

CHAPTER 5

Vector Autoregression and Vector Error-Correction Models

Vector autoregression (VAR) was introduced by Sims (1980) as a technique that could be used by macroeconomists to characterize the joint dynamic behavior of a collection of variables without requiring strong restrictions of the kind needed to identify underlying structural parameters. It has become a prevalent method of time-series modeling.

Although estimating the equations of a VAR does not require strong identification assumptions, some of the most useful applications of the estimates, such as calculating *impulse-response functions* (IRFs) or *variance decompositions* do require identifying restrictions. A typical restriction takes the form of an assumption about the dynamic relationship between a pair of variables, for example, that x affects y only with a lag, or that x does not affect y in the long run.

A VAR system contains a set of m variables, each of which is expressed as a linear function of p lags of itself and of all of the other $m - 1$ variables, plus an error term. (It is possible to include exogenous variables such as seasonal dummies or time trends in a VAR, but we shall focus on the simple case.) With two variables, x and y , an order- p VAR would be the two equations

$$\begin{aligned}y_t &= \beta_{y0} + \beta_{yy1}y_{t-1} + \dots + \beta_{yyp}y_{t-p} + \beta_{yx1}x_{t-1} + \dots + \beta_{xyp}x_{t-p} + v_t^y \\x_t &= \beta_{x0} + \beta_{xy1}y_{t-1} + \dots + \beta_{xyp}y_{t-p} + \beta_{xx1}x_{t-1} + \dots + \beta_{xxp}x_{t-p} + v_t^x.\end{aligned}\tag{5.1}$$

We adopt the subscript convention that β_{xyp} represents the coefficient of y in the equation for x at lag p . If we were to add another variable z to the system, there would be a third equation for z_t and terms involving p lagged values of z , for example, β_{xzp} , would be added to the right-hand side of each of the three equations.

A key feature of equations (5.1) is that no current variables appear on the right-hand side of any of the equations. This makes it plausible, though not always certain, that the regressors of (5.1) are weakly exogenous and that, if all of the variables are stationary and ergodic,

OLS can produce asymptotically desirable estimators. Variables that are known to be exogenous—a common example is seasonal dummy variables—may be added to the right-hand side of the VAR equations without difficulty, and obviously without including additional equations to model them. Our examples will not include such exogenous variables.

The error terms in (5.1) represent the parts of y_t and x_t that are not related to past values of the two variables: the unpredictable “innovation” in each variable. These innovations will, in general, be correlated with one another because there will usually be some tendency for movements in y_t and x_t to be correlated, perhaps because of a contemporaneous causal relationship (or because of the common influence of other variables).

A key distinction in understanding and applying VARs is between the innovation terms v in the VAR and underlying exogenous, orthogonal shocks to each variable, which we shall call ε . The innovation in y_t is the part of y_t that cannot be predicted by *past* values of x and y . Some of this unpredictable variation in y_t that we measure by v_t is surely due to ε_t^y , an exogenous shock to y_t that has no relationship to what is happening with x or any other variable that might be included in the system. However, if x has a contemporaneous effect on y , then some part of v_t^y will be due to the indirect effect of the current shock to x , ε_t^x , which enters the y_t equation in (5.1) through the error term because current x_t is not allowed to be on the right-hand side. We will study in the next section how, by making identifying assumptions, we can identify the exogenous shocks ε from our estimates of the VAR coefficients and residuals.

Correlation between the error terms of two equations, such as that present in (5.1), usually means that we can gain efficiency by using the seemingly unrelated regressions (SUR) system estimator rather than estimating the equations individually by OLS. However, the VAR system conforms to the one exception to that rule: the regressors of all of the equations are identical, meaning that SUR and OLS lead to identical estimators. The only situation in which we gain by estimating the VAR as a system of seemingly unrelated regressions is when we impose restrictions on the coefficients of the VAR, a case that we shall ignore here.

When the variables of a VAR are cointegrated, we use a vector error-correction (VEC) model. A VEC for two variables might look like

$$\begin{aligned}\Delta y_t &= \beta_{y0} + \beta_{y1}\Delta y_{t-1} + \dots + \beta_{yp}\Delta y_{t-p} + \gamma_{y1}\Delta x_{t-1} + \dots + \gamma_{yp}\Delta x_{t-p} \\ &\quad - \lambda_y (y_{t-1} - \alpha_0 - \alpha_1 x_{t-1}) + v_t^y \\ \Delta x_t &= \beta_{x0} + \beta_{x1}\Delta y_{t-1} + \dots + \beta_{xp}\Delta y_{t-p} + \gamma_{x1}\Delta x_{t-1} + \dots + \gamma_{xp}\Delta x_{t-p} \\ &\quad - \lambda_x (y_{t-1} - \alpha_0 - \alpha_1 x_{t-1}) + v_t^x,\end{aligned}\tag{5.2}$$

where $y_t = \alpha_0 + \alpha_1 x_t$ is the long-run cointegrating relationship between the two variables and λ_y and λ_x are the error-correction parameters that measure how y and x react to deviations from long-run equilibrium.

When we apply the VEC model to more than two variables, we must consider the possibility that more than one cointegrating relationship exists among the variables. For example, if x , y , and z all tend to be equal in the long run, then $x_t = y_t$ and $y_t = z_t$ (or, equivalently, $x_t = z_t$) would be two cointegrating relationships. To deal with this situation we need to generalize the procedure for testing for cointegrating relationships to allow more than one cointegrating equation, and we need a model that allows multiple error-correction terms in each equation.

5.1 Forecasting and Granger Causality in a VAR

In order to identify structural shocks and their dynamic effects we must make additional identification assumptions. However, a simple VAR system such as (5.1) can be used for two important econometric tasks *without* making any additional assumptions. We can use (5.1) as a convenient method to generate forecasts for x and y , and we can attempt to infer information about the direction or directions of causality between x and y using the technique of *Granger causality* analysis.

5.1.1 Forecasting with a VAR

The structure of equations (5.1) is designed to model how the values of the variables in period t are related to past values. This makes the VAR a natural for the task of forecasting the future paths of x and y conditional on their past histories.

Suppose that we have a sample of observations on x and y that ends in period T , and that we wish to forecast their values in $T + 1$, $T + 2$, etc. To keep the algebra simple, suppose that $p = 1$, so there is only one lagged value on the right-hand side. For period $T + 1$, our VAR is

$$\begin{aligned} y_{T+1} &= \beta_{y0} + \beta_{yy1}y_T + \beta_{yx1}x_T + v_{T+1}^y \\ x_{T+1} &= \beta_{x0} + \beta_{xy1}y_T + \beta_{xx1}x_T + v_{T+1}^x. \end{aligned} \tag{5.3}$$

Taking the expectation conditional on the relevant information from the sample (x_T and y_T) gives

$$\begin{aligned} E(y_{T+1} | x_T, y_T) &= \beta_{y0} + \beta_{yy1}y_T + \beta_{yx1}x_T + E(v_{T+1}^y | x_T, y_T) \\ E(x_{T+1} | x_T, y_T) &= \beta_{x0} + \beta_{xy1}y_T + \beta_{xx1}x_T + E(v_{T+1}^x | x_T, y_T). \end{aligned} \tag{5.4}$$

The conditional expectation of the VAR error terms on the right-hand side must be zero in order for OLS to estimate the coefficients consistently. Whether or not this assumption is valid will depend on the serial correlation properties of the v terms—we have seen that serially correlated errors and lagged dependent variables of the kind present in the VAR can be a toxic combination.

Thus, we want to make sure that $E(v_t^j | v_{t-1}^x, v_{t-1}^y) = 0$. As we saw in an earlier chapter, adding lagged values of y and x can often eliminate serial correlation of the error, and this method is now more common than using GLS procedures to correct for possible autocorrelation. We assume that our VAR system has sufficient lag length that the error term is not serially correlated, so that the conditional expectation of the error term for all periods after T is zero. This means that the final term on the right-hand side of each equation in (5.4) is zero, so

$$\begin{aligned} E(y_{T+1} | x_T, y_T) &= \beta_{y0} + \beta_{yy1}y_T + \beta_{yx1}x_T \\ E(x_{T+1} | x_T, y_T) &= \beta_{x0} + \beta_{xy1}y_T + \beta_{xx1}x_T. \end{aligned} \quad (5.5)$$

If we knew the β coefficients, we could use (5.5) to calculate a forecast for period $T + 1$. Naturally, we use our estimated VAR coefficients in place of the true values to calculate our predictions

$$\begin{aligned} P(y_{T+1} | x_T, y_T) &\equiv \hat{y}_{T+1|T} = \hat{\beta}_{y0} + \hat{\beta}_{yy1}y_T + \hat{\beta}_{yx1}x_T \\ P(x_{T+1} | x_T, y_T) &\equiv \hat{x}_{T+1|T} = \hat{\beta}_{x0} + \hat{\beta}_{xy1}y_T + \hat{\beta}_{xx1}x_T. \end{aligned} \quad (5.6)$$

The forecast error in the predictions in (5.6) will come from two sources: the unpredictable period $T + 1$ error term and the errors we make in estimating the β coefficients. Formally,

$$\begin{aligned} y_{T+1} - \hat{y}_{T+1|T} &= (\beta_{y0} - \hat{\beta}_{y0}) + (\beta_{yy1} - \hat{\beta}_{yy1})y_T + (\beta_{yx1} - \hat{\beta}_{yx1})x_T + v_{T+1}^y \\ x_{T+1} - \hat{x}_{T+1|T} &= (\beta_{x0} - \hat{\beta}_{x0}) + (\beta_{xy1} - \hat{\beta}_{xy1})y_T + (\beta_{xx1} - \hat{\beta}_{xx1})x_T + v_{T+1}^x. \end{aligned}$$

If our estimates of the β coefficients are consistent and there is no serial correlation in v , then the expectation of the forecast error is asymptotically zero. The variance of the forecast error is

$$\begin{aligned} \text{var}(y_{T+1} - \hat{y}_{T+1|T}) &= \text{var}(\hat{\beta}_{y0}) + \text{var}(\hat{\beta}_{yy1})y_T^2 + \text{var}(\hat{\beta}_{yx1})x_T^2 \\ &\quad + 2\text{cov}(\hat{\beta}_{y0}, \hat{\beta}_{yy1})y_T + 2\text{cov}(\hat{\beta}_{y0}, \hat{\beta}_{yx1})x_T + 2\text{cov}(\hat{\beta}_{yy1}, \hat{\beta}_{yx1})x_T y_T \\ &\quad + \text{var}(v_{T+1}^y) \\ \text{var}(x_{T+1} - \hat{x}_{T+1|T}) &= \text{var}(\hat{\beta}_{x0}) + \text{var}(\hat{\beta}_{xy1})y_T^2 + \text{var}(\hat{\beta}_{xx1})x_T^2 \\ &\quad + 2\text{cov}(\hat{\beta}_{x0}, \hat{\beta}_{xy1})y_T + 2\text{cov}(\hat{\beta}_{x0}, \hat{\beta}_{xx1})x_T + 2\text{cov}(\hat{\beta}_{xy1}, \hat{\beta}_{xx1})x_T y_T \\ &\quad + \text{var}(v_{T+1}^x). \end{aligned}$$

