VAR, SVAR and VECM models

Christopher F Baum

ECON 8823: Applied Econometrics

Boston College, Spring 2015

Vector autoregressive (VAR) models

A p-th order vector autoregression, or $VAR(p)$, with exogenous variables x can be written as:

$$\mathbf{y}_t = \mathbf{v} + \mathbf{A}_1 \mathbf{y}_{t-1} + \cdots + \mathbf{A}_p \mathbf{y}_{t-p} + \mathbf{B}_0 \mathbf{x}_t + \mathbf{B}_1 \mathbf{x}_{t-1} + \cdots + \mathbf{B}_s \mathbf{x}_{t-s} + \mathbf{u}_t$$

where $\mathbf{y}_t$ is a vector of K variables, each modeled as function of p lags of those variables and, optionally, a set of exogenous variables $\mathbf{x}_t$.

We assume that $E(\mathbf{u}_t) = 0$, $E(\mathbf{u}_t \mathbf{u}_t') = \Sigma$ and $E(\mathbf{u}_t \mathbf{u}_s') = 0 \; \forall t \neq s$.

Vector autoregressive (VAR) models

A p-th order vector autoregression, or $VAR(p)$, with exogenous variables x can be written as:

$$\mathbf{y}_t = \mathbf{v} + \mathbf{A}_1 \mathbf{y}_{t-1} + \cdots + \mathbf{A}_p \mathbf{y}_{t-p} + \mathbf{B}_0 \mathbf{x}_t + \mathbf{B}_1 \mathbf{x}_{t-1} + \cdots + \mathbf{B}_s \mathbf{x}_{t-s} + \mathbf{u}_t$$

where $\mathbf{y}_t$ is a vector of K variables, each modeled as function of p lags of those variables and, optionally, a set of exogenous variables $\mathbf{x}_t$.

We assume that $E(\mathbf{u}_t) = 0$, $E(\mathbf{u}_t \mathbf{u}_t') = \Sigma$ and $E(\mathbf{u}_t \mathbf{u}_s') = 0 \; \forall t \neq s$.

If the VAR is stable (see command `varstable`) we can rewrite the VAR in moving average form as:

$$\mathbf{y}_t = \mu + \sum_{i=0}^{\infty} \mathbf{D}_i \mathbf{x}_{t-i} + \sum_{i=0}^{\infty} \Phi_i \mathbf{u}_{t-i}$$

which is the vector moving average (VMA) representation of the VAR, where all past values of y_t have been substituted out. The $\mathbf{D}_i$ matrices are the dynamic multiplier functions, or transfer functions. The sequence of moving average coefficients Φ_i are the simple impulse-response functions (IRFs) at horizon i.

Estimation of the parameters of the VAR requires that the variables in $\mathbf{y}_t$ and $\mathbf{x}_t$ are covariance stationary, with their first two moments finite and time-invariant. If the variables in $\mathbf{y}_t$ are not covariance stationary, but their first differences are, they may be modeled with a vector error correction model, or VECM.

In the absence of exogenous variables, the disturbance variance-covariance matrix Σ contains all relevant information about contemporaneous correlation among the variables in $\mathbf{y}_t$. VARs may be *reduced-form* VARs, which do not account for this contemporaneous correlation. They may be *recursive* VARs, where the K variables are assumed to form a recursive dynamic structural model where each variable only depends upon those above it in the vector $\mathbf{y}_t$. Or, they may be *structural* VARs, where theory is used to place restrictions on the contemporaneous correlations.

Estimation of the parameters of the VAR requires that the variables in $\mathbf{y}_t$ and $\mathbf{x}_t$ are covariance stationary, with their first two moments finite and time-invariant. If the variables in $\mathbf{y}_t$ are not covariance stationary, but their first differences are, they may be modeled with a vector error correction model, or VECM.

In the absence of exogenous variables, the disturbance variance-covariance matrix Σ contains all relevant information about contemporaneous correlation among the variables in $\mathbf{y}_t$. VARs may be *reduced-form* VARs, which do not account for this contemporaneous correlation. They may be *recursive* VARs, where the K variables are assumed to form a recursive dynamic structural model where each variable only depends upon those above it in the vector $\mathbf{y}_t$. Or, they may be *structural* VARs, where theory is used to place restrictions on the contemporaneous correlations.

Stata has a complete suite of commands for fitting and forecasting vector autoregressive (VAR) models and structural vector autoregressive (SVAR) models. Its capabilities include estimating and interpreting impulse response functions (IRFs), dynamic multipliers, and forecast error vector decompositions (FEVDs).

Subsidiary commands allow you to check the stability condition of VAR or SVAR estimates; to compute lag-order selection statistics for VARs; to perform pairwise Granger causality tests for VAR estimates; and to test for residual autocorrelation and normality in the disturbances of VARs.

Dynamic forecasts may be computed and graphed after VAR or SVAR estimation.

Stata has a complete suite of commands for fitting and forecasting vector autoregressive (VAR) models and structural vector autoregressive (SVAR) models. Its capabilities include estimating and interpreting impulse response functions (IRFs), dynamic multipliers, and forecast error vector decompositions (FEVDs).

Subsidiary commands allow you to check the stability condition of VAR or SVAR estimates; to compute lag-order selection statistics for VARs; to perform pairwise Granger causality tests for VAR estimates; and to test for residual autocorrelation and normality in the disturbances of VARs.

Dynamic forecasts may be computed and graphed after VAR or SVAR estimation.

Stata has a complete suite of commands for fitting and forecasting vector autoregressive (VAR) models and structural vector autoregressive (SVAR) models. Its capabilities include estimating and interpreting impulse response functions (IRFs), dynamic multipliers, and forecast error vector decompositions (FEVDs).

Subsidiary commands allow you to check the stability condition of VAR or SVAR estimates; to compute lag-order selection statistics for VARs; to perform pairwise Granger causality tests for VAR estimates; and to test for residual autocorrelation and normality in the disturbances of VARs.

Dynamic forecasts may be computed and graphed after VAR or SVAR estimation.

Stata's `varbasic` command allows you to fit a simple reduced-form VAR without constraints and graph the impulse-response functions (IRFs). The more general `var` command allows for constraints to be placed on the coefficients.

The `varsoc` command allows you to select the appropriate lag order for the VAR; command `varwle` computes Wald tests to determine whether certain lags can be excluded; `varlmar` checks for autocorrelation in the disturbances; and `varstable` checks whether the stability conditions needed to compute IRFs and FEVDs are satisfied.

Stata's `varbasic` command allows you to fit a simple reduced-form VAR without constraints and graph the impulse-response functions (IRFs). The more general `var` command allows for constraints to be placed on the coefficients.

The `varsoc` command allows you to select the appropriate lag order for the VAR; command `varwle` computes Wald tests to determine whether certain lags can be excluded; `varlmar` checks for autocorrelation in the disturbances; and `varstable` checks whether the stability conditions needed to compute IRFs and FEVDs are satisfied.

IRFs, OIRFs and FEVDs

Impulse response functions, or IRFs, measure the effects of a shock to an endogenous variable on itself or on another endogenous variable. Stata's `irf` commands can compute five types of IRFs: *simple* IRFs, *orthogonalized* IRFs, *cumulative* IRFs, *cumulative orthogonalized* IRFs and *structural* IRFs. We defined the simple IRF in an earlier slide.

The forecast error variance decomposition (FEVD) measures the fraction of the forecast error variance of an endogenous variable that can be attributed to orthogonalized shocks to itself or to another endogenous variable.

IRFs, OIRFs and FEVDs

Impulse response functions, or IRFs, measure the effects of a shock to an endogenous variable on itself or on another endogenous variable. Stata's `irf` commands can compute five types of IRFs: *simple* IRFs, *orthogonalized* IRFs, *cumulative* IRFs, *cumulative orthogonalized* IRFs and *structural* IRFs. We defined the simple IRF in an earlier slide.

The forecast error variance decomposition (FEVD) measures the fraction of the forecast error variance of an endogenous variable that can be attributed to orthogonalized shocks to itself or to another endogenous variable.

To analyze IRFs and FEVDs in Stata, you estimate a VAR model and use `irf create` to estimate the IRFs and FEVDs and store them in a file. This step is done automatically by the `varbasic` command, but must be done explicitly after the `var` or `svar` commands. You may then use `irf graph`, `irf table` or other `irf` analysis commands to examine results.

For IRFs to be computed, the VAR must be stable. The simple IRFs shown above have a drawback: they give the effect over time of a one-time unit increase to one of the shocks, holding all else constant. But to the extent the shocks are contemporaneously correlated, the other shocks cannot be held constant, and the VMA form of the VAR cannot have a causal interpretation.

To analyze IRFs and FEVDs in Stata, you estimate a VAR model and use `irf create` to estimate the IRFs and FEVDs and store them in a file. This step is done automatically by the `varbasic` command, but must be done explicitly after the `var` or `svar` commands. You may then use `irf graph`, `irf table` or other `irf` analysis commands to examine results.

For IRFs to be computed, the VAR must be stable. The simple IRFs shown above have a drawback: they give the effect over time of a one-time unit increase to one of the shocks, holding all else constant. But to the extent the shocks are contemporaneously correlated, the other shocks cannot be held constant, and the VMA form of the VAR cannot have a causal interpretation.

Orthogonalized innovations

We can overcome this difficulty by taking $E(u_t u_t') = \Sigma$, the covariance matrix of shocks, and finding a matrix $\mathbf{P}$ such that $\Sigma = \mathbf{P}\mathbf{P}'$ and $\mathbf{P}^{-1}\Sigma\mathbf{P}'^{-1} = \mathbf{I}_K$. The vector of shocks may then be *orthogonalized* by $\mathbf{P}^{-1}$. For a pure VAR, without exogenous variables,

$$\begin{aligned}
\mathbf{y}_t &= \mu + \sum_{i=0}^{\infty} \Phi_i u_{t-i} \\
&= \mu + \sum_{i=0}^{\infty} \Phi_i \mathbf{P}\mathbf{P}^{-1} u_{t-i} \\
&= \mu + \sum_{i=0}^{\infty} \Theta_i \mathbf{P}^{-1} u_{t-i} \\
&= \mu + \sum_{i=0}^{\infty} \Theta_i w_{t-i}
\end{aligned}$$

Sims (*Econometrica*, 1980) suggests that **P** can be written as the Cholesky decomposition of Σ^{-1}, and IRFs based on this choice are known as the *orthogonalized* IRFs. As a VAR can be considered to be the reduced form of a dynamic structural equation (DSE) model, choosing **P** is equivalent to imposing a recursive structure on the corresponding DSE model. The *ordering* of the recursive structure is that imposed in the Cholesky decomposition, which is that in which the endogenous variables appear in the VAR estimation.

As this choice is somewhat arbitrary, you may want to explore the OIRFs resulting from a different ordering. It is not necessary, using `var` and `irf create`, to reestimate the VAR with a different ordering, as the `order()` option of `irf create` will apply the Cholesky decomposition in the specified order.

Just as the OIRFs are sensitive to the ordering of variables, the FEVDs are defined in terms of a particular causal ordering.

If there are additional (strictly) exogenous variables in the VAR, the dynamic multiplier functions or transfer functions can be computed. These measure the impact of a unit change in the exogenous variable on the endogenous variables over time. They are generated by `fcast compute` and graphed with `fcast graph`.

As this choice is somewhat arbitrary, you may want to explore the OIRFs resulting from a different ordering. It is not necessary, using `var` and `irf create`, to reestimate the VAR with a different ordering, as the `order()` option of `irf create` will apply the Cholesky decomposition in the specified order.

Just as the OIRFs are sensitive to the ordering of variables, the FEVDs are defined in terms of a particular causal ordering.

If there are additional (strictly) exogenous variables in the VAR, the dynamic multiplier functions or transfer functions can be computed. These measure the impact of a unit change in the exogenous variable on the endogenous variables over time. They are generated by `fcast compute` and graphed with `fcast graph`.

