EC327: Financial Econometrics, Spring 2013

Wooldridge, Introductory Econometrics (5th ed, 2012)

Chapter 18:
Advanced time series topics

Infinite distributed lag models

Consider a pair of timeseries $[y_t, z_t]$. An *infinite distributed lag* model (IDL) relating y_t to all observed values of z is

$$y_t = \alpha + \delta_0 z_t + \delta_1 z_{t-1} + \delta_2 z_{t-2} + \ldots + u_t$$

where the sum leads back to the infinite past. Unlike a finite DL model, the IDL does not require us to choose a truncation point. In order for this model to make sense, the lag coefficients δ_j must approach zero as $j \to \infty$. They need not do so monotonically, but the

influence of z in the distant past on current y should be negligible.

How can we estimate such a model from a finite sample of data? We must place restrictions on the δ_j parameters that allow us to express the model more parsimoniously.

The dynamic multipliers or impact effects of changes in the z sequence are naturally just the δ coefficients themselves. The long-run effect of a temporary change in z will be zero as $j \to \infty$. The long-run effect of a permanent change: for instance, z increasing by one unit at period t and maintaining that higher level will be the sum of all δ_j coefficients. As long as the coefficients decline toward zero as $j \to \infty$, this infinite sum will be well behaved, and can often be well approximated by a sum of the first p terms.

In order to estimate an IDL with least squares methods, we must make an assumption on the correlation between z and u. Strict exogeneity—where the conditional mean of u is zero, conditioned on past, present and future z—is usually not a workable assumption since it rules out correlations between current u and future z. Such correlations would be expected to appear in a model where a policy response to a sizable shock was expected (e.g., government spending after Hurricane Katrina). A more workable assumption is that u is uncorrelated with current and past z, or $E(u_t|z_t, z_{t-1}, \ldots) = 0$. This assumption does not place restrictions on the serial correlation properties of u (just as in the case of finite DL models).

The Koyck (geometric) distributed lag model

The simplest abstraction from the infeasible IDL model is the Koyck or *geometric* distributed

lag (GDL) model, in which the sequence of δ_j parameters depends on only two parameters:

$$\delta_j = \gamma \rho^j, \ \ |\rho| < 1$$

The parameters γ and ρ may be positive or negative. Convergence of the δ_j to zero occurs at a fast rate, since they are powers of a fraction. The impact multiplier is just $\delta_0 = \gamma$. If $\gamma > 0$ and $\rho > 0$, all δ_j parameters are positive, and decline monotonically with j. If $\rho < 0$ then parameters alternate in sign. The long run propensity can be computed as the sum of the infinite geometric series $1 + \rho + \rho^2 + \ldots = 1/(1 - \rho)$ times γ, and will have the same sign as γ. This definition of the δ_j sequence still leaves us with an infinite sum, but if we write the equation at time t and $t - 1$ we have:

$$y_t = \alpha + \gamma z_t + \gamma \rho z_{t-1} + \gamma \rho^2 z_{t-2} + \ldots + u_t$$

and

$$y_{t-1} = \alpha + \gamma z_{t-1} + \gamma \rho z_{t-2} + \gamma \rho^2 z_{t-3} + \ldots + u_{t-1}$$

If we multiply the second equation by ρ and subtract it from the first, most of the terms cancel:

$$y_t - \rho y_{t-1} = (1 - \rho)\alpha + \gamma z_t + (u_t - \rho u_{t-1})$$

or

$$y_t = \alpha_0 + \gamma z_t + \rho y_{t-1} + (u_t - \rho u_{t-1})$$

where $\alpha_0 = (1 - \rho)\alpha$.

Although this model is simple enough, it is not straightforward to estimate since it has a lagged dependent variable and an error term that is by construction correlated with the LDV (the error term is actually a moving average process of order one, or MA(1)). We can estimate the Koyck lag model with instrumental variables; if the assumption on the conditional mean of u holds, we can use z_{t-1} as an appropriate instrument for y_{t-1}. It is likely that we will want to calculate HAC standard errors for

the equation to account for serial correlation in the composite error term, since the moving average process will not have independent elements.

Alternatively, if we assume that the u process is AR(1) with the same value of ρ,

$$u_t = \rho u_{t-1} + \epsilon_t$$

and the ϵ_t are assumed to be independent of $(z_t, y_{t-1}, z_{t-1}, \ldots)$, then we can write the equation to be estimated as

$$y_t = \alpha_0 + \gamma z_t + \rho y_{t-1} + \epsilon_t$$

which may be consistently estimated by OLS, with no problem of serial correlation in the errors. However, it does depend on the assumption that the ρ in the AR(1) error process is the same ρ governing decay in the δ_j coefficients.

The Koyck model may be extended to multiple explanatory variables, but relies on the

assumption that the same decay parameter ρ applies to each of them.