As our consistent estimates of the β coefficients converge to the true values (as T gets large), all of the terms in this expression converge to zero except the last one. Thus, in calculating

the variance of the forecast error, the error in estimating the coefficients is often neglected, giving

$$\begin{aligned}\text{var}(y_{T+1} - \hat{y}_{T+1|T}) &\approx \text{var}(v_{T+1}^y) \equiv \sigma_{v,y}^2 \\ \text{var}(x_{T+1} - \hat{x}_{T+1|T}) &\approx \text{var}(v_{T+1}^x) \equiv \sigma_{v,x}^2.\end{aligned}\quad (5.7)$$

One of the most useful attributes of the VAR is that it can be used recursively to extend forecasts into the future. For period $T + 2$,

$$\begin{aligned}E(y_{T+2} | x_{T+1}, y_{T+1}) &= \beta_{y0} + \beta_{yy1}y_{T+1} + \beta_{yx1}x_{T+1} \\ E(x_{T+2} | x_{T+1}, y_{T+1}) &= \beta_{x0} + \beta_{xy1}y_{T+1} + \beta_{xx1}x_{T+1},\end{aligned}$$

so by recursive expectations

$$\begin{aligned}E(y_{T+2} | x_T, y_T) &= \beta_{y0} + \beta_{yy1}E(y_{T+1} | x_T, y_T) + \beta_{yx1}E(x_{T+1} | x_T, y_T) \\ &= \beta_{y0} + \beta_{yy1}(\beta_{y0} + \beta_{yy1}y_T + \beta_{yx1}x_T) + \beta_{yx1}(\beta_{x0} + \beta_{xy1}y_T + \beta_{xx1}x_T) \\ E(x_{T+2} | x_T, y_T) &= \beta_{x0} + \beta_{xy1}E(y_{T+1} | x_T, y_T) + \beta_{xx1}E(x_{T+1} | x_T, y_T) \\ &= \beta_{x0} + \beta_{xy1}(\beta_{y0} + \beta_{yy1}y_T + \beta_{yx1}x_T) + \beta_{xx1}(\beta_{x0} + \beta_{xy1}y_T + \beta_{xx1}x_T).\end{aligned}$$

The corresponding forecasts are again obtained by substituting coefficient estimates to get

$$\begin{aligned}P(y_{T+2} | x_T, y_T) &\equiv \hat{y}_{T+2|T} = \hat{\beta}_{y0} + \hat{\beta}_{yy1}\hat{y}_{T+1|T} + \hat{\beta}_{yx1}\hat{x}_{T+1|T} \\ P(x_{T+2} | x_T, y_T) &\equiv \hat{x}_{T+2|T} = \hat{\beta}_{x0} + \hat{\beta}_{xy1}\hat{y}_{T+1|T} + \hat{\beta}_{xx1}\hat{x}_{T+1|T}.\end{aligned}\quad (5.8)$$

If we once again ignore error in estimating the coefficients, then the two-period-ahead forecast error in (5.8) is

$$\begin{aligned}y_{T+2} - \hat{y}_{T+2|T} &\approx \beta_{yy1}(y_{T+1} - y_{T+1|T}) + \beta_{yx1}(x_{T+1} - x_{T+1|T}) + v_{T+2}^y \\ &\approx \beta_{yy1}v_{T+1}^y + \beta_{yx1}v_{T+1}^x + v_{T+2}^y, \\ x_{T+2} - \hat{x}_{T+2|T} &\approx \beta_{xy1}(y_{T+1} - y_{T+1|T}) + \beta_{xx1}(x_{T+1} - x_{T+1|T}) + v_{T+2}^x \\ &\approx \beta_{xy1}v_{T+1}^y + \beta_{xx1}v_{T+1}^x + v_{T+2}^x.\end{aligned}$$

In general, the error terms for period $T + 1$ will be correlated across equations, so the variance of the two-period-ahead forecast is approximately

$$\begin{aligned}\text{var}(y_{T+2} - \hat{y}_{T+2|T}) &\approx \beta_{yy1}^2\sigma_{v,y}^2 + \beta_{yx1}^2\sigma_{v,x}^2 + 2\beta_{yy1}\beta_{yx1}\sigma_{v,xy} + \sigma_{v,y}^2 \\ &= (1 + \beta_{yy1}^2)\sigma_{v,y}^2 + \beta_{yx1}^2\sigma_{v,x}^2 + 2\beta_{yy1}\beta_{yx1}\sigma_{v,xy}, \\ \text{var}(x_{T+2} - \hat{x}_{T+2|T}) &\approx \beta_{xy1}^2\sigma_{v,y}^2 + \beta_{xx1}^2\sigma_{v,x}^2 + 2\beta_{xy1}\beta_{xx1}\sigma_{v,xy} + \sigma_{v,x}^2 \\ &= \beta_{xy1}^2\sigma_{v,y}^2 + (1 + \beta_{xx1}^2)\sigma_{v,x}^2 + 2\beta_{xy1}\beta_{xx1}\sigma_{v,xy}.\end{aligned}\quad (5.9)$$

The two-period-error forecast error has larger variance than the one-period-ahead error because the errors that we make in forecasting period $T + 1$ propagate into errors in the forecast for $T + 2$. As our forecast horizon increases, the variance gets larger and larger, reflecting our inability to forecast a great distance into the future even if (as we have optimistically assumed here) we have accurate estimates of the coefficients.

The calculations in equation (5.9) become increasingly complex as one considers longer forecast horizons. Including more than two variables in the VAR or more than one lag on the right-hand side also increases the number of terms in both (5.8) and (5.9) rapidly. We are fortunate that modern statistical software, including Stata, has automated these tasks for us. We now discuss the basics of estimating a VAR in Stata.

5.1.2 Estimating and forecasting a simple VAR in Stata

At one level, estimating a VAR is a simple task—because it is estimated with OLS, the Stata regression command will handle the estimation. However, for everything we do with a VAR beyond estimation, we need to consider the system as a whole, so Stata provides a family of procedures that are tailored to the VAR application. The two essential VAR commands are `var` and `varbasic`. The latter is easy to use (potentially as easy as listing the variables you want in the system), but lacks the flexibility of the former to deal with asymmetric lag patterns across equations, additional exogenous variables that have no equations of their own, and coefficient constraints across equations. We discuss `var` first; later in the chapter we will go back and show the use of the simpler `varbasic` command. (As always, only simple examples of Stata commands are shown here. The current Stata manual available through the Stata Help menu contains full documentation of all options and variations, along with additional examples.)

To run a simple VAR for variables x and y with two lags or each variable in each equation and no constraints or exogenous variables, we can simply type

```
var x y , lags(1/2)
```

Notice that we need 1/2 rather than just 2 in the lag specification because we want lags 1 through 2, not just the second lag. The output from this command will give the β coefficients from OLS estimation of the two regressions, plus some system and individual-equation goodness-of-fit statistics.

Once we have estimated a VAR model, there are a variety of tests that can be used to help us determine whether we have a good model. In terms of model validation, one important property for our estimates to have desirable asymptotic properties is that the model must be stable in the sense that the estimated coefficients imply that $\partial y_{t+s} / \partial v_t^j$ and $\partial x_{t+s} / \partial v_t^j$ ($j = x, y$) become small as s gets large. If these conditions do not hold, then the VAR implies that x and y are not jointly ergodic: the effects of shocks do not die out.

The Stata command `varstable` (which usually needs no arguments or options) calculates the eigenvalues of a “companion matrix” for the system. If all of the calculated eigenvalues (which can be complex) are less than one (in modulus, if they have imaginary parts), then the model is stable. This condition is the vector extension to the stationarity condition that the roots of an autoregressive polynomial of a single variable lie outside the unit circle.

If the `varstable` command reports an eigenvalue with modulus greater than one, then the VAR is unstable and forecasts will explode. This can arise when the variables in the model are non-stationary or when the model is misspecified. Differencing (and perhaps, after checking for cointegration, using a VEC) may yield a stable system.

If the VAR is stable, then the main issue in specification is lag length. We discussed lag length issues above in the context of single-variable distributed lag models. The issues and methods in a VAR are similar, but apply simultaneously to all of the equations of the model and all of the variables, since we conventionally choose a common length for all lags.

Forecasting with a VAR assumes that there is no serial correlation in the error term. The Stata command `varlmar` implements a VAR version of the Lagrange multiplier test for serial correlation in the residual. This command tests the null hypotheses $\text{cov}(v_t^j, v_{t-s}^j) = 0$ with j indexing the variables of the model. The main option in the `varlmar` command allows you to specify the highest order of autocorrelation (the default is 2) that you want to test in the residual. For example `varlmar , mlag(4)` would perform the above test individually for $s = 1, s = 2, s = 3,$ and $s = 4$. If the Lagrange multiplier test rejects the null hypothesis of no serial correlation, then you may want to include additional lags in the equations and perform the test again.

The Akaike Information Criterion (AIC) and Schwartz-Bayesian Information Criterion (SBIC) are often used to choose the optimal lag length in single-variable distributed-lag models. These and other criteria have been extended to the VAR case and are reported by the `varsoc` command. Typing `varsoc , maxlag(4)` tells Stata to estimate VARs for lag length 0, (just constants), 1, 2, 3, and 4, and compute the log-likelihood function and various information criteria for each choice. The output of the `varsoc` command includes likelihood-ratio test statistics for the null hypothesis that the next lag is zero. The optimal lag length by each criterion is indicated by an asterisk in the table of results. In general, the various criteria will not agree, so you will need to exercise some degree of judgment in choosing among the recommendations.

Another way of deciding on lag length is to use standard (Wald) test statistics to test whether all of the coefficients at each lag are zero. Stata automates this in the `varwle` command, which requires no options.

Once you have settled on a VAR model that includes an appropriate number of lags, is stable, and has serially uncorrelated errors, you can proceed to use the model to generate forecasts. There are two commands for creating and graphing forecasts. The `fcst compute` command calculates the predictions of the VAR and stores them in a set of new vari-

ables. If you want your forecasts to start in the period immediately following the last period of the estimating sample, then the only option you need in the `fcast compute` command is `step(#)`, with which you specify the forecast horizon (how many periods ahead you want to forecast). The forecast variables are stored in variables that attach a prefix that you specify to the names of the VAR variables being forecasted. For example, to forecast your VAR model for 10 periods beginning after the estimating sample and store predicted values of x in `pred_x` and y in `pred_y`, you could type

```
fcast compute pred_ , step(10)
```

The `fcast compute` command also generates standard errors of the forecasts and uses them to calculate upper and lower confidence bounds. After computing the forecasts, you can graph them along with the confidence by typing `fcast graph pred_*`. If you have actual observed values for the variables for the forecast period, they can be added to the graph with the `observed` option (separated from the command by a comma, as always with Stata options).