As this choice is somewhat arbitrary, you may want to explore the OIRFs resulting from a different ordering. It is not necessary, using `var` and `irf create`, to reestimate the VAR with a different ordering, as the `order()` option of `irf create` will apply the Cholesky decomposition in the specified order.

Just as the OIRFs are sensitive to the ordering of variables, the FEVDs are defined in terms of a particular causal ordering.

If there are additional (strictly) exogenous variables in the VAR, the dynamic multiplier functions or transfer functions can be computed. These measure the impact of a unit change in the exogenous variable on the endogenous variables over time. They are generated by `fcast compute` and graphed with `fcast graph`.

varbasic

For a simple VAR estimation, you need only specify the `varbasic` *varlist* command. The number of lags, which is given as a *numlist*, defaults to `(1 2)`. Note that you must list every lag to be included; for instance `lags(4)` would only include the fourth lag, whereas `lags(1/4)` would include the first four lags.

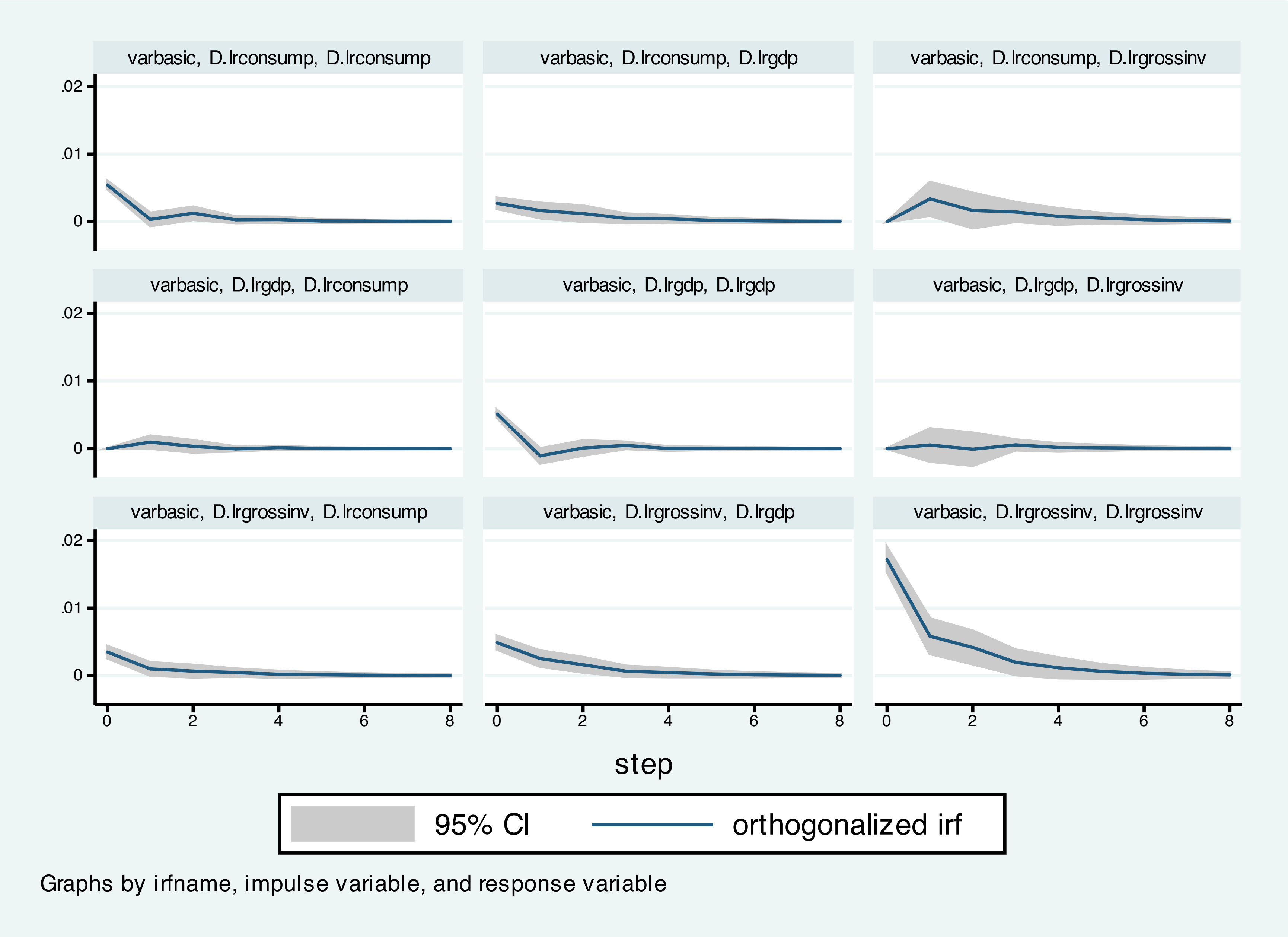
Using the `usmacro1` dataset, let us estimate a basic VAR for the first differences of log real investment, log real consumption and log real income through 2005q4. By default, the command will produce a graph of the orthogonalized IRFs (OIRFs) for 8 steps ahead. You may choose a different horizon with the `step( )` option.

varbasic

For a simple VAR estimation, you need only specify the `varbasic` *varlist* command. The number of lags, which is given as a *numlist*, defaults to `(1 2)`. Note that you must list every lag to be included; for instance `lags(4)` would only include the fourth lag, whereas `lags(1/4)` would include the first four lags.

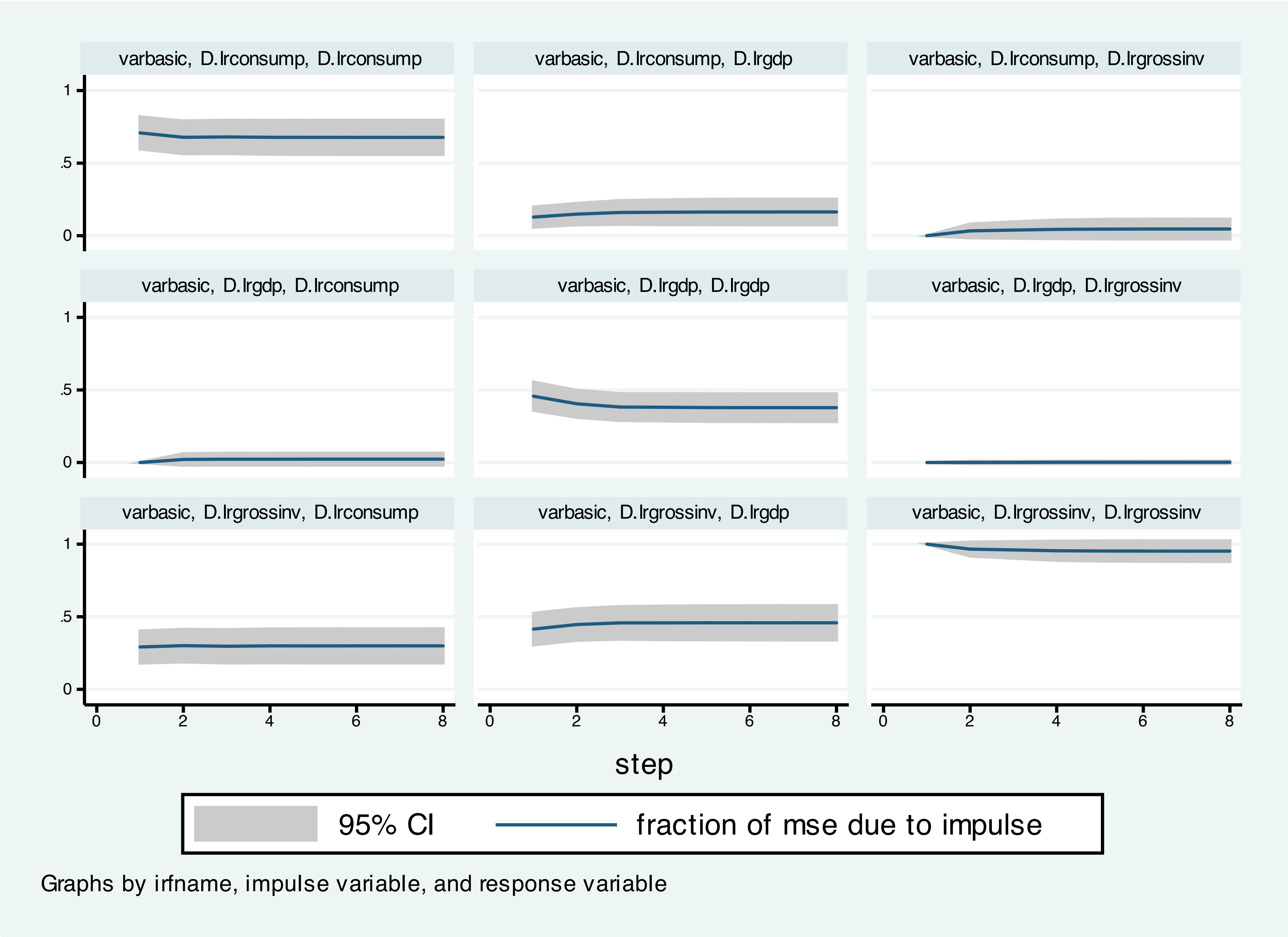
Using the `usmacro1` dataset, let us estimate a basic VAR for the first differences of log real investment, log real consumption and log real income through 2005q4. By default, the command will produce a graph of the orthogonalized IRFs (OIRFs) for 8 steps ahead. You may choose a different horizon with the `step( )` option.

```
 . use usmacro1
. varbasic D.lrgrossinv D.lrconsump D.lrgdp if tin(,2005q4)
Vector autoregression
Sample:  1959q4 - 2005q4                       No. of obs      =       185
Log likelihood =  1905.169                     AIC             =  -20.3694
FPE            =  2.86e-13                     HQIC            = -20.22125
Det(Sigma_ml)  =  2.28e-13                     SBIC            = -20.00385
Equation           Parms      RMSE     R-sq      chi2     P>chi2
----------------------------------------------------------------
D_lrgrossinv          7     .017503   0.2030   47.12655   0.0000
D_lrconsump           7     .006579   0.0994   20.42492   0.0023
D_lrgdp               7     .007722   0.2157   50.88832   0.0000
----------------------------------------------------------------

------------------------------------------------------------------------------
             |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
D_lrgrossinv |
  lrgrossinv |
         LD. |   .1948761   .0977977     1.99   0.046     .0031962    .3865561
        L2D. |   .1271815   .0981167     1.30   0.195    -.0651237    .3194868
             |
   lrconsump |
         LD. |   .5667047   .2556723     2.22   0.027     .0655963    1.067813
        L2D. |   .1771756   .2567412     0.69   0.490     -.326028    .6803791
             |
       lrgdp |
         LD. |   .1051089   .2399165     0.44   0.661    -.3651189    .5753367
        L2D. |  -.1210883   .2349968    -0.52   0.606    -.5816736    .3394969
             |
```


varbasic, D.lrconsump, D.lrconsump
varbasic, D.lrconsump, D.lrgdp
varbasic, D.lrconsump, D.lrgrossinv
varbasic, D.lrgdp, D.lrconsump
varbasic, D.lrgdp, D.lrgdp
varbasic, D.lrgdp, D.lrgrossinv
varbasic, D.lrgrossinv, D.lrconsump
varbasic, D.lrgrossinv, D.lrgdp
varbasic, D.lrgrossinv, D.lrgrossinv
.02
.01
0
0
2
4
6
8
step
95% CI
orthogonalized irf
Graphs by irfname, impulse variable, and response variable

As any of the VAR estimation commands save the estimated IRFs, OIRFs and FEVDs in an `.irf` file, you may examine the FEVDs with a graph command. These items may also be tabulated with the `irf table` and `irf ctable` commands. The latter command allows you to juxtapose tabulated values, such as the OIRF and FEVD for a particular pair of variables, while the `irf cgraph` command allows you to do the same for graphs.

```
. irf graph fevd, lstep(1)
```


varbasic, D.lrconsump, D.lrconsump
varbasic, D.lrconsump, D.lrgdp
varbasic, D.lrconsump, D.lrgrossinv
varbasic, D.lrgdp, D.lrconsump
varbasic, D.lrgdp, D.lrgdp
varbasic, D.lrgdp, D.lrgrossinv
varbasic, D.lrgrossinv, D.lrconsump
varbasic, D.lrgrossinv, D.lrgdp
varbasic, D.lrgrossinv, D.lrgrossinv
0
.5
1
0
2
4
6
8
step
95% CI
fraction of mse due to impulse
Graphs by irfname, impulse variable, and response variable

After producing any graph in Stata, you may save it in Stata's internal format using `graph save` *filename*. This will create a `.gph` file which may be accessed with `graph use`. The file contains all the information necessary to replicate the graph and modify its appearance. However, only Stata can read `.gph` files. If you want to reproduce the graph in a document, use the `graph export` *filename*.*format* command, where *format* is `.eps` or `.pdf`.