Rational distributed lag models

Although the Koyck lag model is useful, it is restrictive in that it assumes that the weights monotonically decline (in absolute value if $\rho < 0$). The GDL is a special case of a more general model of a *rational distributed lag* (RDL). A simple RDL can be constructed by adding a lagged z term to our GDL example:

$$y_t = \alpha_0 + \gamma_0 z_t + \rho y_{t-1} + \gamma_1 z_{t-1} + v_t$$

where v_t is the MA(1) process $u_t - \rho u_{t-1}$. By repeated back-substitution, we can show that this representation is equation to the infinite DL model

$$y_t = \alpha + \gamma_0 z_t + (\rho \gamma_0 + \gamma_1) z_{t-1} + \rho(\rho \gamma_0 + \gamma_1) z_{t-2}$$
$$+ \rho^2(\rho \gamma_0 + \gamma_1) z_{t-3} + \ldots + u_t$$

The impact multiplier is γ_0, while the $j-$period delay multiplier (coefficient on z_{t-j}) is

$$\rho^{j-1}(\rho\gamma_0 + \gamma_1), \; j \geq 1.$$

Thus, the impact propensity can differ in sign from the other lag coefficients (although for $\rho > 0$ the sequence of coefficients on following terms is monotonic with the same sign as $(\rho\gamma_0 + \gamma_1)$. The long-run multiplier for this variant of the RDL model is equal to $(\gamma_0 + \gamma_1)/(1 - \rho)$, with the same sign as $(\gamma_0 + \gamma_1)$, and nonzero as long as $(\gamma_0 + \gamma_1) \neq 0$.

Unit root tests

Given the distinction between trend-stationary and unit root processes, it would seem to be very important to be able to determine whether a particular timeseries which, for instance, generally increases in value is being driven by some underlying trend, or whether its evolution reflects a unit root in its data generating process.

Those who study macroeconomic phenomena will want to know whether economic recessions have permanent consequences for the level of future GDP (as they would if GDP exhibits a unit root), or whether they are merely deviations from a trend rate of growth, temporary downturns that will be offset by the following recovery. Those who are concerned with the stock market want to know whether stock prices really do follow a random walk—i.e. exhibit unit root behavior—rather than some complicated combination of trends and cycles. If stock prices' behavior reflect a unit root, then "technical analysis" or "charting" is no more useful than astrology. On the other hand, if there are no unit roots in stock prices, all of the effort applied by stock analysts to studying the behavior of these series should have a reward.

This concern has given rise to a battery of *unit root* tests: statistical procedures that are designed to render a verdict as to whether a given

sample of timeseries data appears to imply the presence of a unit root in that timeseries, or whether the series may be considered stationary. In terms of our prior terminology, we are trying to discern whether the series exhibits $I(1)$ (unit root) or $I(0)$ (stationary) behavior. It turns out that this is a fairly difficult problem, from a statistical perspective. It might appear sufficient to merely estimate an equation such as $y_t = \phi_1 y_{t-1} + \epsilon_t$, modified to the form

$$\Delta y_t = \gamma y_{t-1} + \epsilon_t \qquad (1)$$

using the available sample of data, and test the null hypothesis that $\gamma = (\phi_1 - 1) = 1$. For various reasons, that turns out to be problematic, in the sense that the distribution of the test statistic is nonstandard under that null. The t test for $\gamma = 0$ does not have a t distribution under the null hypothesis; even as $T \to \infty$, the distribution of this t statistic will not be $N(0, 1)$. Under the alternative

hypothesis, the test statistic is well behaved, but under the null—the point of interest—it follows the "Dickey–Fuller" distribution rather than the Normal or t. The critical points on the D–F distribution, as established by simulation, are considerably larger than those of the equivalent t; whereas a value of -1.645 would be on the borderline of rejection at the 95% level for a one–tailed t test, the D-F critical value would be -1.961 for $T = 1000$.

Of course, the model (1) may not be the appropriate special case of the autoregressive distributed lag model; we may want to allow for an additional term which would become a constant term in a stable autoregressive process, or a drift term in a random walk process. Otherwise, we are specifying a stable autoregressive process with a zero mean under the alternative hypothesis, which may not be sensible

(unless the series has already been demeaned). With that modification, we would test

$$\triangle y_t = \mu + \gamma y_{t-1} + \epsilon_t \qquad (2)$$

which would then allow both a test for a unit root ($\gamma = 0$) and a joint test for a white noise process (an F test for $\gamma = 0$ and $\mu = 0$). Note that the critical values for the t test are not the same as those that would be used in (1); for instance, the D–F critical value for $T = 1000$ in this test is −2.864. One must also note that this model would not be appropriate if there was an obvious trend in the series, since the model under the alternative has no mechanism to generate such a trend (as the RW–with–drift model does under the null).

The most general form of the standard D–F test allows for both a constant in the relationship and a deterministic trend:

$$\triangle y_t = \mu + \gamma y_{t-1} + \beta t + \epsilon_t \qquad (3)$$

Such a model will allow for both a nonzero mean for y (with $\mu \neq 0$) and trending behavior (with $\beta \neq 0$) under the alternative hypothesis, where $\gamma < 0$. The most likely null hypothesis is then that of a RW-with-drift, so that under H_0 $\gamma = 0$ and $\beta = 0$ (no deterministic trend). This null could be rejected for three reasons: (a) there could be no unit root, but a deterministic trend; (b) there could be a unit root, but with a deterministic trend; or (c) there might be no unit root nor deterministic trend. The most general alternative is (a), for which an F test is required (since two restrictions on the parameter vector are implied under the null). The F statistic is calculated in the normal way, but the distribution is again nonstandard, and tabulated values for the "D–F F distribution" must be consulted. More commonly, we consider a t test on γ; once again, the critical values are specific to model (3). For instance, the D–F critical value for $T = 1000$ in this test

is −3.408: larger yet than the critical values in the constant–only model, which in turn exceed those for the original white noise model.