5.1.3 Granger causality

One of the first, and undeniable, maxims that every econometrician or statistician is taught is that “correlation does not imply causality.” Correlation or covariance is a symmetric, bivariate relationship; $\text{cov}(x, y) = \text{cov}(y, x)$. We cannot, in general, infer anything about the existence or direction of causality between x and y by observing non-zero covariance. Even if our statistical analysis is successful in establishing that the covariance is highly unlikely to have occurred by chance, such a relationship could occur because x causes y , because y causes x , because each causes the other, or because x and y are responding to some third variable without any causal relationship between them.

However, Clive Granger defined the concept of *Granger causality*, which, under some controversial assumptions, can be used to shed light on the direction of possible causality between pairs of variables. The formal definition of Granger causality asks whether past values of x aid in the prediction of y_t , conditional on having already accounted for the effects on y_t of past values of y (and perhaps of past values of other variables). If they do, the x is said to “Granger cause” y .

The VAR is a natural framework for examining Granger causality. Consider the two-variable system in equations (5.1). The first equation models y_t as a linear function of its own past values, plus past values of x . If x Granger causes y (which we write as $x \Rightarrow y$), then some or all of the lagged x values have non-zero effects: lagged x affects y_t conditional on the effects of lagged y . Testing for Granger causality in (5.1) amounts to testing the joint blocks of coefficients β_{yx} and β_{yx} to see if they are zero. The null hypothesis $x \not\Rightarrow y$ (x does not Granger cause y) in this VAR is

$$H_0 : \beta_{yx1} = \beta_{yx2} = \dots = \beta_{yxp} = 0,$$

which can be tested using a standard Wald F or χ^2 test. Similarly, the null hypothesis $y \not\Rightarrow x$ can be expressed in the VAR as

$$H_0 : \beta_{xy1} = \beta_{xy2} = \dots = \beta_{xyp} = 0.$$

Running both of these tests can yield four possible outcomes, as shown in Table 5-1: no Granger causality, one-way Granger causality in either direction, or “feedback,” with Granger causality running both ways.

Table 5-1. Granger causality test outcomes

	Fail to reject: $\beta_{yx1} = \beta_{yx2} = \dots = \beta_{yxs} = 0$	Reject: $\beta_{yx1} = \beta_{yx2} = \dots = \beta_{yxs} = 0$
Fail to reject: $\beta_{xy1} = \beta_{xy2} = \dots = \beta_{xys} = 0$	$y \not\Rightarrow x$ $x \not\Rightarrow y$ (no Granger causality)	$y \not\Rightarrow x$ $x \Rightarrow y$ (x Granger causes y)
Reject: $\beta_{xy1} = \beta_{xy2} = \dots = \beta_{xys} = 0$	$y \Rightarrow x$ $x \not\Rightarrow y$ (y Granger causes x)	$y \Rightarrow x$ $x \Rightarrow y$ (bi-directional Granger causality, or feedback)

There are multiple ways to perform Granger causality tests between a pair of variables, so no result is unique or definitive. Within the two-variable VAR, one may obtain different results with different lag lengths p . Moreover, including additional variables in the VAR system may change the outcome of the Wald tests that underpin Granger causality. In a three-variable VAR, there are three pairs of variables, (x, y) , (y, z) , and (x, z) that can be tested for Granger causality in both directions: six tests with 36 possible combinations of outcomes. The effect of lagged x on y_t can disappear when lagged values of a third variable z are added to the regression. For example, if $x \Rightarrow z$ and $z \Rightarrow y$, then omitting z from the VAR system could lead us to conclude that $x \Rightarrow y$ even if there is no direct Granger causality in the larger system.

Is “Granger causality” really “causality”? Obviously, if the maxim about correlation and causality is true, then there must be something tricky happening, and indeed there is. Granger causality tests whether lagged values of one variable conditionally help predict another variable. Under what conditions can we interpret this as “causality”? Two assumptions are sufficient.

First, we use temporal priority in an important way in Granger causality. We interpret correlation between lagged x and the part of current y that is orthogonal to its own lagged values as $x \Rightarrow y$. Could this instead reflect current y “causing” lagged x ? To rule this out, interpreting Granger causality as more general causality requires that we assume that *the future cannot cause the present*. While this may often be a reasonable assumption, modern economic theory has shown us that expectations of future variables (which are likely to be corre-

lated with the future variables themselves) can change agents' current choices, which might result in causality that would appear to violate this assumption.

Second, any causal relationship that is strictly immediate in the sense that a change in x_t leads to a change in y_t but no change in any future values of y would fly under the radar of a Granger causality test, which only measures and tests lagged effects. Most causal economic relationships are dynamic in that effects are not fully realized within a single time period, so this difficulty may not present a practical problem in many cases.

To summarize, we must be very careful in interpreting the result of Granger causality tests to reflect true causality in any non-econometric sense. Only if we can rule out the possibility of the future causing the present and strictly immediate causal effects can we confidently think of “Granger causality” as “causality.”

Stata implements Granger causality tests automatically with `vargranger`, which tests all of the pairs of variables in a VAR for Granger causality. In systems with more than two variables, it also tests the joint hypothesis that all of the other variables fail to Granger cause each variable in turn. This joint test amounts to testing whether all of the lagged terms other than those of the dependent variable have zero effects.

5.2 Identification of Structural Shocks in a VAR System

Two variables that have a dynamic relationship in a VAR system are also likely to have some degree of contemporaneous association. This will be reflected in correlation in the innovation terms ν in (5.1) because there is no other place in the equations for this association to be manifested. It is natural to think of the VAR system as the reduced form of a structural model in which contemporaneous effects among the variables have been “solved out.” We now consider a simple two-equation structural model in which x_t affects y_t contemporaneously but y_t has no immediate effect on x_t .

Suppose that two variables, x and y , evolve over time according to the structural model

$$\begin{aligned} x_t &= \alpha_0 + \alpha_1 x_{t-1} + \theta_1 y_{t-1} + \varepsilon_t^x \\ y_t &= \phi_0 + \phi_1 y_{t-1} + \delta_0 x_t + \delta_1 x_{t-1} + \varepsilon_t^y, \end{aligned} \tag{5.10}$$

where the ε terms are exogenous white-noise shocks to x and y that are “orthogonal” (uncorrelated) to one another: $\text{var}(\varepsilon_t^x) = \sigma_x^2$, $\text{var}(\varepsilon_t^y) = \sigma_y^2$, and $\text{cov}(\varepsilon_t^x, \varepsilon_t^y) = 0$. The ε shocks are changes in the variables that come from outside the VAR system. Because they are (assumed to be) exogenous, we can measure the effect of an exogenous change in x on the path of y and x by looking at the dynamic marginal effects of ε_t^x , for example, $\partial y_{t+s} / \partial \varepsilon_t^x$. This is the key distinction between the VAR error terms ν and the exogenous structural shocks ε —depending on the identifying assumptions we make, we cannot generally interpret a change in ν as an exogenous shock to one variable.

We assume that x and y are stationary and ergodic, which imposes restrictions on the autoregressive coefficients of the model.¹

The first equation of (5.10) is already in the form of a VAR equation: it expresses the current value of x_t as a function of lagged values of x and y . If we solve (5.10) by substituting for x_t in the second equation using the first equation, we get

$$\begin{aligned} y_t &= \phi_0 + \phi_1 y_{t-1} + \delta_0 (\alpha_0 + \alpha_1 x_{t-1} + \theta_1 y_{t-1} + \varepsilon_t^x) + \delta_1 x_{t-1} + \varepsilon_t^y \\ &= (\phi_0 + \delta_0 \alpha_0) + (\phi_1 + \delta_0 \theta_1) y_{t-1} + (\delta_1 + \delta_0 \alpha_1) x_{t-1} + (\varepsilon_t^y + \delta_0 \varepsilon_t^x), \end{aligned} \quad (5.11)$$

which also has the VAR form. Thus, we can write the reduced-form system of (5.10) as

$$\begin{aligned} y_t &= \beta_{y0} + \beta_{yy1} y_{t-1} + \beta_{yx1} x_{t-1} + v_t^y \\ x_t &= \beta_{x0} + \beta_{xy1} y_{t-1} + \beta_{xx1} x_{t-1} + v_t^x, \end{aligned} \quad (5.12)$$

with

$$\begin{aligned} \beta_{y0} &= \phi_0 + \delta_0 \alpha_0 & \beta_{x0} &= \alpha_0 \\ \beta_{yx1} &= \phi_1 + \delta_0 \theta_1 & \beta_{xy1} &= \theta_1 \\ \beta_{yy1} &= \delta_1 + \delta_0 \alpha_1 & \beta_{xx1} &= \alpha_1 \\ v_t^y &= \varepsilon_t^y + \delta_0 \varepsilon_t^x & v_t^x &= \varepsilon_t^x. \end{aligned} \quad (5.13)$$

Given our assumptions about the distributions of the exogenous shocks ε , we can determine the variances and covariance of the VAR error terms ν as

$$\begin{aligned} \text{var}(v_t^x) &= \sigma_x^2 \\ \text{var}(v_t^y) &= \sigma_y^2 + \delta_0^2 \sigma_x^2 \\ \text{cov}(v_t^x, v_t^y) &= E(v_t^x v_t^y) = E[\varepsilon_t^x (\varepsilon_t^y + \delta_0 \varepsilon_t^x)] = \delta_0 \sigma_x^2. \end{aligned} \quad (5.14)$$

Let's now consider what can be estimated using the VAR system (5.12) and to what extent these estimates allow us to infer value of the parameters in the structural system (5.10). In terms of coefficients, there are six β coefficients that can be estimated in the VAR and seven structural coefficients in (5.10). This seems a pessimistic start to the task of identification. However, we can also estimate the variances and covariance of the ν terms using the VAR residuals: $\widehat{\text{var}}(v_t^x)$, $\widehat{\text{var}}(v_t^y)$, and $\widehat{\text{cov}}(v_t^x, v_t^y)$. Conditions (5.14) allow us to estimate three parameters— σ_x^2 , σ_y^2 , δ_0 —from the three estimated variances and covariance:

¹ In a single-variable autoregressive model, we would require that the coefficient ϕ_1 for y_{t-1} be in the range $(-1, 1)$. The corresponding conditions in the vector setting are more involved, but similar in nature.

$$\begin{aligned}
\hat{\sigma}_x^2 &= \widehat{\text{var}}(v_t^x), \\
\hat{\delta}_0 &= \frac{\widehat{\text{cov}}(v_t^x, v_t^y)}{\widehat{\text{var}}(v_t^x)}, \\
\hat{\sigma}_y^2 &= \widehat{\text{var}}(v_t^y) - \hat{\delta}_0^2 \widehat{\text{var}}(v_t^x).
\end{aligned} \tag{5.15}$$

Armed with an estimate of δ_0 from the covariance term, we can now use the six β coefficients to estimate the remaining six structural coefficients using (5.13). The system is just identified.

