1982q4                                        Lags =       5
───────────────────────────────────────────────────────────────────────────────
maximum
  rank    parms       LL       eigenvalue     SBIC       HQIC        AIC
    0      39     1231.1041                -25.12401  -25.76596  -26.20009
    1      44     1245.3882     0.26943    -25.19009  -25.91435  -26.40414
    2      47     1252.5055     0.14480    -25.19781* -25.97144* -26.49463
    3      48     1254.1787     0.03611    -25.18501  -25.97511  -26.50942
───────────────────────────────────────────────────────────────────────────────
```

Both the SBIC and the HQIC estimators suggest that there are two cointegrating equations in the balanced-growth data.

◁

Stored results

`vecrank` stores the following in `e()`:

Scalars

`e(N)`	number of observations
`e(k_eq)`	number of equations in `e(b)`
`e(k_dv)`	number of dependent variables
`e(tmin)`	minimum time
`e(tmax)`	maximum time
`e(n_lags)`	number of lags
`e(k_ce95)`	number of cointegrating equations chosen by multiple trace tests with `level(95)`
`e(k_ce99)`	number of cointegrating equations chosen by multiple trace tests with `level(99)`
`e(k_cesbic)`	number of cointegrating equations chosen by minimizing SBIC
`e(k_cehqic)`	number of cointegrating equations chosen by minimizing HQIC

Macros

`e(cmd)`	`vecrank`
`e(cmdline)`	command as typed
`e(trend)`	trend specified
`e(reduced_lags)`	list of maximum lags to which the model has been reduced
`e(reduce_opt)`	`noreduce`, if `noreduce` is specified
`e(tsfmt)`	format for current time variable

Matrices

`e(max)`	vector of maximum-eigenvalue statistics
`e(trace)`	vector of trace statistics
`e(lambda)`	vector of eigenvalues
`e(k_rank)`	vector of numbers of unconstrained parameters
`e(hqic)`	vector of HQIC values
`e(sbic)`	vector of SBIC values
`e(aic)`	vector of AIC values

Methods and formulas

As shown in *Methods and formulas* of [TS] **vec**, given a lag, trend, and seasonal specification when there are $0 \leq r \leq K$ cointegrating equations, the log likelihood with the Johansen identification restrictions can be written as

$$L = -\frac{1}{2}T\left[K\{\ln(2\pi)+1\}+\ln(|S_{00}|)+\sum_{i=1}^{r}\ln\left(1-\widehat{\lambda}_i\right)\right] \tag{1}$$

where the $(K \times K)$ matrix S_{00} and the eigenvalues $\widehat{\lambda}_i$ are defined in *Methods and formulas* of [TS] **vec**.

The trace statistic compares the null hypothesis that there are r or fewer cointegrating relations with the alternative hypothesis that there are more than r cointegrating equations. Under the alternative hypothesis, the log likelihood is

$$L_A = -\frac{1}{2}T\left[K\left\{\ln(2\pi)+1\right\}+\ln\left(|S_{00}|\right)+\sum_{i=1}^{K}\ln\left(1-\widehat{\lambda}_i\right)\right] \tag{2}$$

Thus the LR test that compares the unrestricted model in (2) with the restricted model in (1) is given by

$$LR_{\text{trace}} = -T\sum_{i=r+1}^{K}\ln\left(1-\widehat{\lambda}_i\right)$$

As discussed by Johansen (1995), the trace statistic has a nonstandard distribution under the null hypothesis because the null hypothesis places restrictions on the coefficients on $\mathbf{y}_{t-1}$, which is assumed to have $K-r$ random-walk components. `vecrank` reports the Osterwald-Lenum (1992) critical values.

The maximum-eigenvalue statistic compares the null model containing r cointegrating relations with the alternative model that has $r+1$ cointegrating relations. Thus using these two values for r in (1) and a few lines of algebra implies that the LR test of this hypothesis is

$$LR_{\max} = -T\ln\left(1-\widehat{\lambda}_{r+1}\right)$$

As for the trace statistic, because this test involves restrictions on the coefficients on a vector of I(1) variables, the test statistic's distribution will be nonstandard. `vecrank` reports the Osterwald-Lenum (1992) critical values.