Any of the forms of this test presume the existence of white noise errors in the regression. If that is implausible, the test will lose significant power. To cope with this issue, any of the 'Dickey–Fuller" tests in practice are usually employed as the "augmented Dickey–Fuller" test, or ADF test, in which a number of lags of the dependent variable are added to the regression to whiten the errors:

$$\Delta y_t = \mu + \gamma y_{t-1} + \vartheta_1 \Delta y_{t-1} + \vartheta_2 \Delta y_{t-2} + \ldots + \beta t + \epsilon_t \tag{4}$$

In this formulation, we consider an $AR(p)$ model as the baseline model, rather than the $AR(1)$ model of the simple Dickey–Fuller framework. The choice of appropriate lag length is likely to depend on the frequency of the data; a

general-to-specific strategy (analogous to the Ng–Perron sequential t procedure) or an information criterion may also be used. We will discuss the use of a modified AIC below.

Phillips–Perron tests

The augmentation of the original D–F regression with lags of the dependent variable is motivated by the need to generate iid errors in that model, since an OLS estimator of the covariance matrix is being employed. An alternative strategy for allowing for errors that are not iid is that of Phillips (1987) and Phillips and Perron (1988), known as the Phillips–Perron (PP) unit root test. The PP test deals with potential serial correlation in the errors by employing a correction factor that estimates the long-run variance of the error process with a variant of the Newey-West formula. Like the

ADF test, use of the PP test requires specification of a lag order; in the latter case, the lag order designates the number of lags to be included in the long–run variance estimate. The PP test allows for dependence among disturbances of either AR or MA form, but have been shown to exhibit serious size distortions in the presence of negative autocorrelations. In principle, the PP tests should be more powerful than the ADF alternative. The same critical values are used for the ADF and PP tests.

The DF-GLS test

Conventional unit root tests are known to lose power dramatically against stationary alternatives with a low order MA process: a characterization that fits well to a number of macroeconomic time series. Consequently, these original tests have been largely supplanted in many researchers' toolkits by improved alternatives.

Along the lines of the ADF test, a more powerful variant is the DFGLS test proposed by Elliott, Rothenberg and Stock (ERS, 1996), originally authored in Stata by Baum and Sperling and available in official Stata as `dfgls`.`dfgls` performs the ERS efficient test for an autoregressive unit root. This test is similar to an (augmented) Dickey-Fuller t test, as performed by `dfuller`, but has the best overall performance in terms of small-sample size and power, dominating the ordinary Dickey-Fuller test. The `dfgls` test "has substantially improved power when an unknown mean or trend is present" (ERS, p.813).

`dfgls` applies a generalized least squares (GLS) detrending (demeaning) step to the `varname`:

$$y_t^d = y_t - \hat{\beta}' z_t$$

For detrending, $z_t = (1, t)'$ and $\hat{\beta}_0, \hat{\beta}_1$ are calculated by regressing $[y_1, (1 - \bar{\alpha} L) y_2, ..., (1 - \bar{\alpha} L) y_T]$

onto $[z_1, (1 - \bar{\alpha}L)\,z_2, ..., (1 - \bar{\alpha}L)\,z_T]$ where $\bar{\alpha} = 1 + \bar{c}/T$ with $\bar{c} = -13.5$, and L is the lag operator. For demeaning, $z_t = (1)'$ and the same regression is run with $\bar{c} = -7.0$. The values of $\bar{c}$ are chosen so that "the test achieves the power envelope against stationary alternatives (is asymptotically MPI *(most powerful invariant))* at 50 percent power" (Stock, 1994, p.2769; emphasis added). The augmented Dickey-Fuller regression is then computed using the y_t^d series:

$$\Delta y_t^d = \alpha + \gamma t + \rho y_{t-1}^d + \sum_{i=1}^{m} \delta_i \Delta y_{t-i}^d + \epsilon_t$$

where m=`maxlag`. The `notrend` option suppresses the time trend in this regression.

Approximate critical values for the GLS detrended test are taken from ERS, Table 1 (p.825). Approximate critical values for the GLS demeaned test are identical to those applicable to

the no–constant, no–trend Dickey–Fuller test, and are computed using the `dfuller` code.

The `dfgls` routine includes a very powerful lag selection criterion, the "modified AIC" (MAIC) criterion proposed by Ng and Perron (2001). They have established that use of this MAIC criterion may provide "huge size improvements" (2001, abstract) in the `dfgls` test. The criterion, indicating the appropriate lag order, is printed on `dfgls`' output, and may be used to select the test statistic from which inference is to be drawn.

It should be noted that all of the lag length criteria employed by `dfgls` (the sequential t test of Ng and Perron (1995), the SC, and the MAIC) are calculated, for various lags, by holding the sample size fixed at that defined for the longest lag. These criteria cannot be meaningfully compared over lag lengths if the

underlying sample is altered to use all available observations. That said, if the optimal lag length (by whatever criterion) is found to be much less than that picked by the Schwert criterion, it would be advisable to rerun the test with the `maxlag` option specifying that optimal lag length, especially when using samples of modest size.