So how did we manage to achieve identification through this “back-door” covariance method? Let’s consider why v_t^x and v_t^y , which are the innovations in x and y that cannot be predicted from past values, might be correlated. In a general model they could be correlated because (1) x_t has an effect on y_t , (2) y_t has an effect on x_t , or (3) the exogenous structural shocks to x and y , ε_t^x and ε_t^y , are correlated with one another. Our VAR estimates give us no method of discriminating among these three possible sources of correlation, so without ruling out two of them we cannot achieve identification. In the model of (5.10), we have ruled out, by assumption, (2) and (3): y_t does not affect x_t and the exogenous shocks to x and y are orthogonal. Thus, we are *interpreting* the covariance between v_t^x and v_t^y as reflecting the contemporaneous effect of x_t on y_t . This allows us to identify the coefficient measuring this effect— δ_0 in (5.10)—based on the $\text{cov}(v_t^x, v_t^y)$ as we do in the second equation of (5.15).

This identifying assumption also allows us to reconstruct estimates of the exogenous structural shocks ε from the residuals of the VAR:

$$\begin{aligned}
\hat{\varepsilon}_t^x &= \hat{v}_t^x \\
\hat{\varepsilon}_t^y &= \hat{v}_t^y - \hat{\delta}_0 \hat{v}_t^x.
\end{aligned}$$

This makes it clear that we are *interpreting* the VAR residual for x to be an exogenous, structural shock to x . In order to extract the structural shock to y , we subtract the part of \hat{v}_t^y , $\hat{\delta}_0 \hat{v}_t^x = \hat{\delta}_0 \hat{\varepsilon}_t^x$, that is due to the effect of the shock to x_t on y_t . From an econometric standpoint, we could equally well make the opposite assumption, assuming that y_t affects x_t rather than vice versa, which would interpret \hat{v}_t^y as $\hat{\varepsilon}_t^y$ and calculate $\hat{\varepsilon}_t^x$ as the part of \hat{v}_t^x that is not explained by \hat{v}_t^y . Choosing which interpretation to use must be done on the basis of theory: which variable is more plausibly exogenous within the immediate period. We may get different results depending on which identification assumption we choose, so if there is no clear choice it may be useful to examine whether results are robust across different choices.

Identification of the underlying structural shocks is necessary if we are to estimate the effects of an exogenous shock to a single variable on the dynamic paths of all of the variables

of the system, which we call *impulse-response functions* (IRFs). We discuss the computation and interpretation of IRFs in the next section.

In our example, we identified shocks by limiting the contemporaneous effects among the variables. With only two variables, there are two possible choices: (1) the assumption we made, that x_t affects y_t immediately but y_t does not have an immediate effect on x_t or (2) the opposite assumption, that y_t affects x_t immediately but x_t does not affect y_t except with a lag. We can think of the choice between these alternatives as an “ordering” of the variables, with the variables lying higher in the order having instantaneous effects on those lower in the order, but the lower variables only affecting those above them with a lag.

This ordering or “orthogonalization” strategy of identification extends directly to VAR systems with more than two variables. Sims’s seminal VAR system had six variables, which he ordered as the money supply, real output, unemployment, wages, prices, and import prices. By adopting this ordering, Sims was imposing an array of identifying restrictions about the contemporaneous effects of shocks on the variables of the system. The money shock, because it was at the top of the list, could affect all of the variables in the system within the current period. The shock to output affects all variables immediately except money, because money lies above it on the list. The variable at the bottom of the list, import prices, is assumed to have no contemporaneous effect on any of the other variables of the system.

Although identification by ordering is still common, subsequent research has shown that other kinds of restrictions can be used. For example, in some macroeconomic models we can assume that changes in a variable such as the money supply would have no long-run effect on another variable such as real output. In a simple system such as (5.10), this might show up as the assumption that $\delta_0 + \delta_1 = 0$, for example. Imposing this condition would allow the seven structural coefficients of (5.10) to be identified from the six β coefficients of the VAR without using restrictions on the covariances.

5.3 Interpreting the Results of Identified VARs

When we can identify the structural shocks to each variable in a VAR, we can perform two kinds of analysis to explain how each shock affects the dynamic path of all of the variables of the system. Impulse-response functions (IRFs) measure the dynamic marginal effects of each shock on all of the variables over time. *Variance decompositions* examine how important each of the shocks is as a component of the overall (unpredictable) variance of each of the variables over time.

It is important to stress that, unlike forecasts and Granger causality tests, both IRFs and variance decompositions can *only* be calculated based on a set of identifying assumptions and that a different set of identification assumptions may lead to different conclusions.

Suppose that we have an n -variable VAR with lags up to order p . If the variables of the system are y^1, y^2, \dots, y^n , then we can write the n equations of the VAR as

$$y_t^i = \beta_{i0} + \sum_{j=1}^n \left(\sum_{s=1}^p \beta_{ijs} y_{t-s}^j \right) + v_t^i, \quad i = 1, 2, \dots, n. \quad (5.16)$$

We assume that we have a set of identifying restrictions on the model—either an ordering of assumed contemporaneous causality or another set of assumptions—so that we can identify the n orthogonal structural shocks ε_t^i from the n VAR error terms v_t^i .

The impulse-response functions are the $n \times n$ set of dynamic marginal effects of a one-time shock to variable j on itself or another variable i :

$$\frac{\partial y_{t+s}^i}{\partial \varepsilon_t^j}, \quad s = 0, 1, 2, \dots \quad (5.17)$$

Note that there is in principle no limit on how far into the future these dynamic impulse responses can extend. If the VAR is stable, then the IRFs should converge to zero as the time from the shock s gets large—one-time shocks should not have permanent effects. As noted above, non-convergent IRFs and unstable VARs are indications of non-stationarity in the variables of the model, which may be corrected by differencing.

IRFs are usually presented graphically with the time lag s running from zero up to some user-set limit S on the horizontal axis and the impact at the s -order lag on the vertical. They can also be expressed in tabular form if the numbers themselves are important. One common format for the entire collection of IRFs corresponding to a VAR is as an $n \times n$ matrix of graphs, with the “impulse variable” (the shock) on one dimension and the “response variable” on the other.

Each of the n^2 IRF graphs tells us how a shock to one variable affects another (or the same) variable. There are two common conventions for determining the size of the shock to the impulse variable. One is to use a shock of magnitude one. Since we can think of the impulse shock as the $\partial \varepsilon$ in the denominator of (5.17), setting the shock to one means that the values reported are the dynamic marginal effects as in (5.17).

However, a shock of size one does not always make economic sense: Suppose that the shock variable is banks’ ratio of reserves to deposits, expressed as a fraction. An increase of one in this variable, say from 0.10 to 1.10, would be implausible. To aid in interpretation, some software packages normalize the size of the shocks to be one standard deviation of the variable rather than one unit. Under this convention, the values plotted are

$$\frac{\partial y_{t+s}^i}{\partial \varepsilon_t^j} \hat{\sigma}_j, \quad s = 0, 1, 2, \dots$$

and are interpreted as the change in each response variable resulting from a one-standard-deviation increase in the impulse variable. This makes the magnitude of the changes in the response variables more realistic, but does not allow the IRF values to be interpreted directly as dynamic marginal effects.

Because the VAR model is linear, the marginal effects in (5.17) are constant, so which normalization to choose for the shocks—one unit or one standard deviation—is arbitrary and should be done to facilitate interpretation. Stata uses the convention of the one-unit impulse in its “simple IRFs” and one standard deviation in its “orthogonalized IRFs.”

If the impulse variable is the same as the response variable, then the IRF tells us how persistent shocks to that variable tend to be. By definition,

$$\frac{\partial y_t^i}{\partial \varepsilon_t^i} = 1,$$

so the zero-order own impulse response is always one. If the VAR is stable, reflecting the stationarity and ergodicity of the underlying variables, then the own impulse responses decay to zero as the time horizon increases:

$$\lim_{s \rightarrow \infty} \frac{\partial y_{t+s}^i}{\partial \varepsilon_t^i} = 0.$$

If the impulse responses decay to zero only slowly then shocks to the variable tend to change its value for many periods, whereas a short impulse response pattern indicates that shocks are more transitory.

For cross-variable effects, where the impulse and response variables are different, general patterns of positive or negative responses are possible. Depending on the identification assumption (the “ordering”), the zero-period response may be zero or non-zero. By assumption, shocks to variables near the bottom of the ordering have no current-period effect on variables higher in the order, so the zero-lag impulse response in such cases is exactly zero.

5.4 A VAR Example: GDP Growth in US and Canada

To illustrate the various applications of VAR analysis, we examine the joint behavior of US and Canadian real GDP growth using a quarterly sample from 1975q1 through 2011q4. Each of the series is an annual continuously-compounded growth rate. For example, $USGR_t = 400 \times (\ln USGDP_t - \ln USGDP_{t-1})$, with the 4 included to express the growth rate as an annual rather than quarterly rate and the 100 to put the rate in percentage terms.

5.4.1 Getting the specification right

As a preliminary check, we verify that both growth series are stationary. To be conservative, we include four lagged differences to eliminate serial correlation in the error term of the Dickey-Fuller regression.

```
. dfuller usgr , lags(4)
```

```
Augmented Dickey-Fuller test for unit root      Number of obs   =      143
```

```

----- Interpolated Dickey-Fuller -----
              Test
              Statistic      1% Critical      5% Critical      10% Critical
                              Value              Value              Value
-----
Z(t)          -4.565          -3.496          -2.887          -2.577
-----
MacKinnon approximate p-value for Z(t) = 0.0001

```

```
. dfuller cgr , lags(4)
```

```
Augmented Dickey-Fuller test for unit root      Number of obs   =      143
```

```

----- Interpolated Dickey-Fuller -----
              Test
              Statistic      1% Critical      5% Critical      10% Critical
                              Value              Value              Value
-----
Z(t)          -4.769          -3.496          -2.887          -2.577
-----
MacKinnon approximate p-value for Z(t) = 0.0001

```

In both cases, we comfortably reject the presence of a unit root in the growth series because the test statistic is more negative than the critical value, even at a 1% level of significance. Phillips-Perron tests lead to similar conclusions. Therefore, we conclude that VAR analysis can be performed on the two growth series without differencing.

To assess the optimal lag length, we use the Stata `varsoc` command with a maximum lag length of four:

```
. varsoc usgr cgr , maxlag(4)
```

```

Selection-order criteria
Sample: 1976q1 - 2011q4
Number of obs   =      144
-----+-----+-----+-----+-----+-----+-----+-----+-----+
| lag |   LL   LR   df   p   FPE   AIC   HQIC   SBIC |
-----+-----+-----+-----+-----+-----+-----+-----+
|  0  | -709.83          67.4086  9.88653  9.90329  9.92777 |
|  1  | -671.726  76.208*  4  0.000  41.9769*  9.41286*  9.46314*  9.5366* |
|  2  | -670.318  2.8151  4  0.589  43.5178  9.44887  9.53267  9.6551 |
|  3  | -667.543   5.55  4  0.235  44.2688  9.46588  9.58321  9.75461 |
|  4  | -664.14  6.8067  4  0.146  44.6449  9.47417  9.62501  9.84539 |
-----+-----+-----+-----+-----+-----+-----+-----+
Endogenous:  usgr cgr
Exogenous:   _cons

```

Note that all of the regressions leading to the numbers in the table are run for a sample beginning in 1976q1, which is the earliest date for which 4 lags are available, even though the regressions with fewer than 4 lags could use a longer sample. In this VAR, all of the criteria support a lag of length one, so that is what we choose.