We now consider a model fit with `var` to the same three variables, adding the change in the log of the real money base as an exogenous variable. We include four lags in the VAR.

```
. var D.lrgrossinv D.lrconsump D.lrgdp if tin(,2005q4), ///
> lags(1/4) exog(D.lrmbase)
Vector autoregression

Sample:  1960q2 - 2005q4                        No. of obs      =        183
Log likelihood =  1907.061                      AIC             =  -20.38318
FPE            =  2.82e-13                      HQIC            =   -20.0846
Det(Sigma_ml)  =  1.78e-13                      SBIC            =  -19.64658

Equation           Parms      RMSE     R-sq      chi2     P>chi2
----------------------------------------------------------------
D_lrgrossinv         14     .017331   0.2426   58.60225   0.0000
D_lrconsump          14     .006487   0.1640   35.90802   0.0006
D_lrgdp              14     .007433   0.2989   78.02177   0.0000
----------------------------------------------------------------

------------------------------------------------------------------------------
             |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
D_lrgrossinv |
  lrgrossinv |
         LD. |   .2337044   .0970048     2.41   0.016     .0435785    .4238303
        L2D. |   .0746063   .0997035     0.75   0.454    -.1208089    .2700215
        L3D. |  -.1986633   .1011362    -1.96   0.049    -.3968866   -.0004401
        L4D. |   .1517106   .1004397     1.51   0.131    -.0451476    .3485688
             |
   lrconsump |
         LD. |   .4716336   .2613373     1.80   0.071     -.040578    .9838452
        L2D. |   .1322693   .2758129     0.48   0.632     -.408314    .6728527
        L3D. |   .2471462   .2697096     0.92   0.359     -.281475    .7757673
        L4D. |  -.0177416   .2558472    -0.07   0.945    -.5191928    .4837097
             |
         LD. |   .1354875   .2455182     0.55   0.581    -.3457193    .6166942
```

To evaluate whether the money base variable should be included in the VAR, we can use `testparm` to construct a joint test of significance of its coefficients:

```
. testparm D.lrmbase
( 1)  [D_lrgrossinv]D.lrmbase = 0
( 2)  [D_lrconsump]D.lrmbase = 0
( 3)  [D_lrgdp]D.lrmbase = 0
           chi2(  3) =    7.95
         Prob > chi2 =    0.0471
```

The variable is marginally significant in the estimated system.

A common diagnostic from a VAR are the set of block *F* tests, or Granger causality tests, that consider whether each variable plays a significant role in each of the equations. These tests may help to establish a sensible causal ordering. They can be performed by `vargranger`:

```
. vargranger
  Granger causality Wald tests
```

Equation	Excluded	chi2	df	Prob > chi2
D_lrgrossinv	D.lrconsump	4.2531	4	0.373
D_lrgrossinv	D.lrgdp	1.0999	4	0.894
D_lrgrossinv	ALL	10.34	8	0.242
D_lrconsump	D.lrgrossinv	5.8806	4	0.208
D_lrconsump	D.lrgdp	8.1826	4	0.085
D_lrconsump	ALL	12.647	8	0.125
D_lrgdp	D.lrgrossinv	22.204	4	0.000
D_lrgdp	D.lrconsump	11.349	4	0.023
D_lrgdp	ALL	42.98	8	0.000

We may also want to compute selection order criteria to gauge whether we have included sufficient lags in the VAR. Introducing too many lags wastes degrees of freedom, while too few lags leave the equations potentially misspecified and are likely to cause autocorrelation in the residuals. The `varsoc` command will produce selection order criteria, and highlight the optimal lag.

```
. varsoc
  Selection-order criteria
  Sample:  1960q2 - 2005q4                              Number of obs      =      183
```

lag	LL	LR	df	p	FPE	AIC	HQIC	SBIC
0	1851.22				3.5e-13	-20.1663	-20.1237	-20.0611
1	1887.29	72.138*	9	0.000	2.6e-13*	-20.4622*	-20.3555*	-20.1991*
2	1894.14	13.716	9	0.133	2.7e-13	-20.4387	-20.2681	-20.0178
3	1902.58	16.866	9	0.051	2.7e-13	-20.4325	-20.1979	-19.8538
4	1907.06	8.9665	9	0.440	2.8e-13	-20.3832	-20.0846	-19.6466

```
  Endogenous:  D.lrgrossinv D.lrconsump D.lrgdp
   Exogenous:  D.lrmbase  _cons
```

We should also be concerned with stability of the VAR, which requires the moduli of the eigenvalues of the dynamic matrix to lie within the unit circle. As there is more than one lag in the VAR we have estimated, it is likely that complex eigenvalues, leading to cycles, will be encountered.

```
. varstable
  Eigenvalue stability condition
```

Eigenvalue	Modulus
.6916791	.691679
-.5793137 + .1840599*i*	.607851
-.5793137 - .1840599*i*	.607851
-.3792302 + .4714717*i*	.605063
-.3792302 - .4714717*i*	.605063
.1193592 + .5921967*i*	.604106
.1193592 - .5921967*i*	.604106
.5317127 + .2672997*i*	.59512
.5317127 - .2672997*i*	.59512
-.4579249	.457925
.1692559 + .3870966*i*	.422482
.1692559 - .3870966*i*	.422482

```
  All the eigenvalues lie inside the unit circle.
  VAR satisfies stability condition.
```

A digression on interpreting the eigenvalues and their moduli: the complex number $\lambda = a + b\,\imath$ with modulus $|\lambda| = \sqrt{a^2 + b^2}$ can be expressed in polar coordinates as $|\lambda| \exp(\imath\,\theta)$, where θ is the angle (in radians) of the line segment $a + b\,\imath$. Note that $\exp(\imath\,\theta) = \cos(\theta) + \imath\,\sin(\theta)$, a periodic function.

The period of this function will be $\frac{2\pi}{\theta}$ time units. For θ we can substitute $\text{atan2}(a, b)$ where $\text{atan2}(\cdot)$ is the variation on the arctangent function available in most programming languages (be careful with the order of arguments, though).

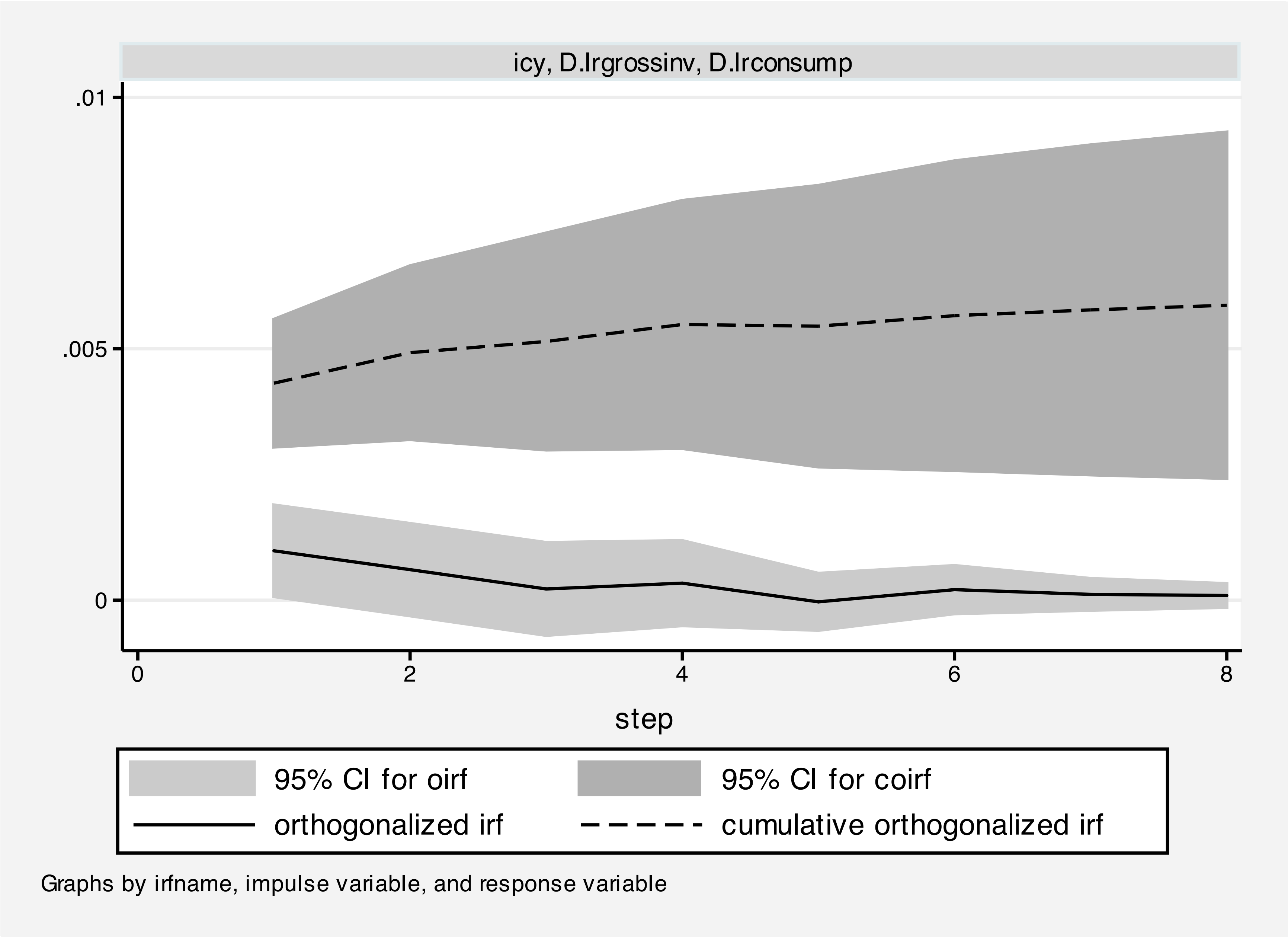
Thus, for the first complex conjugate pair, $-0.579 \pm 0.1841\,\imath$, we have periodicity of 2.217 quarters. For the second, $-0.379 \pm 0.471\,\imath$, we have 2.795 quarters. For the third, $0.119 \pm 0.592\,\imath$, we have 4.580 quarters, and so on.