The KPSS test

An alternative test is that proposed by Kwiatkowski et al. (1992), the so-called KPSS test, which has a null hypothesis of stationarity (that is, $H_0 : y \sim I(0)$). It is also described in Baum (2000) and implemented in Stata. This command is not built in to Stata, but can be readily installed in any version of Stata with access to the Web by using the "ssc install kpss" command. `kpss` performs the Kwiatkowski, Phillips, Schmidt, Shin (KPSS, 1992) test for

stationarity of a time series. The test may be conducted under the null of either trend stationarity (the default) or level stationarity. Inference from this test is complementary to that derived from those based on the Dickey–Fuller distribution (such as `dfgls`, `dfuller` and `pperron`). The KPSS test is often used in conjunction with those tests to investigate the possibility that a series is fractionally integrated (that is, neither *I(1)* nor *I(0)*): see Lee and Schmidt (1996).

The series is detrended (demeaned) by regressing y on $z_t = (1, t)' \left(z_t = (1)' \right)$, yielding residuals e_t. Let the partial sum series of e_t be s_t. Then the zero-order KPSS statistic $k_0 = T^{-2} \sum_{t=1}^{T} s_t^2 \; / \; T^{-1} \sum_{t=1}^{T} e_t^2$. For `maxlag` > 0, the denominator is computed as the Newey-West estimate of the long run variance of the series; see [R] `newey`.

Approximate critical values for the KPSS test are taken from KPSS (1992). The `kpss` routine includes two options recommended by the work of Hobijn et al. (1998). An automatic bandwidth selection routine has been added, rendering it unnecessary to evaluate a range of test statistics for various lags. An option to weight the empirical autocovariance function by the Quadratic Spectral kernel, rather than the Bartlett kernel employed by KPSS, has also been introduced. These options may be used separately or in conjunction.

Combining inference from I(1) and I(0) tests

The two families of unit root tests may be used in conjunction to establish the nature of the data generating process for a given time-series, and in particular to signal the presence

of fractional integration in the series. If inference from the DFGLS test rejects its null hypothesis of unit root behavior, or nonstationarity, while the KPSS test also rejects its null, then we might conclude that both $I(1)$ and $I(0)$ are rejected by the data. That sets the stage for an alternative explanation of the timeseries' behavior: that of fractional integration, or long-range dependence, in which the series may be characterized as $I(d)$, $0 < d < 1$, neither $I(0)$ nor $I(1)$.

Spurious regression

In a cross-section, we speak of "spurious correlation" when the observed correlation between two variables x and y is actually related to their correlations with a third variable z. In that context, regressing x on y shows a significant relationship, but when x is regressed on y and z, the coefficient of y is no longer significant.

This can also happen in a time series context with stationary variables. For instance, we may have two variables that both contain a strong trend, but which are not related to each other. A regression including the trend term will reveal the spurious correlation that would be evident in a y on x regression.

When we are dealing with nonstationary (unit root) processes, characterized as $I(1)$, there is an additional complication: even if there are no trends in the series, we may find that two independent random walks appear to be significantly correlated. Consider the IRWs

$$x_t = x_{t-1} + a_t$$
$$y_t = y_{t-1} + e_t$$

where a_t, e_t are $i.i.d.$ innovations, independent of each other. What if we run the contemporaneous regression

$$\hat{y}_t = b_0 + b_1 x_t$$

and examine the standard t-statistic for b_1? We would hope that since there is no relation between these two variables, the probability limit of b_1 would be zero implying that its t-statistic should be zero 95% of the time. The R^2 should tend to zero as well. However, Granger and Newbold showed that just the opposite result is found in this instance of the **spurious regression problem**. The standard t-statistics will reject the null of $\beta_1 = 0$ over 66% of the time in samples of size 50. As the sample size increases, the problem worsens. This occurs because the error term in this relationship is a random walk under the null, rather than a meanzero $i.i.d.$ error. Including a time trend in the model does not improve this outcome. The R^2 from this regression also is likely to take on a large value. The spurious regression problem will occur with multiple explanatory variables, even if some of them are $I(0)$ and some $I(1)$. This problem is then quite

serious, since it suggests that many of the regressions run on time series data before an understanding of nonstationary processes was developed may indeed be spurious.

Cointegration

The presence of the spurious regression problem suggests that $I(1)$ variables should be used with extreme caution in econometric models. Differencing the variables will generally render them stationary ($I(0)$) but it also limits the kinds of questions we may ask and answer with these data.

The notion of *cointegration*, as proposed by Engle and Granger in 1987, suggests that there are some regressions using $I(1)$ variables that are meaningful. If we have two $I(1)$ processes x_t, y_t, a linear combination such as $y_t - \beta x_t$ is generally an $I(1)$ process for any β. But it is

possible that for some β this expression produces an $I(0)$ process which is covariance stationary: with a constant mean, constant variance and well-behaved autocorrelation function. If such a β exists, we say that x and y are *cointegrated* with cointegrating parameter β. The parameter is not unique, but can be made unique by normalizing the relationship on one variable. There is no notion of causality here, so it does not matter whether we write $y_t - \beta x_t$ or $x_t - \gamma y_t$.

If y_t and x_t are random walk ($I(1)$) processes, then they wander around and may go arbitrarily far from any particular point: after all, they have infinite variance. But if they are cointegrated, then their linear combination will be bounded, and will not wander too far from zero. After all, the linear combination $y_t - \beta x_t$ is of the nature of a regression residual process.