5.4.2 Analysis without making identification assumptions

Although we could accomplish the tasks we desire using `varbasic`, we will use the more general commands to demonstrate their use. To run the VAR regressions, we use `var`:

```
. var usgr cgr , lags(1)
```

```
Vector autoregression
```

```
Sample: 1975q2 - 2011q4                No. of obs   =       147
Log likelihood = -686.0263              AIC          =   9.415324
FPE           =  42.08037                HQIC        =   9.464918
Det(Sigma_ml) =  38.78127                SBIC       =   9.537383
```

Equation	Parms	RMSE	R-sq	chi2	P>chi2
usgr	3	2.95659	0.1789	32.02707	0.0000
cgr	3	2.39083	0.3861	92.43759	0.0000

		Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
usgr							
	usgr						
	L1.	.2512771	.0898504	2.80	0.005	.0751735	.4273807
	cgr						
	L1.	.2341061	.0975328	2.40	0.016	.0429453	.4252668
	_cons	1.494612	.3354005	4.46	0.000	.8372394	2.151985
cgr							
	usgr						
	L1.	.3759117	.0726571	5.17	0.000	.2335065	.5183169
	cgr						
	L1.	.2859551	.0788694	3.63	0.000	.1313739	.4405362
	_cons	.8755421	.2712199	3.23	0.001	.3439609	1.407123

We have not yet attempted any shock identification, so at this point the ordering of the variables in the command is arbitrary. The VAR regressions are run starting the sample at the earliest possible date with one lag, which is 1975q2 because our first available observation is 1975q1. Because it uses three additional observations, the reported AIC and SBIC values from the VAR output do not match those from the `varsoc` table above.

To assess the validity of our VAR, we test for stability and for autocorrelation of the residuals. The `varstable` command examines the dynamic stability of the system. None of the eigenvalues is even close to one, so our system is stable.

```
. varstable

Eigenvalue stability condition
+-----+
|          Eigenvalue          |      Modulus      |
+-----+-----+
|      .5657757                |      .565776     |
|     -.02854353              |      .028544     |
+-----+-----+
All the eigenvalues lie inside the unit circle.
VAR satisfies stability condition.
```

The `varlmar` performs a Lagrange multiplier test for the joint null hypothesis of no autocorrelation of the residuals of the two equations:

```
. varlmar , mlag(4)

Lagrange-multiplier test
+-----+
| lag |      chi2    df   Prob > chi2 |
+-----+-----+
|   1 |      4.6880    4   0.32084 |
|   2 |      5.5347    4   0.23670 |
|   3 |      3.5253    4   0.47404 |
|   4 |      7.1422    4   0.12856 |
+-----+-----+
H0: no autocorrelation at lag order
```

We cannot reject the null of no residual autocorrelation at orders 1 through 4 at any conventional significance level, so we have no evidence to contradict the validity of our VAR.

To determine if the growth rates of the US and Canada affect one another over time, we can perform Granger causality tests using our VAR.

```
. vargranger

Granger causality Wald tests
+-----+-----+
|          Equation          | Excluded |      chi2    df   Prob > chi2 |
+-----+-----+-----+
|          usgr              |    cgr   |  5.7613     1   0.016 |
|          usgr              |    ALL   |  5.7613     1   0.016 |
+-----+-----+-----+
|          cgr               |    usgr  | 26.768     1   0.000 |
|          cgr               |    ALL   | 26.768     1   0.000 |
+-----+-----+-----+
```

We see strong evidence that lagged Canadian growth helps predict US growth (the p -value is 0.016) and overwhelming evidence that lagged US growth helps predict Canadian growth (p -value less than 0.001). It is not surprising, given the relative sizes of the economies, that the US might have a stronger effect on Canada than vice versa. Note that because we have only one lag in our VAR, the Granger causality tests have only one degree of freedom and are

equivalent to the single-coefficient tests in the VAR regression tables. The coefficient of Canadian growth in the US equation has a z value of 2.40, which is the square root of the 5.7613 value reported in the `vargranger` table; both have identical p -values. If we had more than one lag in the regressions, then the z values in the VAR table would be tests of individual lag coefficients and the χ^2 values in the Granger causality table would be joint tests of the blocks of lag coefficients associated with each variable.

We next explore the implications of our VARs for the behavior of GDP in the two countries in 2012 and 2013. What does our model forecast? To find out, we issue that command: `fcast compute p_, step(8)`. This command produces no output, but changes our dataset in two ways. First, it extends the dataset by eight observations to 2013q4, filling in appropriate values for the date variable. Second, it adds eight variables to the dataset with values for those eight quarters. The new variables `p_usgr` and `p_cgr` contain the forecasts, and the variables `p_usgr_SE`, `p_usgr_LB`, and `p_usgr_UB` (and corresponding variables for Canada) contain the standard error, 95% lower bound, and 95% upper bound for the forecasts.

We can examine these forecast values in the data browser or with any statistical commands in the Stata arsenal, but it is often most informative to graph them. The command `fcast graph p_usgr p_cgr` generates the graph shown in Figure 5-1. The graph shows that the confidence bands on our forecasts are very large: our VARs do not forecast very confidently. The point forecasts predict little change in growth rates. The US growth rate is predicted to decline very slightly and then hold steady near its mean; the Canadian growth rate to increase a bit and then converge to its mean. If the goal of our VAR exercise was to obtain insightful and reliable forecasts, we have not succeeded!

Estimation, Granger causality, and forecasting can all be accomplished without any identification assumptions. But this is as far as we can go with our VAR analysis without making some assumptions to allow us to identify the structural shocks to US and Canadian GDP growth.

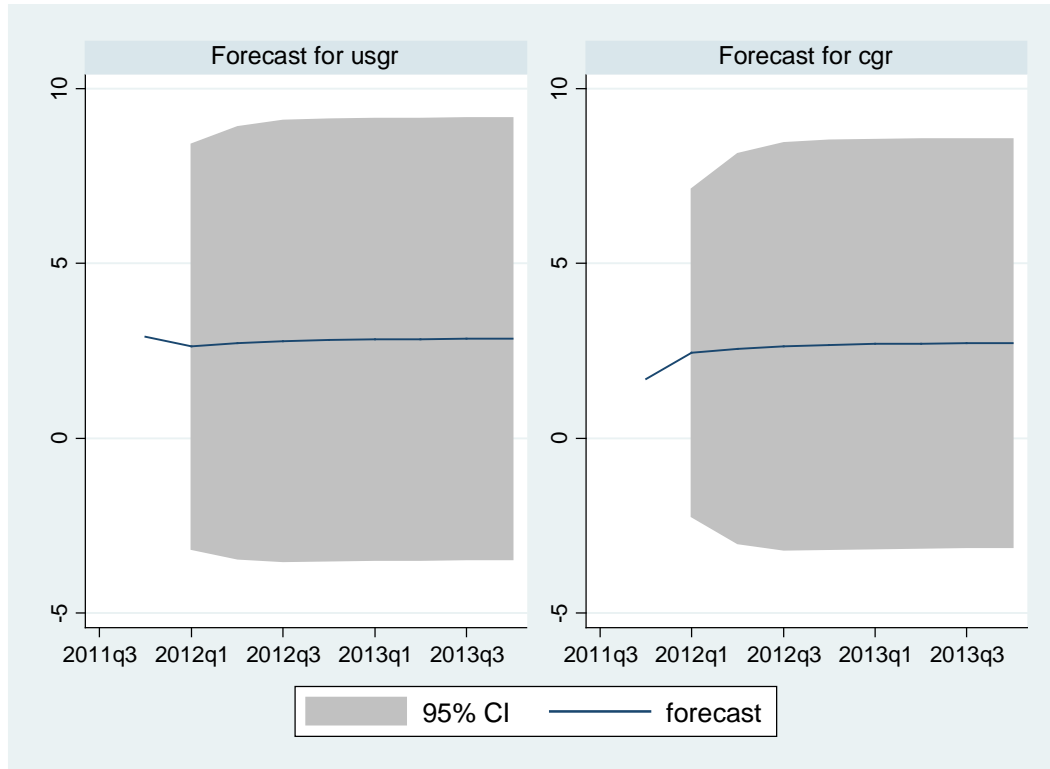


Figure 5-1. Graph of forecasts

5.4.3 Impulse-response functions and variance decompositions

Only if the residuals of the two equations are contemporaneously uncorrelated can we interpret them as structural shocks. We would expect that innovations to US and Canadian growth in any period would tend to be positively correlated, and indeed the cross-equation correlation coefficient for the residuals in our VAR regressions is 0.44.

In order to identify the effects of shocks to US or Canadian growth on the subsequent time paths of both, we must make an assumption about whether the correlation is due to current US growth affecting current Canadian growth or due to Canadian growth affecting US growth. Given the relative sizes of the two economies, theory suggests that US growth would have a stronger effect on Canada than vice versa, and this is supported somewhat by the evidence for lagged effects from our Granger causality tests (although both have strong effects on the other). Thus, we choose as our preferred identification pattern interpreting the contemporaneous correlation as the effect of US growth on Canadian growth: US growth is first in our ordering and Canadian growth is second.

We begin by creating a “.irf” file called “uscan.irf” to contain our impulse-response functions:

```
. irf set "uscan"
(file uscan.irf created)
(file uscan.irf now active)
```

We then create an IRF entry in the file called “L1” to hold the results of our one-lag VARs, running the IRF effect horizon out 20 quarters (five years):

```
. irf create L1 , step(20) order(usgr cgr)
(file uscan.irf updated)
```

We specified the ordering explicitly in creating the IRF. However, because the ordering is the same as the order in which the variables were listed in the `var` command itself, Stata would have chosen this ordering by default.

We can now use the `irf graph` command to produce impulse-response function and variance decomposition graphs. To get the (orthogonalized) IRFs, we type

```
. irf graph oirf , irf(L1) ustep(8)
```

It is important to specify `oirf` rather than `irf` because the latter gives impulse responses assuming (counterfactually) that the VAR residuals are uncorrelated. The resulting IRFs are shown in Figure 5-2.