As the estimated VAR appears stable, we can produce IRFs and FEVDs in tabular or graphical form:

```
 . irf create icy, step(8) set(res1)
(file res1.irf created)
(file res1.irf now active)
(file res1.irf updated)
. irf table oirf coirf, impulse(D.lrgrossinv) response(D.lrconsump) noci stderr
> or
                    Results from icy
```

step	(1) oirf	(1) S.E.	(1) coirf	(1) S.E.
0	.003334	.000427	.003334	.000427
1	.000981	.000465	.004315	.000648
2	.000607	.000468	.004922	.000882
3	.000223	.000471	.005145	.001101
4	.000338	.000431	.005483	.001258
5	-.000034	.000289	.005449	.001428
6	.000209	.000244	.005658	.001571
7	.000115	.000161	.005773	.001674
8	.000092	.00012	.005865	.001757

```
(1) irfname = icy, impulse = D.lrgrossinv, and response = D.lrconsump
. irf graph oirf coirf, impulse(D.lrgrossinv) response(D.lrconsump) ///
>     lstep(1) scheme(s2mono)
```


icy, D.lrgrossinv, D.lrconsump
.01
.005
0
0
2
4
6
8
step
95% CI for oirf
95% CI for coirf
orthogonalized irf
cumulative orthogonalized irf
Graphs by irfname, impulse variable, and response variable

Structural VAR estimation

All of the capabilities we have illustrated for reduced-form VARs are also available for *structural* VARs, which are estimated with the `svar` command. In the SVAR framework, the orthogonalization matrix **P** is not constructed manually as the Cholesky decomposition of the error covariance matrix. Instead, restrictions are placed on the **P** matrix, either in terms of short-run restrictions on the contemporaneous covariances between shocks, or in terms of restrictions on the long-run accumulated effects of the shocks.

Short-run SVAR models

A short-run SVAR model without exogenous variables can be written as

$$\mathbf{A}(\mathbf{I}_K - \mathbf{A}_1 L - \mathbf{A}_2 L^2 - \cdots - \mathbf{A}_p L^p)\mathbf{y}_t = \mathbf{A}\epsilon_t = \mathbf{B}\mathbf{e}_t$$

where L is the lag operator. The vector ϵ_t refers to the original shocks in the model, with covariance matrix Σ, while the vector $\mathbf{e}_t$ are a set of orthogonalized disturbances with covariance matrix $\mathbf{I}_K$.

In a short-run SVAR, we obtain identification by placing restrictions on the matrices $\mathbf{A}$ and $\mathbf{B}$, which are assumed to be nonsingular. The orthgonalization matrix $\mathbf{P}_{sr} = \mathbf{A}^{-1}\mathbf{B}$ is then related to the error covariance matrix by $\Sigma = \mathbf{P}_{sr}\mathbf{P}_{sr}'$.

Short-run SVAR models

A short-run SVAR model without exogenous variables can be written as

$$\mathbf{A}(\mathbf{I}_K - \mathbf{A}_1 L - \mathbf{A}_2 L^2 - \cdots - \mathbf{A}_p L^p)\mathbf{y}_t = \mathbf{A}\epsilon_t = \mathbf{B}\mathbf{e}_t$$

where L is the lag operator. The vector ϵ_t refers to the original shocks in the model, with covariance matrix Σ, while the vector $\mathbf{e}_t$ are a set of orthogonalized disturbances with covariance matrix $\mathbf{I}_K$.

In a short-run SVAR, we obtain identification by placing restrictions on the matrices $\mathbf{A}$ and $\mathbf{B}$, which are assumed to be nonsingular. The orthgonalization matrix $\mathbf{P}_{sr} = \mathbf{A}^{-1}\mathbf{B}$ is then related to the error covariance matrix by $\Sigma = \mathbf{P}_{sr}\mathbf{P}_{sr}'$.

As there are $K(K+1)/2$ free parameters in Σ, given its symmetric nature, only that many parameters may be estimated in the **A** and **B** matrices. As there are $2K^2$ parameters in **A** and **B**, the order condition for identification requires that $2K^2 - K(K+1)/2$ restrictions be placed on the elements of these matrices.

For instance, we could reproduce the effect of the Cholesky decomposition by defining matrices **A** and **B** appropriately. In the syntax of `svar`, a missing value in a matrix is a free parameter to be estimated. The form of the **A** matrix imposes the recursive structure, while the diagonal **B** orthogonalizes the effects of innovations.

```
. matrix A = (1, 0, 0 \ ., 1, 0 \ ., ., 1)
. matrix B = (., 0, 0 \ 0, ., 0 \ 0, 0, 1)
. matrix list A
A[3,3]
    c1  c2  c3
r1   1   0   0
r2   .   1   0
r3   .   .   1
. matrix list B
symmetric B[3,3]
    c1  c2  c3
r1   .
r2   0   .
r3   0   0   1
```

```
. svar D.lrgrossinv D.lrconsump D.lrgdp if tin(,2005q4), aeq(A) beq(B) nolog
Estimating short-run parameters
Structural vector autoregression
 ( 1)  [a_1_1]_cons = 1
 ( 2)  [a_1_2]_cons = 0
 ( 3)  [a_1_3]_cons = 0
 ( 4)  [a_2_2]_cons = 1
 ( 5)  [a_2_3]_cons = 0
 ( 6)  [a_3_3]_cons = 1
 ( 7)  [b_1_2]_cons = 0
 ( 8)  [b_1_3]_cons = 0
 ( 9)  [b_2_1]_cons = 0
 (10)  [b_2_3]_cons = 0
 (11)  [b_3_1]_cons = 0
 (12)  [b_3_2]_cons = 0
Sample:  1959q4 - 2005q4                          No. of obs      =       185
Exactly identified model                          Log likelihood  =  1905.169

------------------------------------------------------------------------------
             |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       /a_1_1|          1          .        .       .            .           .
       /a_2_1|  -.2030461   .0232562    -8.73   0.000    -.2486274   -.1574649
       /a_3_1|  -.1827889   .0260518    -7.02   0.000    -.2338495   -.1317283
       /a_1_2|  (omitted)
       /a_2_2|          1          .        .       .            .           .
       /a_3_2|  -.4994815    .069309    -7.21   0.000    -.6353246   -.3636384
       /a_1_3|  (omitted)
       /a_2_3|  (omitted)
       /a_3_3|          1          .        .       .            .           .
       /b_1_1|   .0171686   .0008926    19.24   0.000     .0154193     .018918
```

The output from the VAR can also be displayed with the `var` option. This model is exactly identified; if we impose additional restrictions on the parameters, it would be an overidentified model, and the overidentifying restrictions could be tested.

For instance, we could impose the restriction that $\mathbf{A}_{2,1} = 0$ by placing a zero in that cell of the matrix rather than a missing value. This implies that changes in the first variable (`D.lrgrossinv`) do not contemporaneously affect the second variable, (`D.lrconsump`).

```
. matrix Arest = (1, 0, 0 \ 0, 1, 0 \ ., ., 1)
. matrix list Arest
Arest[3,3]
    c1  c2  c3
r1   1   0   0
r2   0   1   0
r3   .   .   1
```

The output from the VAR can also be displayed with the `var` option. This model is exactly identified; if we impose additional restrictions on the parameters, it would be an overidentified model, and the overidentifying restrictions could be tested.

For instance, we could impose the restriction that $\mathbf{A}_{2,1} = 0$ by placing a zero in that cell of the matrix rather than a missing value. This implies that changes in the first variable (`D.lrgrossinv`) do not contemporaneously affect the second variable, (`D.lrconsump`).

```
 . matrix Arest = (1, 0, 0 \ 0, 1, 0 \ ., ., 1)
. matrix list Arest
Arest[3,3]
    c1  c2  c3
r1   1   0   0
r2   0   1   0
r3   .   .   1
```

```
. svar D.lrgrossinv D.lrconsump D.lrgdp if tin(,2005q4), aeq(Arest) beq(B) nolog
Estimating short-run parameters
Structural vector autoregression
...
Sample:  1959q4 - 2005q4                          No. of obs      =        185
Overidentified model                              Log likelihood  =   1873.254

------------------------------------------------------------------------------
             |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      /a_1_1 |          1          .        .       .            .           .
      /a_2_1 |  (omitted)
      /a_3_1 |  -.1827926   .0219237    -8.34   0.000    -.2257622   -.1398229
      /a_1_2 |  (omitted)
      /a_2_2 |          1          .        .       .            .           .
      /a_3_2 |   -.499383   .0583265    -8.56   0.000    -.6137008   -.3850652
      /a_1_3 |  (omitted)
      /a_2_3 |  (omitted)
      /a_3_3 |          1          .        .       .            .           .
...
------------------------------------------------------------------------------
LR test of identifying restrictions:  chi2(  1)=    63.83  Prob > chi2 = 0.000
```

As we would expect from the significant coefficient in the exactly identified VAR, the overidentifying restriction is clearly rejected.

```
. svar D.lrgrossinv D.lrconsump D.lrgdp if tin(,2005q4), aeq(Arest) beq(B) nolog
Estimating short-run parameters
Structural vector autoregression
...
Sample:  1959q4 - 2005q4                          No. of obs      =       185
Overidentified model                              Log likelihood  =  1873.254

------------------------------------------------------------------------------
             |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      /a_1_1 |          1          .        .       .            .           .
      /a_2_1 |  (omitted)
      /a_3_1 |  -.1827926   .0219237    -8.34   0.000    -.2257622   -.1398229
      /a_1_2 |  (omitted)
      /a_2_2 |          1          .        .       .            .           .
      /a_3_2 |   -.499383   .0583265    -8.56   0.000    -.6137008   -.3850652
      /a_1_3 |  (omitted)
      /a_2_3 |  (omitted)
      /a_3_3 |          1          .        .       .            .           .
...
------------------------------------------------------------------------------
LR test of identifying restrictions:  chi2(  1)=    63.83  Prob > chi2 = 0.000
```

As we would expect from the significant coefficient in the exactly identified VAR, the overidentifying restriction is clearly rejected.

Long-run SVAR models

A short-run SVAR model without exogenous variables can be written as

$$\mathbf{A}(\mathbf{I}_K - \mathbf{A}_1 L - \mathbf{A}_2 L^2 - \cdots - \mathbf{A}_p L^p)\mathbf{y}_t = \mathbf{A}\bar{\mathbf{A}}\,\mathbf{y}_t = \mathbf{B}\,\mathbf{e}_t$$

where $\bar{\mathbf{A}}$ is the parenthesized expression. If we set $\mathbf{A} = \mathbf{I}$, we can write this equation as

$$\mathbf{y}_t = \bar{\mathbf{A}}^{-1}\mathbf{B}\,\mathbf{e}_t = \mathbf{C}\,\mathbf{e}_t$$

In a long-run SVAR, constraints are placed on elements of the $\mathbf{C}$ matrix. These constraints are often exclusion restrictions. For instance, constraining $\mathbf{C}_{1,2} = 0$ forces the long-run response of variable 1 to a shock to variable 2 to zero.

Long-run SVAR models

A short-run SVAR model without exogenous variables can be written as

$$\mathbf{A}(\mathbf{I}_K - \mathbf{A}_1 L - \mathbf{A}_2 L^2 - \cdots - \mathbf{A}_p L^p)\mathbf{y}_t = \mathbf{A}\bar{\mathbf{A}}\,\mathbf{y}_t = \mathbf{B}\,\mathbf{e}_t$$

where $\bar{\mathbf{A}}$ is the parenthesized expression. If we set $\mathbf{A} = \mathbf{I}$, we can write this equation as

$$\mathbf{y}_t = \bar{\mathbf{A}}^{-1}\mathbf{B}\,\mathbf{e}_t = \mathbf{C}\,\mathbf{e}_t$$

In a long-run SVAR, constraints are placed on elements of the $\mathbf{C}$ matrix. These constraints are often exclusion restrictions. For instance, constraining $\mathbf{C}_{1,2} = 0$ forces the long-run response of variable 1 to a shock to variable 2 to zero.