As an example, consider two interest rate series. (Indeed, interest rates cannot travel arbitrarily far in reality: they cannot be negative, and cannot be too large unless inflation is very high). But the annualized six-month and three-month Treasury bill yields both appear to be $I(1)$ processes when measured at a quarterly frequency. But if we consider the spread between them—the special linear combination $r6_t - r3_t$—it appears to be a stationary process. We thus may conclude that $r6$ and $r3$ are cointegrated.

As in many cases of cointegration of economic and financial series, this finding is implied by economic theory. The spread between the three- and six-month rates can be used to define the forward rate, and those two instruments may be used to purchase a three-month forward T-bill. That bill, like any portfolio of Treasury securities, must have a positive yield, and

that puts restrictions on the values which forward rates may assume. The three-month forward rate is the rate which equates the six-month "buy and hold" strategy with the strategy of buying the three-month bill and rolling it over to another three-month bill at that time. By the expectations hypothesis of the term structure of interest rates, those two strategies should have the same expected return. We could violate the expectations hypothesis in the presence of liquidity premia——for instance, investors might pay extra for the convenience of the shorter-tenor instrument——but the two yields cannot be unrelated as long as arbitrage between them can exist. Theory, then, argues that the six-month and three-month Treasury bill yields *should* be cointegrated. There can be temporary deviations from equilibrium——for instance, when the stock market is very volatile, the demand for very short-term Treasuries may be very high——but

cointegration says that in the long run the equilibrium relationship projected by the expectations hypothesis (or its variant with a risk premium) must hold. In this case, we considered the spread itself, in effect setting β to unity; alternatively, we could allow the data to determine β by running an OLS regression of y on x and testing the resulting b against unity. In any case it is the residual from that "Engle–Granger" regression which must be examined for stationarity via a Dickey-Fuller or DF-GLS test.

Under the null hypothesis that x_t and y_t are independent random walks—not cointegrated—the OLS regression of y on x (or x on y) is a spurious regression. Under the alternative hypothesis of cointegration, this regression provides a consistent estimate of the cointegrating parameter. The test must be done then versus the "Dickey–Fuller" distribution, with different

critical values than those appropriate for the standard D–F test to account for the estimation of β. This regression also may be run with a linear trend, although the strict definition of cointegration does not allow for a time trend. If y and x are cointegrated, then we can estimate a model using the levels of those series, as we now discuss.

Error correction models

If two series are stationary, ($I(0)$), then we could write down a model such as we considered in describing the rational distributed lag:

$$y_t = \alpha_0 + \gamma_0 x_t + \rho y_{t-1} + \gamma_1 x_{t-1} + u_t$$

If both y and x are unit root ($I(1)$) processes which are not cointegrated, this model would be inappropriate because it would represent a spurious regression. We could in that case

work with the stationary forms of these series, obtained by differencing:

$$\Delta y_t = \lambda_0 + \gamma_0 \Delta x_t + \rho \Delta y_{t-1} + \gamma_1 \Delta x_{t-1} + v_t$$

where for generality we have added a constant term (corresponding to a trend in the levels equation). We could estimate this equation without difficulty, and derive the parameters of interest (impact effect, delay multipliers and long-run effect) from it.

But if we find that x and y are cointegrated, we have additional $I(0)$ variables that can be included in the relationship. Define $s_t = y_t - \beta x_t$ as the error in the equilibrium relationship at time t. Now we can introduce a lag of s_t in the equation:

$$\Delta y_t = \lambda_0 + \gamma_0 \Delta x_t + \rho \Delta y_{t-1} + \gamma_1 \Delta x_{t-1} + \delta s_t v_t$$

$$\Delta y_t = \lambda_0 + \gamma_0 \Delta x_t + \rho \Delta y_{t-1} + \gamma_1 \Delta x_{t-1} + \\ \delta (y_{t-1} - \beta x_{t-1}) + v_t$$

This is the so-called *error correction* model, which takes into account the fact that there is an equilibrium relationship between the non-stationary level variables x and y. The error correction term s_t expresses the impact of disequilibrium on the relationship. To consider this, work with an even simpler form of the model in differences:

$$\Delta y_t = \lambda_0 + \gamma_0 \Delta x_t + \delta(y_{t-1} - \beta x_{t-1}) + v_t$$

with $\delta < 0$, representing the principle of *negative feedback*. If last period $y_{t-1} > \beta x_{t-1}$, y is above its equilibrium level, so that the error correction term pushes y down. If $y_{t-1} < \beta x_{t-1}$, the error correction term serves to increase y this period. To estimate this model, we merely regress Δy_t on Δx_t and s_t, having estimated s_t in the test for cointegration (and the same for the expanded model above with lags of both y and x in the regression. In this

case, we would be ignoring important information about the joint time series process generating y_t, x_t if we were to run the regression in differences. Including the error correction term allows us to take that information into account.