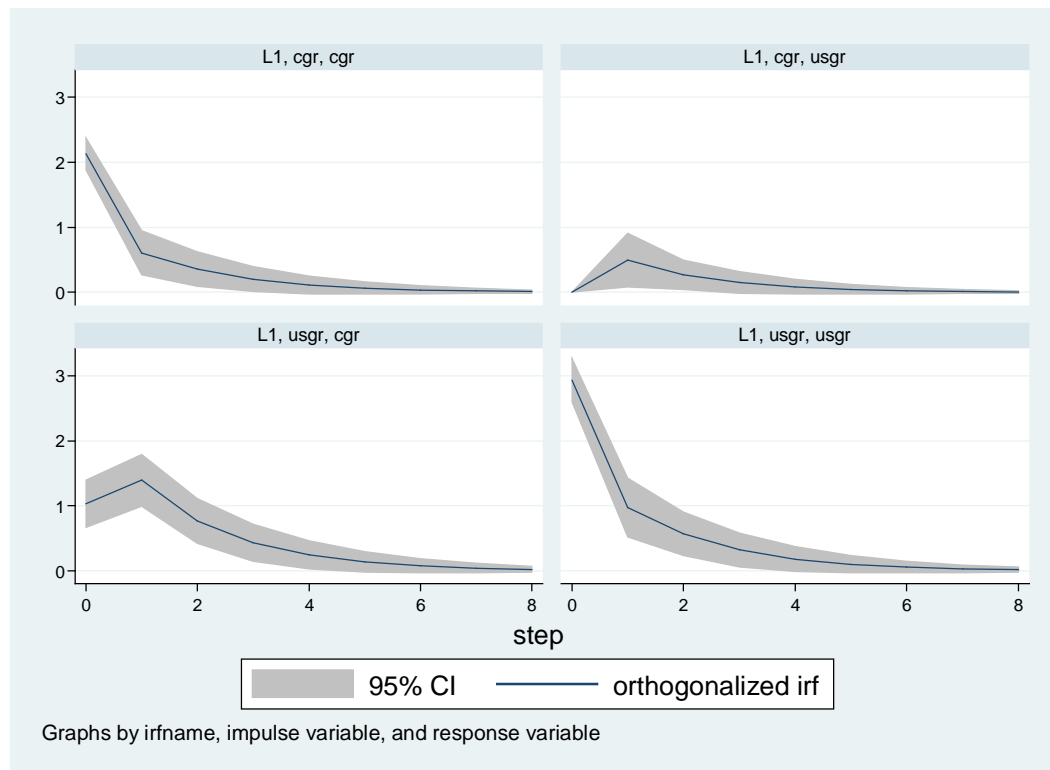


Figure 5-2. IRFs for preferred ordering

The diagonal panels in Figure 5-2 show the effects of shocks to each country's GDP growth on future values of its own growth. In both cases, the shock dies out quickly, reflecting the stationarity of the variables. A one-standard deviation shock to Canadian GDP growth in the top-left panel is just over 2 percent; a corresponding shock to U.S. growth is close to 3 percent.

The off-diagonal panels (bottom-left and top-right) show the effects of a growth shock in one country on the path of growth in the other. In the bottom-left panel, we see that a one-standard-deviation (about 3 percentage points) shock in U.S. growth raises Canadian growth by about 1 percentage point in the current quarter, then by a bit more in the next quarter as the lagged effect kicks in. From the second lag on, the effect decays rapidly to zero, with the statistical significance of the effect vanishing at about one year.

In the top-right panel, we see the estimated effects of a shock to Canadian growth on growth in the United States. The first thing to notice is that the effect is zero in the current period (at zero lag). This is a direct result of our identification assumption: we imposed the condition that Canadian growth has no immediate effect on U.S. growth in order to identify the shocks. The second noteworthy result is that the dynamic effect that occurs in the second period is much smaller than the effect of the U.S. on Canada. This is as we expected.

But how much of this greater dependence of Canada on the United States is really the data speaking and how much is our assumption that contemporaneous correlation in shocks runs only from the U.S. to Canada? Recall that our identification assumption imposes the condition that any "common shocks" that affect both countries are assumed to be U.S. shocks, with Canada shocks being the part of the Canadian VAR innovation that is *not* explained by the common shock. This might cause the Canadian shocks to have smaller variance (which it does in Figure 5-2) and might also overestimate the effect of the U.S. shocks on Canada.

To assess the sensitivity of our conclusions to the ordering assumption, we examine the IRFs making the opposite assumption: that contemporaneous correlation in the innovations is due to Canada shocks affecting the U.S. Figure 5-3 shows the graphs of the reverse-ordering IRFs. As expected, the effect of the U.S. on Canada (lower left) now begins at zero and the effect of Canada on the U.S. (upper right) does not.

Beyond this, there are a couple of interesting changes when we reverse the order. First, note that both shocks now have a standard deviation of about 2.5 rather than the U.S. shock having a much larger standard deviation. This occurs because we now attribute the "common" part of the innovation to the Canadian shock rather than the U.S. shock. Second, after the initial period in which the U.S.-to-Canada effect is constrained to be zero, the two effects are of similar magnitudes and die out in a similar way.

This example shows the difficulty of identifying impulse responses in VARs. The implications can depend on the identification assumption we make, so if we are not sure which assumption is better we may be left with considerable uncertainty in interpreting our results.

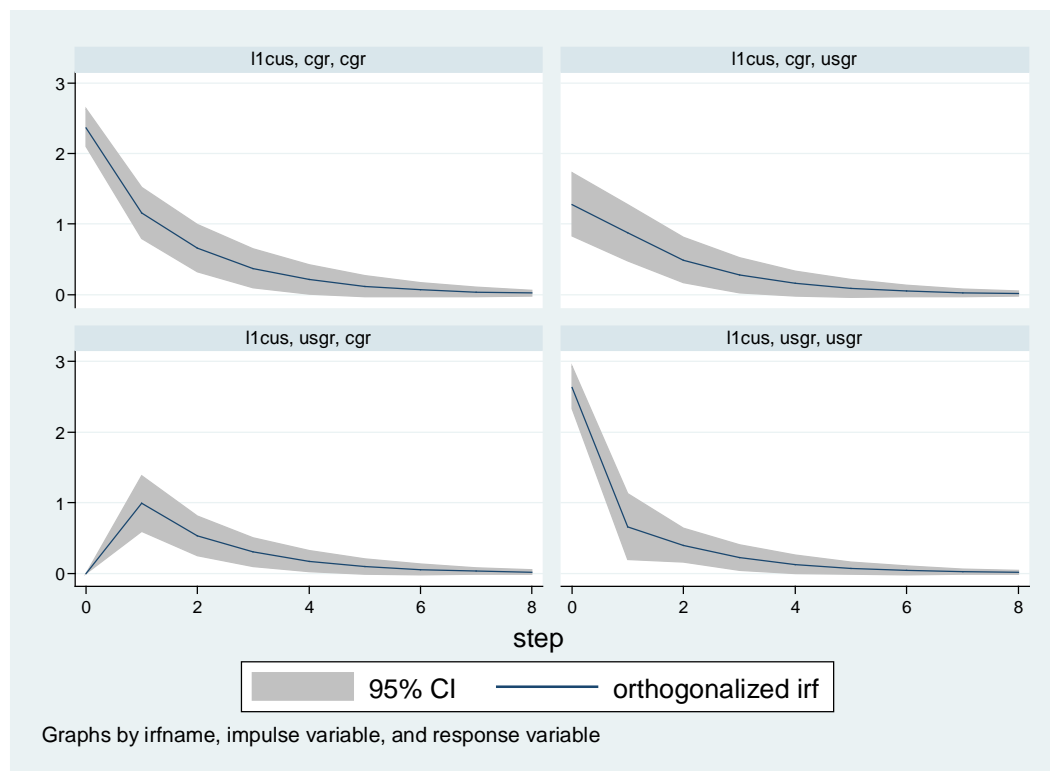


Figure 5-3. IRFs with reversed ordering

We can also use Stata's `irf graph` command to plot the cumulative effect of a permanent shock to one of the variables. For the preferred ordering this looks like Figure 5-4. Using the top-left panel, a permanent positive shock of one standard deviation to Canada's growth—an exogenous increase of about 2 percentage points of growth that is sustained over time—would eventually cause Canadian growth to be about 3.5 percentage points higher. This magnification comes from two effects. First, shocks to Canadian growth tend to persist for a period or two after the shock, so growth increases more as a result. Second, a positive shock to Canadian growth increases U.S. growth (even with no exogenous shock in the U.S.), which feeds back positively on Canadian growth. The same multiplier effect happens in the other panels.

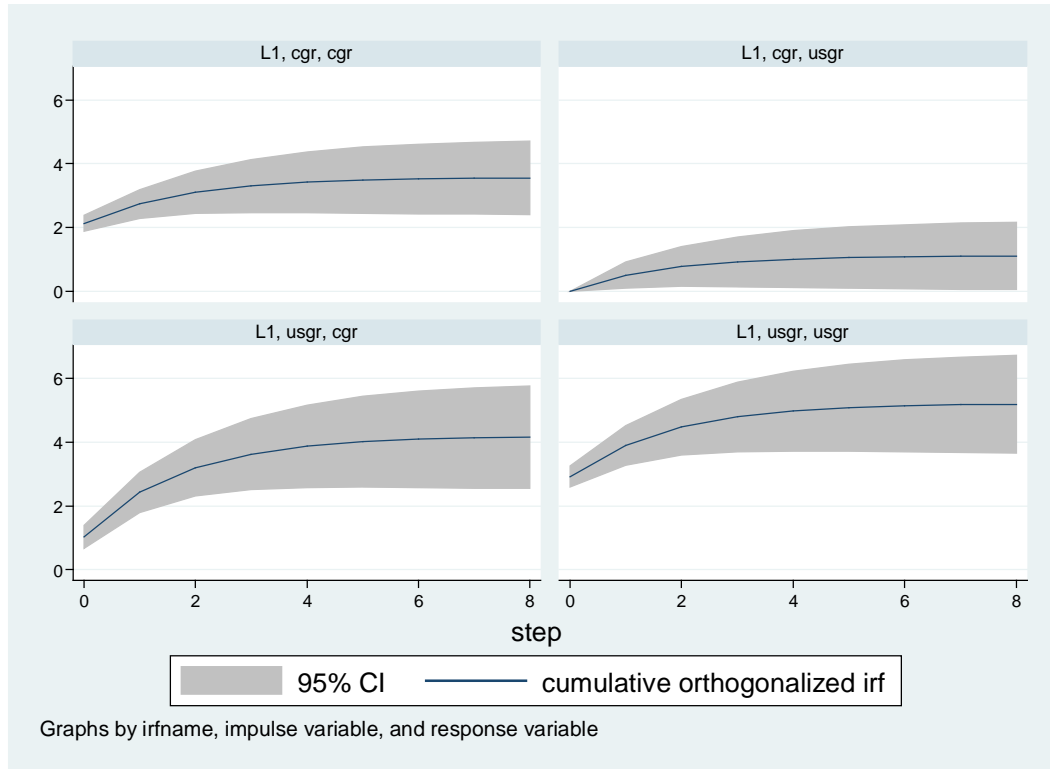


Figure 5-4. Cumulative IRFs with preferred ordering

Another tool that is available for analysis of identified VARs is the forecast-error variance decomposition, which measures the extent to which each shock contributes to unexplained movements (forecast errors) in each variable. Figure 5-5 results from the Stata command: `irf graph fevd , irf(L1) ustep(8)` and shows how each shock contributes to the variation in each variable. All variance decompositions start at zero because there is no forecast error at a zero lag.