We illustrate with a two-variable SVAR in the first differences in the logs of real money and real GDP. The long-run restrictions of a diagonal **C** matrix implies that shocks to the money supply process have no long-run effects on GDP growth, and shocks to the GDP process have no long-run effects on the money supply.

```
 . matrix lr = (., 0\0, .)
. matrix list lr
symmetric lr[2,2]
    c1  c2
r1   .
r2   0   .
```

```
. svar D.lrmbase D.lrgdp, lags(4) lreq(lr) nolog
Estimating long-run parameters
Structural vector autoregression
 ( 1)  [c_1_2]_cons = 0
 ( 2)  [c_2_1]_cons = 0
Sample:  1960q2 - 2010q3                          No. of obs      =        202
Overidentified model                              Log likelihood  =   1020.662

------------------------------------------------------------------------------
             |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      /c_1_1 |   .0524697   .0026105    20.10   0.000     .0473532    .0575861
      /c_2_1 |  (omitted)
      /c_1_2 |  (omitted)
      /c_2_2 |   .0093022   .0004628    20.10   0.000     .0083951    .0102092
------------------------------------------------------------------------------
LR test of identifying restrictions:  chi2(  1)=    1.448  Prob > chi2 = 0.229
```

The test of overidentifying restrictions cannot reject the validity of the constraints imposed on the long-run responses.

```
. svar D.lrmbase D.lrgdp, lags(4) lreq(lr) nolog
Estimating long-run parameters
Structural vector autoregression
 ( 1)  [c_1_2]_cons = 0
 ( 2)  [c_2_1]_cons = 0
Sample:  1960q2 - 2010q3                         No. of obs      =        202
Overidentified model                             Log likelihood  =   1020.662

------------------------------------------------------------------------------
             |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      /c_1_1 |   .0524697   .0026105    20.10   0.000     .0473532    .0575861
      /c_2_1 |  (omitted)
      /c_1_2 |  (omitted)
      /c_2_2 |   .0093022   .0004628    20.10   0.000     .0083951    .0102092
------------------------------------------------------------------------------
LR test of identifying restrictions:  chi2(  1)=    1.448  Prob > chi2 = 0.229
```

The test of overidentifying restrictions cannot reject the validity of the constraints imposed on the long-run responses.

Vector error correction models (VECMs)

VECMs may be estimated by Stata's `vec` command. These models are employed because many economic time series appear to be 'first-difference stationary,' with their levels exhibiting unit root or nonstationary behavior. Conventional regression estimators, including VARs, have good properties when applied to covariance-stationary time series, but encounter difficulties when applied to nonstationary or integrated processes.

These difficulties were illustrated by Granger and Newbold (*J. Econometrics*, 1974) when they introduced the concept of *spurious regressions*. If you have two independent random walk processes, a regression of one on the other will yield a significant coefficient, even though they are not related in any way.

Vector error correction models (VECMs)

VECMs may be estimated by Stata's `vec` command. These models are employed because many economic time series appear to be 'first-difference stationary,' with their levels exhibiting unit root or nonstationary behavior. Conventional regression estimators, including VARs, have good properties when applied to covariance-stationary time series, but encounter difficulties when applied to nonstationary or integrated processes.

These difficulties were illustrated by Granger and Newbold (*J. Econometrics*, 1974) when they introduced the concept of *spurious regressions*. If you have two independent random walk processes, a regression of one on the other will yield a significant coefficient, even though they are not related in any way.

This insight, and Nelson and Plosser's findings (*J. Mon. Ec.*, 1982) that unit roots might be present in a wide variety of macroeconomic series in levels or logarithms, gave rise to the industry of unit root testing, and the implication that variables should be rendered stationary by differencing before they are included in an econometric model.

Further theoretical developments by Granger and Engle in their celebrated paper (*Econometrica*, 1987) raised the possibility that two or more integrated, nonstationary time series might be *cointegrated*, so that some linear combination of these series could be stationary even though each series is not.

This insight, and Nelson and Plosser's findings (*J. Mon. Ec.*, 1982) that unit roots might be present in a wide variety of macroeconomic series in levels or logarithms, gave rise to the industry of unit root testing, and the implication that variables should be rendered stationary by differencing before they are included in an econometric model.

Further theoretical developments by Granger and Engle in their celebrated paper (*Econometrica*, 1987) raised the possibility that two or more integrated, nonstationary time series might be *cointegrated*, so that some linear combination of these series could be stationary even though each series is not.

If two series are both integrated (of order one, or $I(1)$) we could model their interrelationship by taking first differences of each series and including the differences in a VAR or a structural model.

However, this approach would be suboptimal if it was determined that these series are indeed cointegrated. In that case, the VAR would only express the short-run responses of these series to innovations in each series. This implies that the simple regression in first differences is misspecified.

If the series are cointegrated, they move together in the long run. A VAR in first differences, although properly specified in terms of covariance-stationary series, will not capture those long-run tendences.

If two series are both integrated (of order one, or $I(1)$) we could model their interrelationship by taking first differences of each series and including the differences in a VAR or a structural model.

However, this approach would be suboptimal if it was determined that these series are indeed cointegrated. In that case, the VAR would only express the short-run responses of these series to innovations in each series. This implies that the simple regression in first differences is misspecified.

If the series are cointegrated, they move together in the long run. A VAR in first differences, although properly specified in terms of covariance-stationary series, will not capture those long-run tendences.

If two series are both integrated (of order one, or $I(1)$) we could model their interrelationship by taking first differences of each series and including the differences in a VAR or a structural model.

However, this approach would be suboptimal if it was determined that these series are indeed cointegrated. In that case, the VAR would only express the short-run responses of these series to innovations in each series. This implies that the simple regression in first differences is misspecified.

If the series are cointegrated, they move together in the long run. A VAR in first differences, although properly specified in terms of covariance-stationary series, will not capture those long-run tendences.

Accordingly, the VAR concept may be extended to the vector error-correction model, or VECM, where there is evidence of cointegration among two or more series. The model is fit to the first differences of the nonstationary variables, but a lagged *error-correction term* is added to the relationship.

In the case of two variables, this term is the lagged residual from the cointegrating regression, of one of the series on the other in levels. It expresses the prior disequilibrium from the long-run relationship, in which that residual would be zero.

In the case of multiple variables, there is a vector of error-correction terms, of length equal to the number of cointegrating relationships, or *cointegrating vectors*, among the series.

Accordingly, the VAR concept may be extended to the vector error-correction model, or VECM, where there is evidence of cointegration among two or more series. The model is fit to the first differences of the nonstationary variables, but a lagged *error-correction term* is added to the relationship.

In the case of two variables, this term is the lagged residual from the cointegrating regression, of one of the series on the other in levels. It expresses the prior disequilibrium from the long-run relationship, in which that residual would be zero.

In the case of multiple variables, there is a vector of error-correction terms, of length equal to the number of cointegrating relationships, or *cointegrating vectors*, among the series.

Accordingly, the VAR concept may be extended to the vector error-correction model, or VECM, where there is evidence of cointegration among two or more series. The model is fit to the first differences of the nonstationary variables, but a lagged *error-correction term* is added to the relationship.

In the case of two variables, this term is the lagged residual from the cointegrating regression, of one of the series on the other in levels. It expresses the prior disequilibrium from the long-run relationship, in which that residual would be zero.

In the case of multiple variables, there is a vector of error-correction terms, of length equal to the number of cointegrating relationships, or *cointegrating vectors*, among the series.

In terms of economic content, we might expect that there is some long-run value of the dividend/price ratio for common equities. During market 'bubbles', the stock price index may be high and the ratio low, but we would expect a market correction to return the ratio to its long-run value. A similar rationale can be offered about the ratio of rents to housing prices in a housing market where there is potential to construct new rental housing as well as single-family homes.

To extend the concept to more than two variables, we might rely on the concept of purchasing power parity (PPP) in international trade, which defines a relationship between the nominal exchange rate and the price indices in the foreign and domestic economies. We might find episodes where a currency appears over- or undervalued, but in the absence of central bank intervention and effective exchange controls, we expect that the 'law of one price' will provide some long-run anchor to these three measures' relationship.

In terms of economic content, we might expect that there is some long-run value of the dividend/price ratio for common equities. During market 'bubbles', the stock price index may be high and the ratio low, but we would expect a market correction to return the ratio to its long-run value. A similar rationale can be offered about the ratio of rents to housing prices in a housing market where there is potential to construct new rental housing as well as single-family homes.

To extend the concept to more than two variables, we might rely on the concept of purchasing power parity (PPP) in international trade, which defines a relationship between the nominal exchange rate and the price indices in the foreign and domestic economies. We might find episodes where a currency appears over- or undervalued, but in the absence of central bank intervention and effective exchange controls, we expect that the 'law of one price' will provide some long-run anchor to these three measures' relationship.

Consider two series, y_t and x_t, that obey the following equations:

$$\begin{aligned} y_t + \beta x_t &= \epsilon_t, \quad \epsilon_t = \epsilon_{t-1} + \omega_t \\ y_t + \alpha x_t &= \nu_t, \quad \nu_t = \rho \nu_{t-1} + \zeta_t, \quad |\rho| < 1 \end{aligned}$$

Assume that ω_t and ζ_t are *i.i.d.* disturbances, correlated with each other. The random-walk nature of ϵ_t implies that both y_t and x_t are also $I(1)$, or nonstationary, as each side of the equation must have the same order of integration. By the same token, the stationary nature of the ν_t process implies that the linear combination $(y_t + \alpha x_t)$ must also be stationary, or $I(0)$.

Thus y_t and x_t cointegrate, with a cointegrating vector $(1, \alpha)$.

Consider two series, y_t and x_t, that obey the following equations:

$$
\begin{aligned}
y_t + \beta x_t &= \epsilon_t, \quad \epsilon_t = \epsilon_{t-1} + \omega_t \\
y_t + \alpha x_t &= \nu_t, \quad \nu_t = \rho \nu_{t-1} + \zeta_t, \quad |\rho| < 1
\end{aligned}
$$

Assume that ω_t and ζ_t are *i.i.d.* disturbances, correlated with each other. The random-walk nature of ϵ_t implies that both y_t and x_t are also $I(1)$, or nonstationary, as each side of the equation must have the same order of integration. By the same token, the stationary nature of the ν_t process implies that the linear combination $(y_t + \alpha x_t)$ must also be stationary, or $I(0)$.

Thus y_t and x_t cointegrate, with a cointegrating vector $(1, \alpha)$.

We can rewrite the system as

$$\begin{aligned}\Delta y_t &= \beta\delta z_{t-1} + \eta_{1t}\\ \Delta x_t &= -\delta z_{t-1} + \eta_{2t}\end{aligned}$$

where $\delta = (1-\rho)/(\alpha-\beta)$, $z_t = y_t + \alpha x_t$, and the errors (η_{1t}, η_{2t}) are stationary linear combinations of (ω_t, ζ_t).

When y_t and x_t are in equilibrium, $z_t = 0$. The coefficients on z_t indicate how the system responds to disequilibrium. A stable dynamic system must exhibit *negative feedback*: for instance, in a functioning market, excess demand must cause the price to rise to clear the market.

We can rewrite the system as

$$\begin{aligned}\Delta y_t &= \beta\delta z_{t-1} + \eta_{1t}\\ \Delta x_t &= -\delta z_{t-1} + \eta_{2t}\end{aligned}$$

where $\delta = (1-\rho)/(\alpha-\beta),\ z_t = y_t + \alpha x_t$, and the errors (η_{1t}, η_{2t}) are stationary linear combinations of (ω_t, ζ_t).

When y_t and x_t are in equilibrium, $z_t = 0$. The coefficients on z_t indicate how the system responds to disequilibrium. A stable dynamic system must exhibit *negative feedback*: for instance, in a functioning market, excess demand must cause the price to rise to clear the market.

In the case of two nonstationary ($I(1)$) variables y_t and x_t, if there are two nonzero values (a, b) such that $ay_t + bx_t$ is stationary, or $I(0)$, then the variables are cointegrated. To identify the cointegrating vector, we set one of the values (a, b) to 1 and estimate the other. As Granger and Engle showed, this can be done by a regression in levels. If the residuals from that 'Granger–Engle' regression are stationary, cointegration is established.

In the general case of K variables, there may be 1, 2,..., (K-1) cointegrating vectors representing stationary linear combinations. That is, if $\mathbf{y}_t$ is a vector of $I(1)$ variables and there exists a vector β such that $\beta\mathbf{y}_t$ is a vector of $I(0)$ variables, then the variables in $\mathbf{y}_t$ are said to be cointegrated with cointegrating vector β. In that case we need to estimate the number of cointegrating relationships, not merely whether cointegration exists among these series.