Vector autoregressions

The *vector autoregression* model (VAR) is a framework used for modeling multivariate relationships: in its pure form, without any restrictions posed by economic theory. It is an extension of the univariate autoregressive model. If we consider two economic timeseries y_1 and y_2, we might model each as a univariate autoregression on several of their respective lagged values. That would ignore any possible interactions between the variables, though. We can capture those interactions—still in the spirit of a proper OLS regression framework, with no

simultaneity taken into account—with a VAR model. A VAR is characterised by two parameters: p, the longest lag in the autoregression, and k, the number of variables in the VAR. The simplest VAR would set $p = 1, k = 2$:

$$\begin{pmatrix} y_{1t} \\ y_{2t} \end{pmatrix} = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} + \begin{bmatrix} \pi_{11.1} & \pi_{12.1} \\ \pi_{21.1} & \pi_{22.1} \end{bmatrix} \begin{pmatrix} y_{1t-1} \\ y_{2t-1} \end{pmatrix} + \begin{pmatrix} \epsilon_{1t} \\ \epsilon_{2t} \end{pmatrix}$$

or in matrix terms

$$y_t = \mu + \Pi_1 y_{t-1} + \epsilon_t$$

A pth order VAR can likewise be written as

$$y_t = \mu + \Pi_1 y_{t-1} + + \Pi_2 y_{t-2} + \ldots + + \Pi_p y_{t-p} + \epsilon_t$$

where each of the Π matrices are estimated coefficients on a particular lag of the set of y variables. This can be viewed as a *reduced form* of a model in which the y variables are jointly determined, and then solved for the reduced form:

$$AY_t = BX_t + \mu + \eta$$

Any linear structural econometric model (including a number of exogenous or predetermined variables X, which may include lagged y variables) can be written in this form, where the off-diagonal coefficients in the A matrix represent the contemporaneous effects of each y on another. If we use economic theory to impose identifying restrictions on the A matrix, we will arrive at coefficient matrices A and B to be estimated by instrumental variables methods. Those coefficient estimates may then be manipulated to yield the structural model's reduced form:

$$
\begin{aligned}
Y_t &= A^{-1}BX_t + A^{-1}\mu + A^{-1}\eta \\
Y_t &= \Phi X_t + \lambda + \zeta
\end{aligned}
$$

This is known as the *restricted reduced form* of the economic model, in that it embodies the restrictions (usually zero restrictions) placed on the elements of A and B by economic theory. We may contrast it with the *unrestricted*

reduced form defined by the VAR model, which does not place any restrictions on the coefficients of the reduced form, but estimates them directly from the data. We have written the VAR above as a pure VAR, with only lagged values of ys as regressors, but a VAR can be written in terms of a number of exogenous variables X as well.

In any case, we may estimate the VAR by a sequence of OLS regressions. Since the VAR has the same variables on the right hand side of each equation, there is no gain to systems estimation. Computationally, it makes sense to take advantage of that common set of regressors in estimating the VAR but that is purely a matter of convenience. But if you did not have a computer program that knew what a VAR was, you could estimate each equation of the VAR with a standard OLS routine. In Stata, there is a suite of commands named

`var...` which are used to estimate and analyze VARs.

What we often want to do with the VAR though, is to evaluate the effects of shocks, or *innovations*. If we have a three-variable VAR system—with, for instance, inflation, unemployment and interest rates—we want to evaluate the effect on each variable of a shock, and trace the time form of the response. Will a shock to inflation caused by oil prices have a permanent effect on any of the three variables, or a transitory effect? If its effect is transitory, how long will it take to dissipate? Those questions may be answered following estimation of the VAR by computing the IRFs—*impulse response functions*—that allow us to trace out these interrelationships.

Before computing these effects, we must be concerned with *stability* of the VAR. A stable VAR is stationary in terms of having time-invariant first and second moments. If the response to a shock does not die out over time,

then the VAR is not stable. (For instance, an oil price shock may have a permanent effect on the price *level*, but it should not have a permanent effect on the rate of inflation). Just as in univariate modeling, if there is a unit root in the VAR representation, then a shock will have infinite memory. We might in this circumstance rely on a VAR fit to the differences of nonstationary variables, or we might better take advantage of a finding of cointegration among those variables and build an enhanced model: the so-called *vector error correction model*, or VECM, as we discuss below.

Consider stability in the context of a first order VAR such as

$$y_t = \mu + \Pi_1 y_{t-1} + \epsilon_t$$

This is a vector difference equation, the solution of which may be written as

$$y_t = (I + \Pi_1 + \Pi_1^2 + \ldots + \Pi_1^{t-1})\mu + \Pi_1^t y_0$$
$$+ \sum_{i=0}^{t-1} \Pi_1^i \epsilon_{t-i}$$

If the model is stable, then the bracketed sum of terms in Π_1 has a finite limit as $t \to \infty$ of $(I - \Pi_1)^{-1}$. That expression is the matrix of long run multipliers, representing the steady state effect of a shock. That effect may be other than zero, but if the VAR is to be stable, the dynamic multiplier representing the additional effect of the shock in a later period must eventually go to zero. The condition that ensures that this finite limit exists relates to the eigenvalues of the matrix Π_1, which all must have a modulus less than one.