The left-column panels show that (with the preferred identification assumption) the Canadian shock contributes about 80% of the variance in the one-period-ahead forecast error for Canadian growth, with the U.S. shock contributing the other 20%. As our forecast horizon moves further into the future, the effect of the U.S. shock on Canadian growth increases and the shares converge to less than 60% of variation in Canadian growth being due to the Canadian shock and more than 40% due to the U.S. shock. The right-column panels indicate that very little (less than 5%) of the variation in U.S. growth is attributable to Canadian growth shocks in the short run or long run.

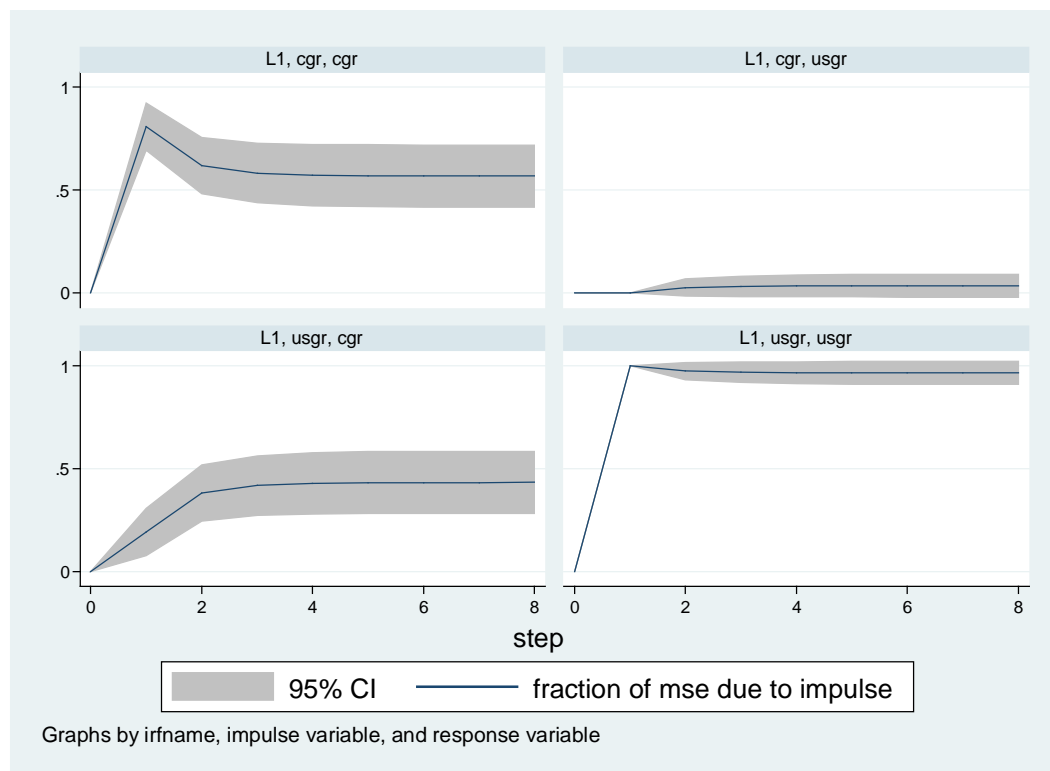


Figure 5-5. Variance decompositions with preferred ordering

Like IRFs, variance decompositions can be sensitive to the identification assumptions we make. If we compute IRFs and variance decompositions for alternative orderings and find that the results are similar, then we gain confidence that our conclusions are not sensitive to the (perhaps arbitrary) assumptions we make about contemporaneous causality. If alternative assumptions lead to different conclusions, we must be more careful about drawing conclusions.

Figure 5-6 shows the very different results that we get when we reverse the contemporaneous causal ordering. Now the Canadian shock (which includes the shock that is common to both countries under this assumption) explains most (80%) of the variation in Canadian growth and much (30%) of the variation in growth in the United States.

It may seem frustrating to reach quite different conclusions depending on a potentially arbitrary assumption about the direction of immediate causation. In this case, though, the differences between the results suggest some possible interpretations.

First, the United States has a stronger effect on Canada than vice versa. Interpreting the VAR results in favor of Canada's effect (by putting them first in the order) gives Canada a substantial effect on the U.S. but the U.S. shocks are clearly still important for both countries, but interpreting them in favor of the U.S. effect virtually wipes out the effect of Canada on the United States.

Second, because of the way the results vary between orderings, it is clear that much of the variation in growth in both countries is due to a common shock. Whichever country is (perhaps arbitrarily) assigned ownership of this shock seems to have a large effect relative with the other. While this doesn't resolve the "causality question," it is very useful information about the co-movement of U.S. and Canadian growth.

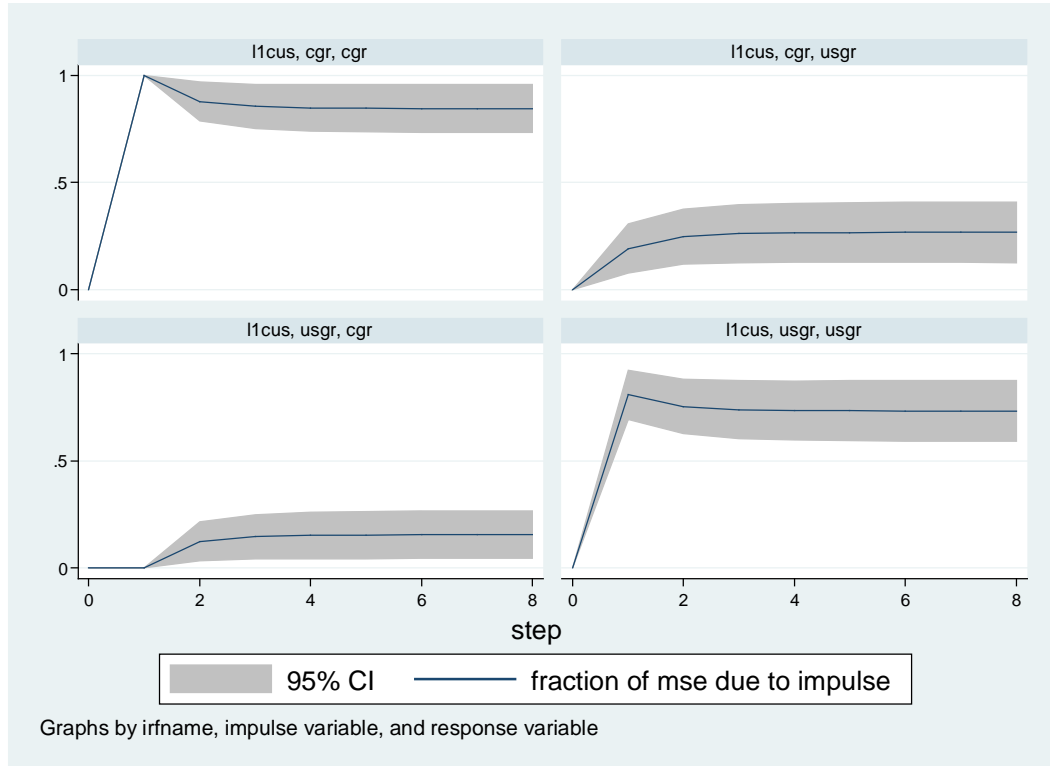


Figure 5-6. Variance decomposition with reversed ordering

5.5 Cointegration in a VAR: Vector Error-Correction Models

In our analysis of vector autoregressions, we have assumed that the variables of the model are stationary and ergodic. We saw in the previous chapter that variables that are individually non-stationary may be cointegrated: two (or more) variables may have common underlying stochastic trends along which they move together on a non-stationary path. For the simple case of two variables and one cointegrating relationship, we saw that an error-correction model is the appropriate econometric specification. In this model, the equation is differenced and an error-correction term measuring the previous period's deviation from long-run equilibrium is included.

We now consider how cointegrated variables can be used in a VAR using a *vector error-correction (VEC)* model. First we examine the two-variable case, which extends the simple

single-equation error-correction model to two equations in a straightforward way. We then generalize the model to more than two variables and equations, which allows for the possibility of more than one cointegrating relationship. This requires a new test for cointegration and a generalization of the error-correction model to include multiple error-correction terms.

5.5.1 A two-variable VEC model

If two $I(1)$ series x and y are cointegrated, then there exist unique α_0 and α_1 such that $u_t \equiv y_t - \alpha_0 - \alpha_1 x_t$ is $I(0)$. In the single-equation model of cointegration where we thought of y as the dependent variable and x as an exogenous regressor, we saw that the error-correction model

$$\Delta y_t = \beta_0 + \beta_1 \Delta x_t + \lambda u_{t-1} + \varepsilon_t = \beta_0 + \beta_1 \Delta x_t + \lambda (y_{t-1} - \alpha_0 - \alpha_1 x_{t-1}) + \varepsilon_t \quad (5.18)$$

was an appropriate specification. All terms in equation (5.18) are $I(0)$ as long as the α coefficients (the “cointegrating vector”) are known or at least consistently estimated. The u_{t-1} term is the magnitude by which y was above or below its long-run equilibrium value in the previous period. The coefficient λ (which we expect to be negative) represents the amount of “correction” of this period- $(t-1)$ disequilibrium that happens in period t . For example, if λ is -0.25 , then one quarter of the gap between y_{t-1} and its equilibrium value would tend (all else equal) to be reversed (because the sign is negative) in period t .

The VEC model extends this single-equation error-correction model to allow y and x to evolve jointly over time as in a VAR system. In the two-variable case, there can be only one cointegrating relationship and the y equation of the VEC system is similar to (5.18), except that we mirror the VAR specification by putting lagged differences of y and x on the right-hand side. With only one lagged difference (there can be more) the bivariate VEC can be written

$$\begin{aligned} \Delta y_t &= \beta_{y0} + \beta_{yy1} \Delta y_{t-1} + \beta_{yx1} \Delta x_{t-1} + \lambda_y (y_{t-1} - \alpha_0 - \alpha_1 x_{t-1}) + v_t^y, \\ \Delta x_t &= \beta_{x0} + \beta_{xy1} \Delta y_{t-1} + \beta_{xx1} \Delta x_{t-1} + \lambda_x (y_{t-1} - \alpha_0 - \alpha_1 x_{t-1}) + v_t^x. \end{aligned} \quad (5.19)$$

As in (5.18), all of the terms in both equations of (5.19) are $I(0)$ if the variables are cointegrated with cointegrating vector $(1, -\alpha_0, -\alpha_1)$, in other words, if $y_t - \alpha_0 - \alpha_1 x_t$ is stationary. The λ coefficients are again the error-correction coefficients, measuring the response of each variable to the degree of deviation from long-run equilibrium in the previous period. We expect $\lambda_y < 0$ for the same reason as above: if y_{t-1} is above its long-run value in relation to x_{t-1} then the error-correction term in parentheses is positive and this should lead, other things constant, to downward movement in y in period t . The expected sign of λ_x depends on the sign of α_1 . We expect $\partial \Delta x_t / \partial x_{t-1} = -\lambda_x \alpha_1 < 0$ for the same reason that we expect $\partial \Delta y_t / \partial y_{t-1} = \lambda_y < 0$: if x_{t-1} is above its long-run relation to y , then we expect Δx_t to be negative, other things constant.