In the case of two nonstationary ($I(1)$) variables y_t and x_t, if there are two nonzero values (a, b) such that $ay_t + bx_t$ is stationary, or $I(0)$, then the variables are cointegrated. To identify the cointegrating vector, we set one of the values (a, b) to 1 and estimate the other. As Granger and Engle showed, this can be done by a regression in levels. If the residuals from that 'Granger–Engle' regression are stationary, cointegration is established.

In the general case of K variables, there may be 1, 2,. . . ,(K-1) cointegrating vectors representing stationary linear combinations. That is, if $\mathbf{y}_t$ is a vector of $I(1)$ variables and there exists a vector β such that $\beta\mathbf{y}_t$ is a vector of $I(0)$ variables, then the variables in $\mathbf{y}_t$ are said to be cointegrated with cointegrating vector β. In that case we need to estimate the number of cointegrating relationships, not merely whether cointegration exists among these series.

For a K-variable VAR with p lags,

$$\mathbf{y}_t = \mathbf{v} + \mathbf{A}_1 \mathbf{y}_{t-1} + \cdots + \mathbf{A}_p \mathbf{y}_{t-p} + \epsilon_t$$

let ϵ_t be *i.i.d.* normal over time with covariance matrix Σ. We may rewrite the VAR as a VECM:

$$\Delta \mathbf{y}_t = v + \Pi \mathbf{y}_{t-1} + \sum_{i=1}^{p-1} \Gamma_i \Delta \mathbf{y}_{t-i} + \epsilon_t$$

where $\Pi = \sum_{j=1}^{j=p} \mathbf{A}_j - \mathbf{I}_k$ and $\Gamma_i = -\sum_{j=i+1}^{j=p} \mathbf{A}_j$.

If all variables in $\mathbf{y}_t$ are $I(1)$, the matrix Π has rank $0 \leq r < K$, where r is the number of linearly independent cointegrating vectors. If the variables are cointegrated ($r > 0$) the VAR in first differences is misspecified as it excludes the error correction term.

If the rank of $\Pi = 0$, there is no cointegration among the nonstationary variables, and a VAR in their first differences is consistent.

If the rank of $\Pi = K$, all of the variables in $\mathbf{y}_t$ are $I(0)$ and a VAR in their levels is consistent.

If the rank of Π is $r > 0$, it may be expressed as $\Pi = \alpha\beta'$, where α and β are $(K \times r)$ matrices of rank r. We must place restrictions on these matrices' elements in order to identify the system.

If all variables in $\mathbf{y}_t$ are $I(1)$, the matrix Π has rank $0 \leq r < K$, where r is the number of linearly independent cointegrating vectors. If the variables are cointegrated ($r > 0$) the VAR in first differences is misspecified as it excludes the error correction term.

If the rank of $\Pi = 0$, there is no cointegration among the nonstationary variables, and a VAR in their first differences is consistent.

If the rank of $\Pi = K$, all of the variables in $\mathbf{y}_t$ are $I(0)$ and a VAR in their levels is consistent.

If the rank of Π is $r > 0$, it may be expressed as $\Pi = \alpha\beta'$, where α and β are $(K \times r)$ matrices of rank r. We must place restrictions on these matrices' elements in order to identify the system.

If all variables in $\mathbf{y}_t$ are $I(1)$, the matrix Π has rank $0 \leq r < K$, where r is the number of linearly independent cointegrating vectors. If the variables are cointegrated ($r > 0$) the VAR in first differences is misspecified as it excludes the error correction term.

If the rank of $\Pi = 0$, there is no cointegration among the nonstationary variables, and a VAR in their first differences is consistent.

If the rank of $\Pi = K$, all of the variables in $\mathbf{y}_t$ are $I(0)$ and a VAR in their levels is consistent.

If the rank of Π is $r > 0$, it may be expressed as $\Pi = \alpha\beta'$, where α and β are $(K \times r)$ matrices of rank r. We must place restrictions on these matrices' elements in order to identify the system.

If all variables in $\mathbf{y}_t$ are $I(1)$, the matrix Π has rank $0 \leq r < K$, where r is the number of linearly independent cointegrating vectors. If the variables are cointegrated ($r > 0$) the VAR in first differences is misspecified as it excludes the error correction term.

If the rank of $\Pi = 0$, there is no cointegration among the nonstationary variables, and a VAR in their first differences is consistent.

If the rank of $\Pi = K$, all of the variables in $\mathbf{y}_t$ are $I(0)$ and a VAR in their levels is consistent.

If the rank of Π is $r > 0$, it may be expressed as $\Pi = \alpha\beta'$, where α and β are $(K \times r)$ matrices of rank r. We must place restrictions on these matrices' elements in order to identify the system.

Stata's implementation of VECM modeling is based on the maximum likelihood framework of Johansen (*J. Ec. Dyn. Ctrl.*, 1988 and subsequent works). In that framework, deterministic trends can appear in the means of the differenced series, or in the mean of the cointegrating relationship. The constant term in the VECM implies a linear trend in the levels of the variables. Thus, a time trend in the equation implies quadratic trends in the level data.

Writing the matrix of coefficients on the vector error correction term $\mathbf{y}_{t-1}$ as $\Pi = \alpha\beta'$, we can incorporate a trend in the cointegrating relationship and the equation itself as

$$\Delta \mathbf{y}_t = \alpha(\beta' \mathbf{y}_{t-1} + \mu + \rho t) + \sum_{i=1}^{p-1} \Gamma_i \Delta \mathbf{y}_{t-i} + \gamma + \tau t + \epsilon_t$$

Stata's implementation of VECM modeling is based on the maximum likelihood framework of Johansen (*J. Ec. Dyn. Ctrl.*, 1988 and subsequent works). In that framework, deterministic trends can appear in the means of the differenced series, or in the mean of the cointegrating relationship. The constant term in the VECM implies a linear trend in the levels of the variables. Thus, a time trend in the equation implies quadratic trends in the level data.

Writing the matrix of coefficients on the vector error correction term $\mathbf{y}_{t-1}$ as $\Pi = \alpha\beta'$, we can incorporate a trend in the cointegrating relationship and the equation itself as

$$\Delta \mathbf{y}_t = \alpha(\beta'\mathbf{y}_{t-1} + \mu + \rho t) + \sum_{i=1}^{p-1} \Gamma_i \Delta \mathbf{y}_{t-i} + \gamma + \tau t + \epsilon_t$$

Johansen spells out five cases for estimation of the VECM:

1. Unrestricted trend: estimated as shown, cointegrating equations are trend stationary
2. Restricted trend, $\tau = 0$: cointegrating equations are trend stationary, and trends in levels are linear but not quadratic
3. Unrestricted constant: $\tau = \rho = 0$: cointegrating equations are stationary around constant means, linear trend in levels
4. Restricted constant: $\tau = \rho = \gamma = 0$: cointegrating equations are stationary around constant means, no linear time trends in the data
5. No trend: $\tau = \rho = \gamma = \mu = 0$: cointegrating equations, levels and differences of the data have means of zero

We have not illustrated VECMs with additional (strictly) exogenous variables, but they may be added, just as in a VAR model.

Johansen spells out five cases for estimation of the VECM:

1. Unrestricted trend: estimated as shown, cointegrating equations are trend stationary
2. Restricted trend, $\tau = 0$: cointegrating equations are trend stationary, and trends in levels are linear but not quadratic
3. Unrestricted constant: $\tau = \rho = 0$: cointegrating equations are stationary around constant means, linear trend in levels
4. Restricted constant: $\tau = \rho = \gamma = 0$: cointegrating equations are stationary around constant means, no linear time trends in the data
5. No trend: $\tau = \rho = \gamma = \mu = 0$: cointegrating equations, levels and differences of the data have means of zero

We have not illustrated VECMs with additional (strictly) exogenous variables, but they may be added, just as in a VAR model.

Johansen spells out five cases for estimation of the VECM:

1. Unrestricted trend: estimated as shown, cointegrating equations are trend stationary
2. Restricted trend, $\tau = 0$: cointegrating equations are trend stationary, and trends in levels are linear but not quadratic
3. Unrestricted constant: $\tau = \rho = 0$: cointegrating equations are stationary around constant means, linear trend in levels
4. Restricted constant: $\tau = \rho = \gamma = 0$: cointegrating equations are stationary around constant means, no linear time trends in the data
5. No trend: $\tau = \rho = \gamma = \mu = 0$: cointegrating equations, levels and differences of the data have means of zero

We have not illustrated VECMs with additional (strictly) exogenous variables, but they may be added, just as in a VAR model.

Johansen spells out five cases for estimation of the VECM:

1. Unrestricted trend: estimated as shown, cointegrating equations are trend stationary
2. Restricted trend, $\tau = 0$: cointegrating equations are trend stationary, and trends in levels are linear but not quadratic
3. Unrestricted constant: $\tau = \rho = 0$: cointegrating equations are stationary around constant means, linear trend in levels
4. Restricted constant: $\tau = \rho = \gamma = 0$: cointegrating equations are stationary around constant means, no linear time trends in the data
5. No trend: $\tau = \rho = \gamma = \mu = 0$: cointegrating equations, levels and differences of the data have means of zero

We have not illustrated VECMs with additional (strictly) exogenous variables, but they may be added, just as in a VAR model.

Johansen spells out five cases for estimation of the VECM:

1. Unrestricted trend: estimated as shown, cointegrating equations are trend stationary
2. Restricted trend, $\tau = 0$: cointegrating equations are trend stationary, and trends in levels are linear but not quadratic
3. Unrestricted constant: $\tau = \rho = 0$: cointegrating equations are stationary around constant means, linear trend in levels
4. Restricted constant: $\tau = \rho = \gamma = 0$: cointegrating equations are stationary around constant means, no linear time trends in the data
5. No trend: $\tau = \rho = \gamma = \mu = 0$: cointegrating equations, levels and differences of the data have means of zero

We have not illustrated VECMs with additional (strictly) exogenous variables, but they may be added, just as in a VAR model.

Johansen spells out five cases for estimation of the VECM:

1. Unrestricted trend: estimated as shown, cointegrating equations are trend stationary
2. Restricted trend, $\tau = 0$: cointegrating equations are trend stationary, and trends in levels are linear but not quadratic
3. Unrestricted constant: $\tau = \rho = 0$: cointegrating equations are stationary around constant means, linear trend in levels
4. Restricted constant: $\tau = \rho = \gamma = 0$: cointegrating equations are stationary around constant means, no linear time trends in the data
5. No trend: $\tau = \rho = \gamma = \mu = 0$: cointegrating equations, levels and differences of the data have means of zero

We have not illustrated VECMs with additional (strictly) exogenous variables, but they may be added, just as in a VAR model.

Johansen spells out five cases for estimation of the VECM:

1. Unrestricted trend: estimated as shown, cointegrating equations are trend stationary
2. Restricted trend, $\tau = 0$: cointegrating equations are trend stationary, and trends in levels are linear but not quadratic
3. Unrestricted constant: $\tau = \rho = 0$: cointegrating equations are stationary around constant means, linear trend in levels
4. Restricted constant: $\tau = \rho = \gamma = 0$: cointegrating equations are stationary around constant means, no linear time trends in the data
5. No trend: $\tau = \rho = \gamma = \mu = 0$: cointegrating equations, levels and differences of the data have means of zero

We have not illustrated VECMs with additional (strictly) exogenous variables, but they may be added, just as in a VAR model.

To consistently test for cointegration, we must choose the appropriate lag length. The `varsoc` command is capable of making that determination, as illustrated earlier. We may then use the `vecrank` command to test for cointegration via Johansen's max-eigenvalue statistic and trace statistic.

We illustrate a simple VECM using the Penn World Tables data. In that data set, the price index is the relative price vs. the US, and the nominal exchange rate is expressed as local currency units per US dollar. If the real exchange rate is a cointegrating combination, the logs of the price index and the nominal exchange rate should be cointegrated. We test this hypothesis with respect to the UK, using Stata's default of an unrestricted constant in the taxonomy given above.