In a univariate context, consider the second order autoregression

$$
\begin{aligned}
y_t &= \phi_1 y_{t-1} + \phi_2 y_{t-2} + \epsilon_t \\
&= (\phi_1 L + \phi_2 L^2) y_t + \epsilon_t \\
(1 - (\phi_1 L + \phi_2 L^2)) y_t &= \epsilon_t \\
\phi(L) y_t &= \epsilon_t
\end{aligned}
$$

Stability requires that the roots of this polynomial in the lag operator have modulus greater

than one. If $\phi_1 = 3/4, \phi_2 = 1/4$ then the model is

$$(1 - 3/4L - 1/4L^2)y_t = \epsilon_t$$

If we evaluate the function at $L = 1$, it equals zero. The polynomial $\phi(L)$ may be factored into $(1 - L)(1 + 1/4L)$, which has roots of $+1$ and -4. The root of $+1$ is a unit root. We could then difference the data:

$$(1 + 1/4L)[(1 - L)y_t = \epsilon_t$$

so that we now have a first-order autoregression in the *differences* of y_t,

$$\triangle y_t = -1/4 \triangle y_{t-1} + \epsilon_t$$

which will be stable. However, merely taking differences may not render a model stable; if we have the model

$$(1 + 3L - 4L^2)y_t = \epsilon_t$$

the lag polynomial factors into $(1 - L)(1 + 4L)$. The differences of y_t are not a stable autoregressive process, since they have a coefficient

of 4. The model is unstable in both levels and first differences.

Let us now consider how this works in the multivariate domain of a VAR. The condition for stability in a multivariate system can be stated in terms of the *characteristic polynomial*,

$$|\pi_1 - \lambda I| = 0$$

All eigenvalues (roots) of this polynomial must have moduli less than one. For instance, if

$$\pi_1 = \begin{pmatrix} 5/8 & 1/2 \\ 1/4 & 5/8 \end{pmatrix}$$

Then we seek the roots of the polynomial

$$(5/8 - \lambda)^2 - 1/8 = 0$$
$$(0.978 - \lambda)(0.271 - \lambda) = 0$$

with real roots of 0.978 and 0.271. This VAR is stable. If on the other hand the (2,2) element of $\pi_1 = 2/3$ rather than 5/8, the VAR

would have roots of (1, 7/24), and would therefore fail to meet the stability condition. The Stata command `varstable` may be used after estimation of a VAR to evaluate the stability properties of the system.

We can always rewrite a 2nd, 3rd, … order VAR system in first order form. E.g.,

$$y_t = \mu + \pi_1 y_{t-1} + \pi_2 y_{t-2} + \epsilon_t$$

can be rewritten as the first order system

$$\begin{pmatrix} y_t \\ y_{t-1} \end{pmatrix} = \begin{pmatrix} \mu \\ \mathbf{0} \end{pmatrix} + \begin{pmatrix} \pi_1 & \pi_2 \\ \mathbf{I} & \mathbf{0} \end{pmatrix} \begin{pmatrix} y_{t-1} \\ y_{t-2} \end{pmatrix} + \begin{pmatrix} \epsilon_t \\ \mathbf{0} \end{pmatrix}$$

and the stability condition can be evaluated by examining the dynamic matrix of this first order system and calculating its eigenvalues.

Given that the dynamic matrix is square but non-symmetric, it need not have real eigenvalues; there may be one or more complex conjugate pairs. If so, we must evaluate their moduli (square root of the sum of squares of the

real and imaginary parts of the eigenvalues). We will see an example of this below. In a dynamic system, complex eigenvalues give rise to cycles, while real eigenvalues imply monotonic behavior.

VARs come in three flavors: *reduced form, recursive* and *structural*. We have considered the reduced form VAR in which only lagged values of the dependent variables (and possibly exogenous variables) appear on the right hand side. The difficulty with reduced form VARs is that the error terms in the equations will in general be correlated across equations. This makes difficult the notion of tracing out the effects of a one-unit shock to one equation, since in the data such a shock will generally be correlated with shocks to the other equations. This is usually dealt with by performing a Choleski factorization of the error covariance matrix into upper triangular form, which essentially *orthogonalizes* the innovations. That

is, the new innovations are uncorrelated by construction with each other. This approach comes with a cost: the resulting computations of the shocks' effects is sensitive to the order of the equations in the VAR. A Choleski factorization of the equations listed in a different order will yield different results.

The second approach to VARs, *recursive* estimation, constructs the error terms in each equation to be uncorrelated with errors in the prior equations. Here, too, we must choose a causal ordering. Consider a three-variable VAR in inflation, the unemployment rate and the interest rate, in that order. The inflation rate, as the first equation, is modeled as a pure reduced form. In the unemployment rate equation, we include the lagged values of all variables *plus* current inflation. In the third equation, we include the lagged values of all variiables plus the current values of inflation and the unemployment rate. This is equivalent to stating

that inflation has a *contemporaneous* effect on unemployment, but not *vice versa*, and that inflation and unemployment have contemporaneous effects on the interest rate, but not *vice versa*. The results of analyzing a recursive VAR will thus depend on the causal ordering chosen. With N variables in the VAR, there are $N!$ possible orderings.

The third flavor of VAR is the *structural VAR* or SVAR. Here we use economic theory to sort out the contemporaneous relationships between the variables. For instance, we may assume that the Fed reacts to contemporaneous deviations of inflation and unemployment from their target values by adjusting the current period's interest rate via a *Taylor rule*. The Fed's behavior could also be modeled as forward-looking, reacting to forecasts of inflation and unemployment over several quarters' horizon.