A simple, concrete example may help clarify the role of the error-correction terms in a VEC model. Let the long-run cointegrating relationship be $y_t = x_t$, so that $\alpha_0 = 0$ and $\alpha_1 = -1$. The parenthetical error-correction term in each equation of (5.19) is now $y_{t-1} - x_{t-1}$, the difference between y and x in the previous period. Suppose that because of previous shocks, $y_{t-1} = x_{t-1} + 1$ so that y is above its long-run equilibrium relationship to x by one unit (or, equivalently, x is below its long-run equilibrium relationship to y by one unit). To move toward long-run equilibrium in period t , we expect (if there are no other changes) $\Delta y_t < 0$ and $\Delta x_t > 0$. Using equation (5.19), Δy_t changes in response to this disequilibrium by $\lambda_y (y_{t-1} - x_{t-1}) = \lambda_y$, so for stable adjustment to occur $\lambda_y < 0$; y is “too high” so it must decrease in response to the disequilibrium. The corresponding change in Δx_t from equation (5.19) is $\lambda_x (y_{t-1} - x_{t-1}) = \lambda_x$. Since x is “too low,” stable adjustment requires that the response in x be *positive*, so we need $\lambda_x > 0$. Note that if the long-run relationship between y and x were inverse ($\alpha_1 < 0$), then x would need to *decrease* in order to move toward equilibrium and we would need $\lambda_x < 0$. The expected sign on λ_x depends on the sign of α_1 .

If theory tells us the coefficients α_0 and α_1 of the cointegrating relationship, as in the case of purchasing-power parity, then we can calculate the error-correction term in (5.19) and estimate it as a standard VAR. However, we usually do not know these coefficients, so they must be estimated.

Single-equation cointegrated models can be estimated either directly or in two steps. We can use OLS to estimate the cointegrating relationship—the cointegrating vector $(1, \alpha_0, \alpha_1)$ —and impose these estimates on the error-correction model, or we can estimate the α coefficients jointly with the β coefficients on the differences. In a VAR/VEC system, separate estimation of the cointegrating relationship is likely to be problematic because it may be implausible to assume that either x or y is even weakly exogenous. (However, note that the identifying restrictions we impose to calculate IRFs require exactly this assumption.) Thus, it is common to estimate both sets of coefficients simultaneously in a VEC model. We shall see an example of VEC estimation and interpretation shortly, but first we consider models with more than two variables and equations because these models raise important additional considerations.

5.5.2 A three-variable VEC with (partially) known cointegrating relationships

We now consider a vector error-correction model with three variables x , y , and z . This situation is more complex because the number of linear combinations of the three variables that are stationary could be 0, 1, or 2. In other words, there could be zero, one, or two common trends among the three variables.

If there are no cointegrating relationships, then the series are not cointegrated and a VAR in differences is the appropriate specification. There is no long-run relationship to which the levels of the variables tend to return, so there is no basis for an error-correction term in any equation.

There would be one cointegrating relationship among the three variables if there is one long-run equilibrium condition tying the levels of the variables together. An example would be the purchasing-power-parity (PPP) condition between two countries under floating exchange rates. Suppose that P^1 is the price of a basket of goods in country one, P^2 is the price of the same basket in country 2, and X is the exchange rate: the number of units of country one's currency that buys one unit of country two's. If goods are to cost the same in both countries—purchasing-power parity—then $X = P^1 / P^2$. Any increase in prices in country one should be reflected, in long-run equilibrium, by an increase of equal proportion in the amount of country-one currency needed to buy a unit of country-two's currency.

In practice, economists have to rely on price indexes whose market baskets differ across countries, so the PPP equation would need a constant of proportionality to reflect this difference: $X = A_0 P^1 / P^2$. Denoting logs of the variables by small letters, this implies that

$$x = \alpha_0 + p^1 - p^2$$

is a long-run equilibrium condition toward which the variables should tend.

We could estimate a VEC system (with one lag, for simplicity) for the evolution of the three variables x , p^1 , and p^2 with one cointegrating relationship (with some known coefficients):

$$\begin{aligned}\Delta x_t &= \beta_{x0} + \beta_{xx1} \Delta x_{t-1} + \beta_{x11} \Delta p_{t-1}^1 + \beta_{x21} \Delta p_{t-1}^2 + \lambda_x (x_{t-1} - \alpha_0 - p_{t-1}^1 + p_{t-1}^2) + v_t^x \\ \Delta p_t^1 &= \beta_{10} + \beta_{1x1} \Delta x_{t-1} + \beta_{111} \Delta p_{t-1}^1 + \beta_{121} \Delta p_{t-1}^2 + \lambda_1 (x_{t-1} - \alpha_0 - p_{t-1}^1 + p_{t-1}^2) + v_t^1 \\ \Delta p_t^2 &= \beta_{20} + \beta_{2x1} \Delta x_{t-1} + \beta_{211} \Delta p_{t-1}^1 + \beta_{221} \Delta p_{t-1}^2 + \lambda_2 (x_{t-1} - \alpha_0 - p_{t-1}^1 + p_{t-1}^2) + v_t^2.\end{aligned}\tag{5.20}$$

If the exchange rate is out of equilibrium, say, too high, then we expect some combination of adjustments in x , p^1 , and p^2 to move back toward long-run equilibrium. The error-correction coefficients λ_x , λ_1 , and λ_2 measure these responses. Using the logic described above, we would expect λ_x and λ_2 to be negative and λ_1 to be positive.

This does not exhaust the possible degree of cointegration among these variables, however. Suppose that country one is on a gold standard, so that the price level in that country tends to be constant in the long run.² This would impose a second long-run equilibrium condition—a second cointegrating relationship—on the variables: $p^1 = \alpha_1$. The VEC system incorporating both cointegrating relationships would look like

² Another example we could use would be a fixed-exchange-rate system in which one country keeps x near a constant level in the long run.

$$\begin{aligned}
\Delta x_t &= \beta_{x0} + \beta_{xx1} \Delta x_{t-1} + \beta_{x11} \Delta p_{t-1}^1 + \beta_{x21} \Delta p_{t-1}^2 + \lambda_x (x_{t-1} - \alpha_0 - p_{t-1}^1 + p_{t-1}^2) + \mu_x (p_{t-1}^1 - \alpha_1) + v_t^x \\
\Delta p_t^1 &= \beta_{10} + \beta_{1x1} \Delta x_{t-1} + \beta_{111} \Delta p_{t-1}^1 + \beta_{121} \Delta p_{t-1}^2 + \lambda_1 (x_{t-1} - \alpha_0 - p_{t-1}^1 + p_{t-1}^2) + \mu_1 (p_{t-1}^1 - \alpha_1) + v_t^1 \\
\Delta p_t^2 &= \beta_{20} + \beta_{2x1} \Delta x_{t-1} + \beta_{211} \Delta p_{t-1}^1 + \beta_{221} \Delta p_{t-1}^2 + \lambda_2 (x_{t-1} - \alpha_0 - p_{t-1}^1 + p_{t-1}^2) + \mu_2 (p_{t-1}^1 - \alpha_1) + v_t^2.
\end{aligned} \tag{5.21}$$

In equations (5.20) and (5.21) we started with a system in which we “knew” the nature of the cointegrating relationship(s) among the variables. It is more common that we must test for the possibility of cointegration (and determine how many cointegrating relationships exist) and estimate a full set of α parameters for them. We now turn to that process, then to an example of an estimated VEC model.

5.5.3 Testing for and estimating cointegrating relationships in a VEC

The most common tests to determine the number of cointegrating relationships among the series in a VAR/VEC are due to Johansen (1995). Although the mathematics of the tests involve methods that are beyond our reach, the intuition is very similar to testing for unit roots in the polynomial representing an AR process.

If we have $n I(1)$ variables that are modeled jointly in a dynamic system, there can be up to $n - 1$ cointegrating relationships linking them. Stock and Watson (ref) think of each cointegrating relationship as a *common trend* linking some or all of the series in the system. We shall think of “cointegrating relationship” and “common trend” as synonymous. The *cointegrating rank* of the system is the number of such common trends, or the number of cointegrating relationships.³

To determine the cointegrating rank r , we perform a sequence of tests. First we test the null hypothesis of $r = 0$ against $r \geq 1$ to determine if there is at least one cointegrating relationship. If we fail to reject $r = 0$, then we conclude that there are no cointegrating relationships or common trends among the series. In this case, we do not need a VEC model and can simply use a VAR in the differences of the series.

If we reject $r = 0$ at the initial stage then at least some of the series are cointegrated and we want to determine the number of cointegrating relationships. We proceed to a second step to test the null hypothesis that $r \leq 1$ against $r \geq 2$. If we cannot reject the hypothesis of no more than one common trend, then we estimate a VEC system with one cointegrating relationship, such as (5.20).

If we reject the hypothesis that $r \leq 1$, then we proceed further to test $r \leq 2$ against $r \geq 3$, and so on. We choose r to be the smallest value at which we fail to reject the null hypothesis that there are no additional cointegrating relationships.

³ For those familiar with linear algebra, the term “rank” refers to the rank of a matrix characterizing the dynamic system. If a dynamic system of n variables has r cointegrating relationships, then the rank of the matrix is $n - r$. This means that the matrix has r eigenvalues that are zero and $n - r$ that are not. The Johansen tests are based on determining the number of nonzero eigenvalues.

Johansen proposed several related tests that can be used at each stage. The most common (and the default in Stata) is the *trace statistic*. The Stata command `vecrank` prints the trace statistic or, alternatively, the maximum-eigenvalue statistic (with the `max` option) or various information criteria (with the `ic` option).

The Johansen procedure invoked in Stata by the `vec` command estimates both the parameters of the adjustment process (the β coefficients on the lagged changes in all variables) and the long-run cointegrating relationships themselves (the α coefficients on the long-run relationships) by maximum likelihood. We must tell Stata whether to include constant terms in the differenced VEC regressions—remember that a constant term in a differenced equation corresponds to a trend term in the levels—or perhaps trend terms (which would be a quadratic trend in the levels). It is also possible to include seasonal variables where appropriate or to impose constraints on the coefficients of either the cointegrating relationships or the adjustment equations.

Once the VEC system has been estimated, we can proceed to calculate IRFs and variance decompositions, or to generate forecasts as we would with a VAR.

References

- Johansen, Soren. 1995. *Likelihood-Based Inference in Cointegrated Vector Autoregressive Models*. Oxford: Clarendon Press.
- Sims, Christopher A. 1980. Macroeconomics and Reality. *Econometrica* 48 (1):1-48.