To consistently test for cointegration, we must choose the appropriate lag length. The `varsoc` command is capable of making that determination, as illustrated earlier. We may then use the `vecrank` command to test for cointegration via Johansen's max-eigenvalue statistic and trace statistic.

We illustrate a simple VECM using the Penn World Tables data. In that data set, the price index is the relative price vs. the US, and the nominal exchange rate is expressed as local currency units per US dollar. If the real exchange rate is a cointegrating combination, the logs of the price index and the nominal exchange rate should be cointegrated. We test this hypothesis with respect to the UK, using Stata's default of an unrestricted constant in the taxonomy given above.

```
. use pwt6_3, clear
(Penn World Tables 6.3, August 2009)
. keep if inlist(isocode,"GBR")
(10962 observations deleted)
. // p already defined as UK/US relative price
. g lp = log(p)
. // xrat is nominal exchange rate, GBP per USD
. g lxrat = log(xrat)
. varsoc lp lxrat if tin(,2002)
   Selection-order criteria
   Sample:  1954 - 2002                          Number of obs      =        49
```

lag	LL	LR	df	p	FPE	AIC	HQIC	SBIC
0	19.4466				.001682	-.712107	-.682811	-.63489
1	173.914	308.93	4	0.000	3.6e-06	-6.85363	-6.76575	-6.62198
2	206.551	65.275*	4	0.000	1.1e-06*	-8.02251*	-7.87603*	-7.63642*
3	210.351	7.5993	4	0.107	1.1e-06	-8.01433	-7.80926	-7.47381
4	214.265	7.827	4	0.098	1.1e-06	-8.0108	-7.74714	-7.31585

```
   Endogenous:  lp lxrat
    Exogenous:  _cons
```

Two lags are selected by most of the criteria.

```
. use pwt6_3, clear
(Penn World Tables 6.3, August 2009)

. keep if inlist(isocode,"GBR")
(10962 observations deleted)

. // p already defined as UK/US relative price
. g lp = log(p)

. // xrat is nominal exchange rate, GBP per USD
. g lxrat = log(xrat)

. varsoc lp lxrat if tin(,2002)

   Selection-order criteria
   Sample:  1954 - 2002                          Number of obs      =        49
```

lag	LL	LR	df	p	FPE	AIC	HQIC	SBIC
0	19.4466				.001682	-.712107	-.682811	-.63489
1	173.914	308.93	4	0.000	3.6e-06	-6.85363	-6.76575	-6.62198
2	206.551	65.275*	4	0.000	1.1e-06*	-8.02251*	-7.87603*	-7.63642*
3	210.351	7.5993	4	0.107	1.1e-06	-8.01433	-7.80926	-7.47381
4	214.265	7.827	4	0.098	1.1e-06	-8.0108	-7.74714	-7.31585

```
   Endogenous:  lp lxrat
    Exogenous:  _cons
```

Two lags are selected by most of the criteria.

```
. vecrank lp lxrat if tin(,2002)
                         Johansen tests for cointegration
Trend: constant                                         Number of obs =      51
Sample:  1952 - 2002                                             Lags =       2
-------------------------------------------------------------------------------
                                                        5%
maximum                                      trace    critical
  rank    parms      LL       eigenvalue  statistic    value
    0       6     202.92635       .        22.9305     15.41
    1       9     213.94024    0.35074      0.9028*     3.76
    2      10     214.39162    0.01755
-------------------------------------------------------------------------------
```

We can reject the null of 0 cointegrating vectors in favor of > 0 via the trace statistic. We cannot reject the null of 1 cointegrating vector in favor of > 1. Thus, we conclude that there is one cointegrating vector. For two series, this could have also been determined by a Granger–Engle regression in levels.

```
. vecrank lp lxrat if tin(,2002)
                       Johansen tests for cointegration
Trend: constant                                          Number of obs =      51
Sample:  1952 - 2002                                              Lags =       2
-------------------------------------------------------------------------------
                                                           5%
maximum                                      trace     critical
  rank    parms      LL        eigenvalue  statistic    value
    0       6     202.92635          .      22.9305     15.41
    1       9     213.94024     0.35074      0.9028*     3.76
    2      10     214.39162     0.01755
-------------------------------------------------------------------------------
```

We can reject the null of 0 cointegrating vectors in favor of > 0 via the trace statistic. We cannot reject the null of 1 cointegrating vector in favor of > 1. Thus, we conclude that there is one cointegrating vector. For two series, this could have also been determined by a Granger–Engle regression in levels.

```
. vec lp lxrat if tin(,2002), lags(2)
Vector error-correction model
Sample:  1952 - 2002                              No. of obs      =        51
                                                  AIC             = -8.036872
Log likelihood =  213.9402                        HQIC            =   -7.9066
Det(Sigma_ml)  =  7.79e-07                        SBIC            = -7.695962
Equation           Parms      RMSE     R-sq      chi2     P>chi2
----------------------------------------------------------------
D_lp                 4      .057538   0.4363   36.37753   0.0000
D_lxrat              4      .055753   0.4496   38.38598   0.0000
----------------------------------------------------------------

------------------------------------------------------------------------------
             |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
D_lp         |
        _ce1 |
         L1. |    -.26966   .0536001    -5.03   0.000    -.3747143   -.1646057
             |
          lp |
         LD. |   .4083733    .324227     1.26   0.208    -.2270999    1.043847
             |
       lxrat |
         LD. |  -.1750804   .3309682    -0.53   0.597    -.8237663    .4736054
             |
       _cons |   .0027061   .0111043     0.24   0.807     -.019058    .0244702
-------------+----------------------------------------------------------------
...
```

```
--------------+----------------------------------------------------------------
D_lxrat       |
          _ce1|
           L1.|    .2537426    .0519368     4.89   0.000      .1519484    .3555369
              |
            lp|
           LD.|    .3566706    .3141656     1.14   0.256     -.2590827    .9724239
              |
         lxrat|
           LD.|    .8975872    .3206977     2.80   0.005      .2690313    1.526143
              |
         _cons|    .0028758    .0107597     0.27   0.789     -.0182129    .0239645
-------------------------------------------------------------------------------

Cointegrating equations
Equation           Parms    chi2     P>chi2
-------------------------------------------
_ce1                   1   44.70585   0.0000
-------------------------------------------

Identification:  beta is exactly identified
                 Johansen normalization restriction imposed
-------------------------------------------------------------------------------
         beta |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
--------------+----------------------------------------------------------------
_ce1          |
            lp|          1          .        .       .            .           .
         lxrat|   -.7842433    .1172921    -6.69   0.000     -1.014131   -.5543551
         _cons|   -4.982628          .        .       .            .           .
-------------------------------------------------------------------------------
```

In the `lp` equation, the `L1._ce1` term is the lagged error correction term. It is significantly negative, representing the negative feedback necessary in relative prices to bring the real exchange rate back to equilibrium. The short-run coefficients in this equation are not significantly different from zero.

In the `lxrat` equation, the lagged error correction term is positive, as it must be for the other variable in the relationship: that is, if $(\log p - \log e)$ is above long-run equilibrium, either p must fall or e must rise. The short-run coefficient on the exchange rate is positive and significant.

In the `lp` equation, the `L1._ce1` term is the lagged error correction term. It is significantly negative, representing the negative feedback necessary in relative prices to bring the real exchange rate back to equilibrium. The short-run coefficients in this equation are not significantly different from zero.

In the `lxrat` equation, the lagged error correction term is positive, as it must be for the other variable in the relationship: that is, if $(\log p - \log e)$ is above long-run equilibrium, either p must fall or e must rise. The short-run coefficient on the exchange rate is positive and significant.

The estimated cointegrating vector is listed at the foot of the output, normalized with a coefficient of unity on `lp` and an estimated coefficient of -0.78 on `lxrat`, significantly different from zero. The constant term corresponds to the μ term in the representation given above.

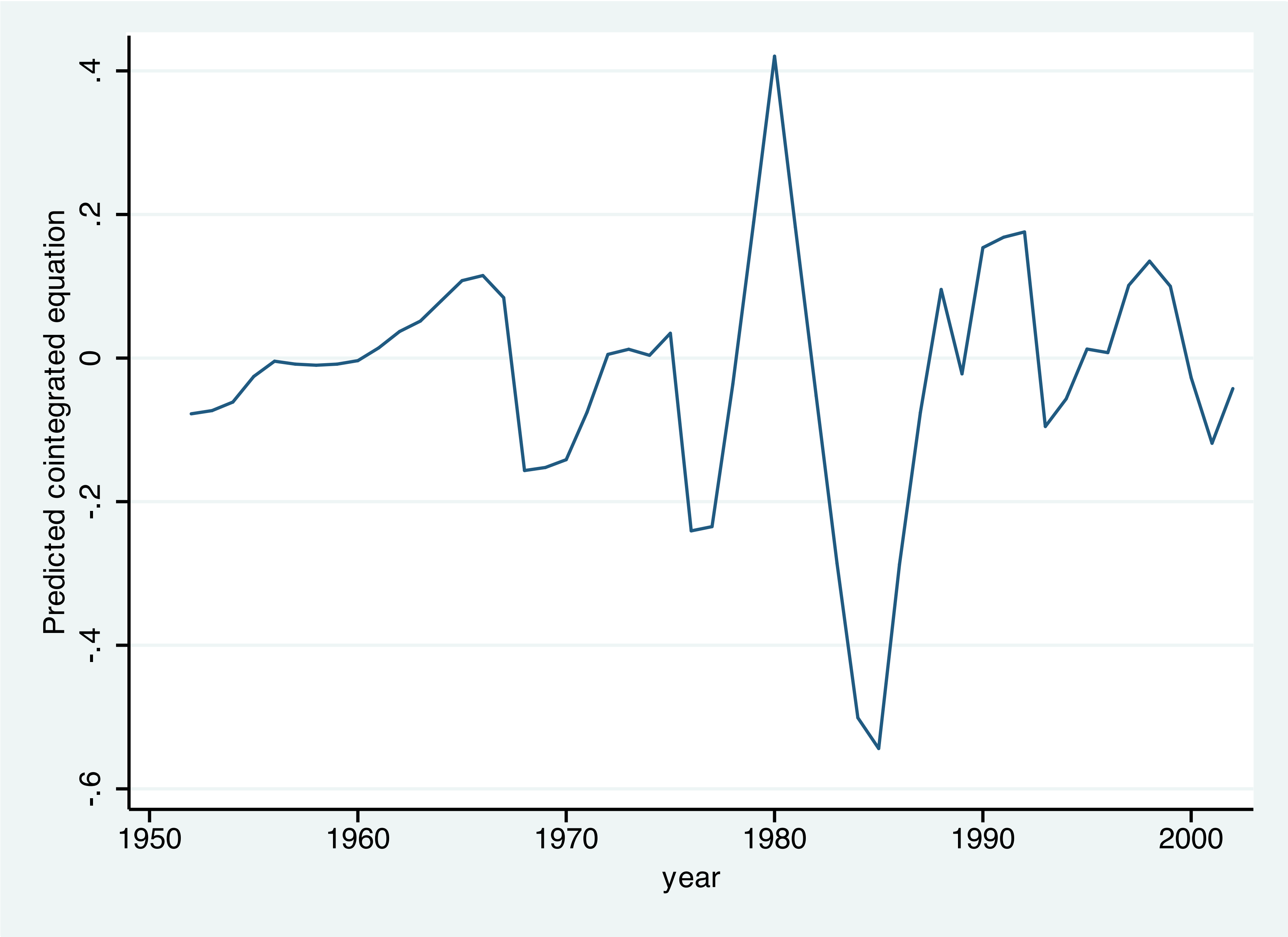
The significance of the lagged error correction term in this equation, and the significant coefficient estimated in the cointegrating vector, indicates that a VAR in first differences of these variables would yield inconsistent estimates due to misspecification.

The estimated cointegrating vector is listed at the foot of the output, normalized with a coefficient of unity on `lp` and an estimated coefficient of -0.78 on `lxrat`, significantly different from zero. The constant term corresponds to the μ term in the representation given above.