The formulas for the AIC, SBIC, and HQIC are given in *Methods and formulas* of [TS] **vec**.

Søren Johansen (1939–) earned degrees in mathematical statistics at the University of Copenhagen, where he is now based. In addition to making contributions to mathematical statistics, probability theory, and medical statistics, he has worked mostly in econometrics—in particular, on the theory of cointegration.

References

Anderson, T. W. 1951. Estimating linear restrictions on regression coefficients for multivariate normal distributions. *Annals of Mathematical Statistics* 22: 327–351.

Aznar, A., and M. Salvador. 2002. Selecting the rank of the cointegration space and the form of the intercept using an information criterion. *Econometric Theory* 18: 926–947.

Engle, R. F., and C. W. J. Granger. 1987. Co-integration and error correction: Representation, estimation, and testing. *Econometrica* 55: 251–276.

Gonzalo, J., and J.-Y. Pitarakis. 1998. Specification via model selection in vector error correction models. *Economics Letters* 60: 321–328.

Hamilton, J. D. 1994. *Time Series Analysis*. Princeton: Princeton University Press.

Hubrich, K., H. Lütkepohl, and P. Saikkonen. 2001. A review of systems cointegration tests. *Econometric Reviews* 20: 247–318.

Johansen, S. 1988. Statistical analysis of cointegration vectors. *Journal of Economic Dynamics and Control* 12: 231–254.

——. 1991. Estimation and hypothesis testing of cointegration vectors in Gaussian vector autoregressive models. *Econometrica* 59: 1551–1580.

——. 1995. *Likelihood-Based Inference in Cointegrated Vector Autoregressive Models*. Oxford: Oxford University Press.

King, R. G., C. I. Plosser, J. H. Stock, and M. W. Watson. 1991. Stochastic trends and economic fluctuations. *American Economic Review* 81: 819–840.

Lütkepohl, H. 2005. *New Introduction to Multiple Time Series Analysis*. New York: Springer.

Maddala, G. S., and I.-M. Kim. 1998. *Unit Roots, Cointegration, and Structural Change*. Cambridge: Cambridge University Press.

Osterwald-Lenum, M. G. 1992. A note with quantiles of the asymptotic distribution of the maximum likelihood cointegration rank test statistics. *Oxford Bulletin of Economics and Statistics* 54: 461–472.

Park, J. Y., and P. C. B. Phillips. 1988. Statistical inference in regressions with integrated processes: Part I. *Econometric Theory* 4: 468–497.

——. 1989. Statistical inference in regressions with integrated processes: Part II. *Econometric Theory* 5: 95–131.

Phillips, P. C. B. 1986. Understanding spurious regressions in econometrics. *Journal of Econometrics* 33: 311–340.

Phillips, P. C. B., and S. N. Durlauf. 1986. Multiple time series regressions with integrated processes. *Review of Economic Studies* 53: 473–495.

Sims, C. A., J. H. Stock, and M. W. Watson. 1990. Inference in linear time series models with some unit roots. *Econometrica* 58: 113–144.

Stock, J. H. 1987. Asymptotic properties of least squares estimators of cointegrating vectors. *Econometrica* 55: 1035–1056.

Stock, J. H., and M. W. Watson. 1988. Testing for common trends. *Journal of the American Statistical Association* 83: 1097–1107.

Watson, M. W. 1994. Vector autoregressions and cointegration. In Vol. 4 of *Handbook of Econometrics*, ed. R. F. Engle and D. L. McFadden. Amsterdam: Elsevier.

Also see

[TS] **tsset** — Declare data to be time-series data

[TS] **vec** — Vector error-correction models

[TS] **vec intro** — Introduction to vector error-correction models