Granger causality

One concept often applied in VAR analysis is *Granger causality*: do lagged values of one variable (y_1) help us to predict the current values of another variable (y_2) after the past values of y_2 have been controlled for? If they do, then we say that y_1 Granger-causes y_2, and *vice versa*. In the absence of Granger causality, past values of y_1 are not systematically related to the current value of y_2. We can determine whether there are significant effects using a "block F" test of the coefficients on lagged y_1 in the y_2 equation of a VAR. The Stata command `vargranger` conducts tests for Granger causality following estimation of a VAR.

Impulse response functions

A second tool used to analyze the VAR's implications is the *impulse response function*, or IRF. A pth-order VAR can be written, using the lag operator L, as:

$$(I - \pi_1 L - \pi_2 L^2 - \ldots - \pi_p L^p) y_t = \mu + \epsilon_t$$

If the VAR satisfies the stability conditions, then we may compute the inverse of the lag polynomial and rewrite the system:

$$y_t = (I - \pi_1 L - \pi_2 L^2 - \ldots - \pi_p L^p)^{-1} (\mu + \epsilon_t)$$

which expresses y_t as an infinite moving average ($MA(\infty)$) of the entire history of the ϵ process. The IRF is then the sequence of derivatives

$$\frac{\partial y_{i,t+s}}{\partial \epsilon_{jt}}$$

for $i, j = 1, N$ and $s > 0$. These derivatives trace out the response of current and future values of each of the variables to a one-unit increase in the current value of one of the VAR errors under the assumption that this error returns to its expected value of zero in subsequent periods and that all other errors are equal to zero. There will be N^2 such sets of derivatives, since they are defined for each combination of dependent variable and innovation. If we consider a reduced form VAR

in which the Choleski decomposition has been applied to the error covariance matrix, we have an *orthogonalized IRF* in which the original VAR innovations $(\epsilon_1, \ldots, \epsilon_N)$ have been decomposed into a set of uncorrelated components $(u_1, \ldots, u_N)$. The $OIRF$ is then the set of derivatives

$$\frac{\partial y_{i,t+s}}{\partial u_{jt}}$$

which express the dynamic multipliers of the system to a set of individual shocks which can be considered, in turn, as independent of the other shocks. A suite of Stata commands named `irf...` are available to estimate IRFs and OIRFs after VAR estimation, and present them in either tabular or graphical form.

Forecast error variance decompositions

The third tool commonly employed in VAR analysis is the *forecast error variance decomposition* ($FEVD$). This is a form of accounting

that evaluates the percentage of the variance of the error made in forecasting a variable in the VAR due to a specific shock at a given horizon. The forecast error variance decomposition is like a partial R^2 for the forecast error, given a forecast horizon. It expresses the importance of each type of shock in the forecast errors encountered, in-sample, for short, medium and longer-term forecasts. The `fevd` option on Stata commands can be used to generate FEVDs in tabular or graphical form.

Vector error correction models

Consider a set of y variables which are determined to possess unit roots. We could formulate a VAR in first differences of the variables:

$$\Delta y_t = \mu + \pi_1^* \Delta y_{t-1} + \epsilon_t^*$$

But such a formulation provides no information about the relationship between the level

of the variables in the VAR, and it is this aspect on which economic theory is usually the most informative.

An alternative arises when the variables in y_t are cointegrated: that is, there is a linear combination of the $I(1)$ variables in y_t that is itself $I(0)$. So the $I(0)$ variables that may be modeled include not only the first differences of the variables but the cointegrating combinations formed from the $I(1)$ variables. When $k > 2$ there may be more than one linear combination of $I(1)$ variables that is stationary, each of which is a candidate regressor. A model in first differences that encompasses not only past differences but past disequilibria between the level variables is capable of representing not only short-run responses but also long-run relationships. This model is known as the *vector error correction model*, or VECM, as proposed by Engle and Granger (1987). If the k-vector

of variables y_t are cointegrated ($CI(1,1)$) then there exists an error correction representation of the general form:

$$\Delta y_t = \alpha z_{t-1} + \Gamma_1 \Delta y_{t-1} + \Gamma_2 \Delta y_{t-2} + \ldots$$
$$+ \Gamma_p \Delta y_{t-(p-1)} + \epsilon_t^*$$

where $z_{t-1} = \beta' y_{t-1}$ are the r linear cointegrating combinations among the k variables, with β is the $k \times r$ matrix of cointegrating vectors.

The long-run or equilibrium relationships among the levels of the variables are captured by the cointegrating combinations, $z_{t-1} = \beta' y_{t-1}$. Nonzero values of z_{t-1} indicate lagged disequilibria which are eradicated by the adjustment coefficients in α, a $k \times r$ matrix of coefficients. Each column of α is associated with one of the r stationary cointegrating combinations. Short-run dynamic adjustments are captured by nonzero values of the elements in Γ_i.

In Stata, there is a suite of commands named `vec...` which are used to estimate and analyze VECMs. In a model containing more than two y_t variables, the Engle–Granger regression can detect the existence of cointegration, but cannot establish the dimension of the *cointegrating space*: the parameter r above, indicating the number of cointegrating vectors. In a k-variable system, there can be up to $(k-1)$ cointegrating vectors. Methods developed by Johansen and Juselius are used to test for the dimensionality of the cointegrating space (in Stata, the `vecrank` command).