The significance of the lagged error correction term in this equation, and the significant coefficient estimated in the cointegrating vector, indicates that a VAR in first differences of these variables would yield inconsistent estimates due to misspecification.

We can evaluate the cointegrating equation by using `predict` to generate its in-sample values:

```
. predict ce1 if e(sample), ce equ(#1)
. tsline ce1 if e(sample)
```


Predicted cointegrated equation
.4
.2
0
-.2
-.4
-.6
1950
1960
1970
1980
1990
2000
year

We should also evaluate the stability of the estimated VECM. For a K-variable model with r cointegrating relationships, the companion matrix will have $K - r$ unit eigenvalues. For stability, the moduli of the remaining r eigenvalues should be strictly less than unity.

```
. vecstable, graph
  Eigenvalue stability condition
```

Eigenvalue	Modulus
1	1
.7660493	.766049
.5356276 + .522604*i*	.748339
.5356276 - .522604*i*	.748339

```
  The VECM specification imposes a unit modulus.
```

The eigenvalues meet the stability condition.

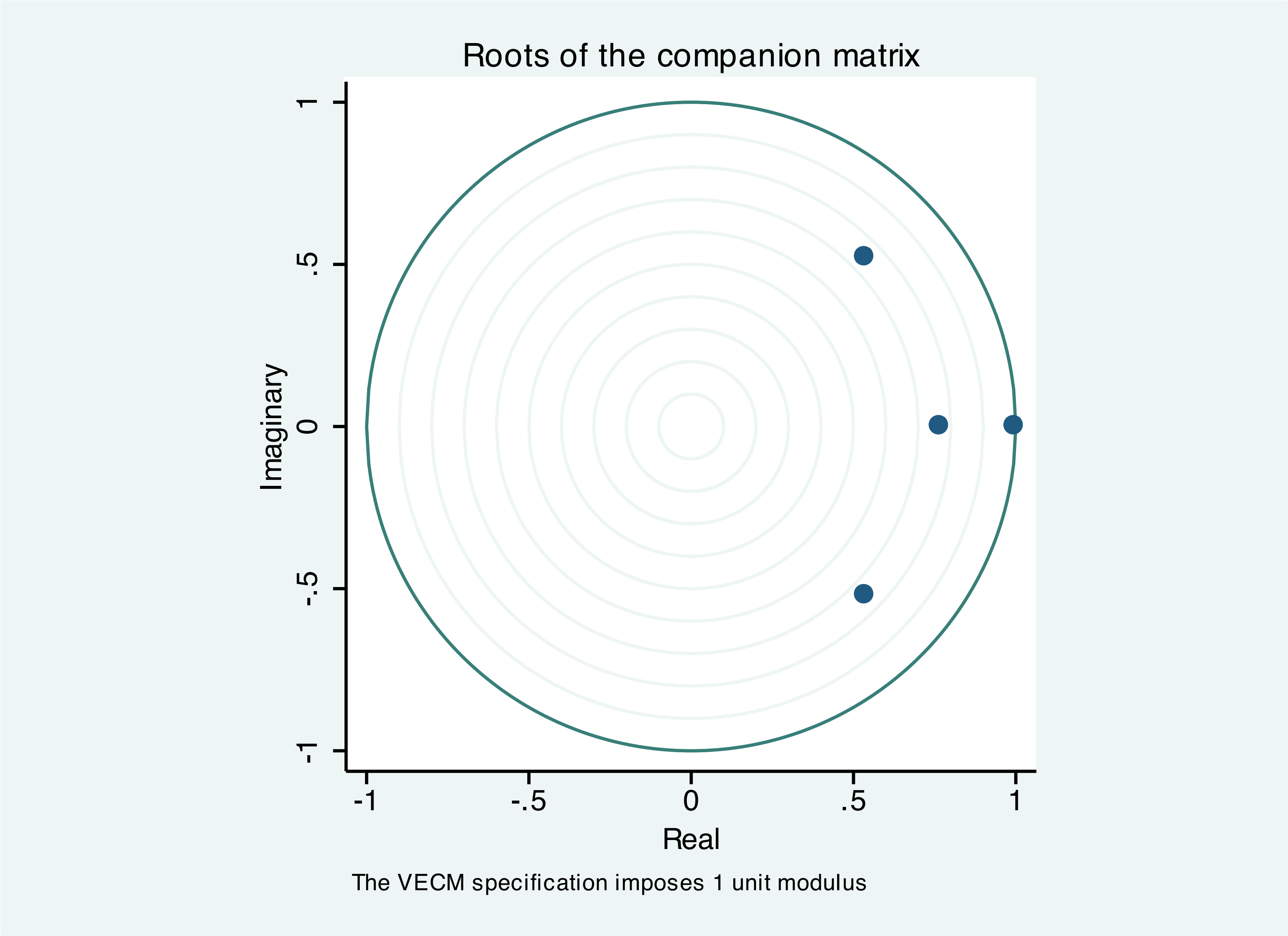
We should also evaluate the stability of the estimated VECM. For a K-variable model with r cointegrating relationships, the companion matrix will have $K - r$ unit eigenvalues. For stability, the moduli of the remaining r eigenvalues should be strictly less than unity.

```
. vecstable, graph
  Eigenvalue stability condition
```

Eigenvalue	Modulus
1	1
.7660493	.766049
.5356276 + .522604*i*	.748339
.5356276 − .522604*i*	.748339

```
  The VECM specification imposes a unit modulus.
```

The eigenvalues meet the stability condition.


Roots of the companion matrix
Imaginary
1
.5
0
-.5
-1
-1
-.5
0
.5
1
Real
The VECM specification imposes 1 unit modulus

We can use much of the same post-estimation apparatus as developed for VARs for VECMs. Impulse response functions, orthogonalized IRFs, FEVDs, and the like can be constructed for VECMs. However, the presence of the integrated variables (and unit moduli) in the VECM representation implies that shocks may be permanent as well as transitory.

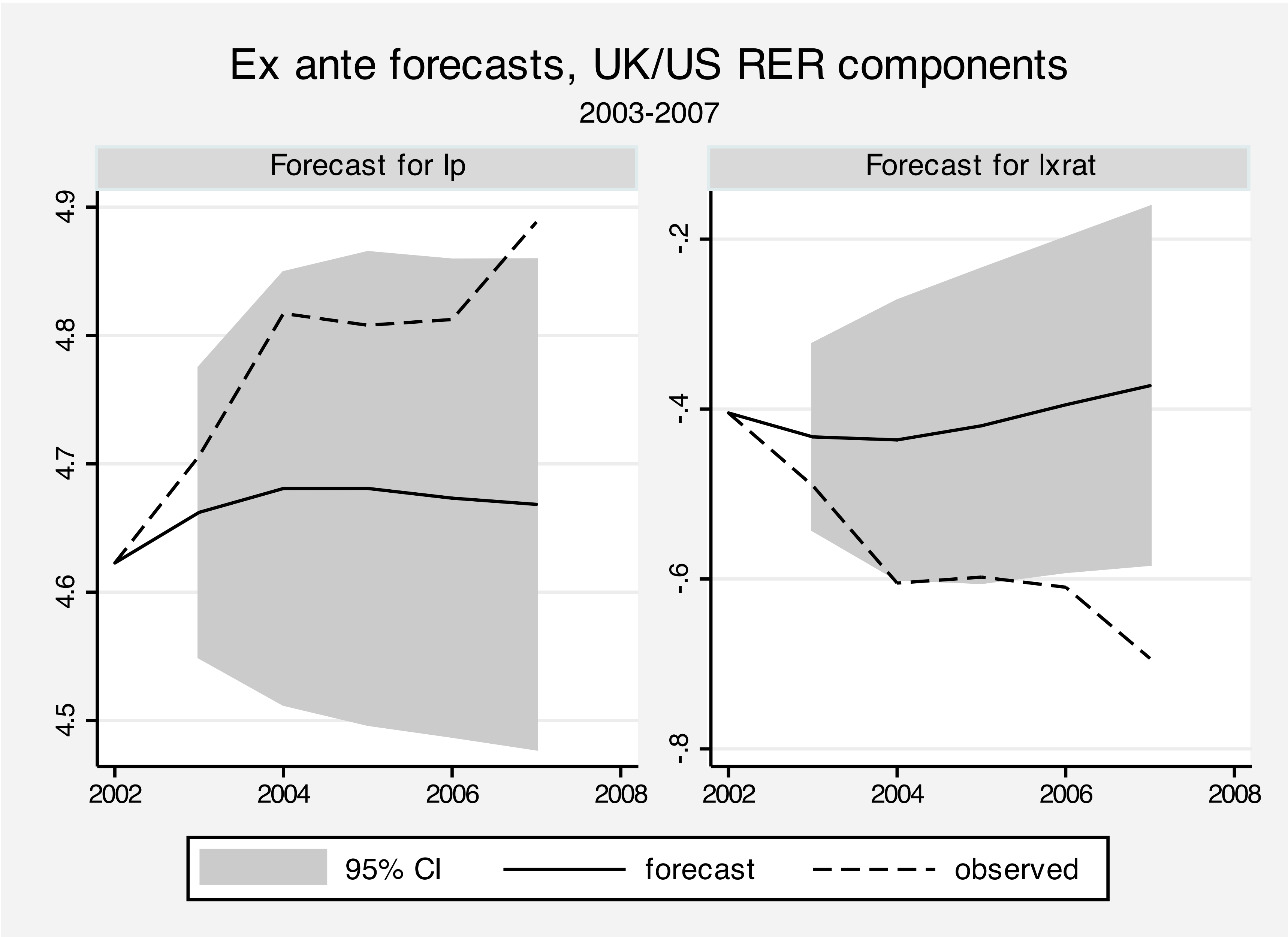
We illustrate here one feature of Stata's `vec` suite: the capability to compute dynamic forecasts from a VECM. We estimated the model on annual data through 2002, and now forecast through the end of available data in 2007:

```
 . tsset year
        time variable:  year, 1950 to 2007
                delta:  1 year

. fcast compute ppp_, step(5)

. fcast graph ppp_lp ppp_lxrat, observed scheme(s2mono) legend(rows(1)) ///
> byopts(ti("Ex ante forecasts, UK/US RER components") t2("2003-2007"))
```

We can use much of the same post-estimation apparatus as developed for VARs for VECMs. Impulse response functions, orthogonalized IRFs, FEVDs, and the like can be constructed for VECMs. However, the presence of the integrated variables (and unit moduli) in the VECM representation implies that shocks may be permanent as well as transitory.

We illustrate here one feature of Stata's `vec` suite: the capability to compute dynamic forecasts from a VECM. We estimated the model on annual data through 2002, and now forecast through the end of available data in 2007:

```
 . tsset year
        time variable:  year, 1950 to 2007
                delta:  1 year
. fcast compute ppp_, step(5)
. fcast graph ppp_lp ppp_lxrat, observed scheme(s2mono) legend(rows(1)) ///
> byopts(ti("Ex ante forecasts, UK/US RER components") t2("2003-2007"))
```


Ex ante forecasts, UK/US RER components
2003-2007
Forecast for lp
Forecast for lxrat
4.5
4.6
4.7
4.8
4.9
-.8
-.6
-.4
-.2
2002
2004
2006
2008
95% CI
forecast
observed

We see that the model's predicted log relative price was considerably lower than that observed, while the predicted log nominal exchange rate was considerably higher than that observed over this out-of-sample period.

Consult the online Stata *Time Series* manual for much greater detail on Stata's VECM capabilities, applications to multiple-variable systems and alternative treatments of deterministic trends in the VECM context.

We see that the model's predicted log relative price was considerably lower than that observed, while the predicted log nominal exchange rate was considerably higher than that observed over this out-of-sample period.

Consult the online Stata *Time Series* manual for much greater detail on Stata's VECM capabilities, applications to multiple-variable systems and alternative treatments of deterministic trends in the VECM context.