Notes Section 6

Cointegration

Consider the system

$$X_t = X_{t-1} + \epsilon_t$$
$$Y_t = \alpha + \beta X_t + \omega_t$$

with ϵ_t and ω_t each iid, standard Normal variates. There are no dynamics in the DGP, and we assume there is no contemporaneous correlation between the two error processes. In this case, OLS will find unbiased and consistent estimates of α and β, despite the fact that the variance of X is not bounded. X and Y are said to be cointegrated, or to possess a cointegrating relationship. Y will be $I(1)$, since it is composed of an $I(1)$ variable plus a stationary error.

On the other hand, consider two independent random walks, X and Y, and the regression of Y upon X. In that regression, the true slope coefficient is zero, since there is no relationship whatsoever between these two integrated processes. But the regression (run either way) will yield a nonzero estimate of the slope coefficient, and the significance of that coefficient will not diminish with sample size. Indeed, the probability of rejecting the (true) null will increase with sample size. In a simple Monte Carlo experiment, the rejection rate for the t statistic is 72% in samples of size 50, but 80% for $T = 100$, and 91% for $T = 500$, whereas the rejection rate should be 10%. Furthermore, R^2 of such a regression becomes a random variable in a regression of unrelated nonstationary variables, and the likelihood of finding a sizable R^2 in that context is quite sizable.

Demonstrably, OLS does not yield consistent estimates of the true slope parameter in this instance: the case of a **spurious regression**, as defined by Granger and Newbold (1974). A theoretical explanation of this phenomenon was presented by Phillips (1986). The problem of spurious regressions appears with $I(1)$ variables, so the determination of unit root processes is essential. Furthermore, the problem will not arise if the series are cointegrated, so that determining whether cointegration exists is important as well.

What is cointegration? The notion that two (or more) nonstationary processes may be following the same stochastic trend, or may share an underlying common factor. Although X and Y are both $I(1)$, a linear combination of those nonstationary variables may exist which is itself $I(0)$. The coefficients in that linear combination form the cointegrating vector, which will have one element normalized to unity, since the CI vector is defined up to a factor of proportionality. Additionally, the vector may include a constant, to allow for unequal means of the two variables, so

$$Y_t = \varphi_1 + \varphi_2 X_t + \xi_t$$

and the notion is that ξ_t will be a stationary process in the presence of cointegration.

The concept may be extended to higher orders of integration, and more than two variables. If X and Y are each $I(2)$, then a linear combination of them might be $I(1)$, or even $I(0)$. We can speak of series as being cointegrated $CI(d,b)$, where d is the common order of integration of the variables, and b is the reduction in the order of integration of the cointegrating combination. Thus, the case above would be $CI(1,1)$, with two $I(1)$ variables forming a combination one order lower, of $I(0)$. A linear combination of $I(1)$ variables is not a spurious regression if it is stationary.

The Engle–Granger approach

The original approach to testing for cointegration is that of Engle and Granger (1987). In this classic paper, they demonstrate that if one regresses an $I(1)$ variable upon another $I(1)$ variable (in what is termed a "balanced" regression, in which all variables share the same order of integration), the residuals from that regression may then be subjected to a unit root test. The null hypothesis, in this case, is that of **noncointegration**: that is, failing to reject a unit root in the errors in a Dickey-Fuller style test will yield the conclusion of nonCI, whereas rejection in favor of stationarity in the error process will be evidence of a cointegrating relationship among the variables. Stock (1987) has shown that the OLS estimates in this regression have the desirable property of "superconsistency"—that is, they are not only consistent estimates of the underlying parameters of the DGP, but they converge on the population values more quickly than OLS estimates in the context of stationary regressors. The ADF-type test applied in this instance will not contain a constant term, since the OLS residuals will be mean zero with a constant included in the CI regression. The critical values in this case are not the same as those of the standard D–F distribution, since the timeseries being tested is a generated series. They are larger negative values than those provided for the D–F distribution, and like D–F critical values, are obtained by simulation. Note also that the regression could be run with either variable on the left hand side, since it is not a structural relationship; if a CI vector exists, it may be renormalized on the other variable. If the R^2 in the CI regression is low, however, (less than 0.8), then the inference may differ depending upon normalization.

How would one operate with more than two variables? One may still form a cointegrating vector among three or more $I(1)$ series, and estimate the CI regression. As in the two–variable case, the estimates of the CI coefficients will be superconsistent. However, with more than two variables, the weakness of the E-G approach emerges: the test can determine whether a CI vector exists that yields stationary errors, but it will generate only one of the possible multiple CI vectors that could exist in this setting. When there are three variables, for instance, they could all be driven by the same common factor (stochastic trend), or their behavior could reflect two common factors–coinciding with the existence of two CI vectors. The E–G approach is not capable of finding more than one CI vector; other approaches, such as Johansen and Juselius' ML approach, can do so. (Of course, it could be that there are zero common factors underlying these three variables' dynamics, in which case the E–G approach will correctly reflect the absence of CI among them).

Cointegration and the error correction model

A simple error correction model (ECM) may be written as

$$Y_t - Y_{t-1} = \theta_1 \Delta X_t + \theta_2 (Y_{t-1} - Y_{t-1}^*) + \epsilon_t$$

In this formulation, also known as a partial adjustment scheme, there is adjustment of Y toward a target Y^* which depends on the lagged disequilibrium. Imagine that there is a constant ratio, in equilibrium, between consumption and income (in logs) c and y, so that $c_t = k + y_t$. Then a measure of disequilibrium may be written as $\zeta_t = c_t - (k + y_t)$. An error correction scheme might be written as

$$\Delta c_t = \theta_1 \Delta y_t + \theta_2 \zeta_{t-1} + \epsilon_t$$

where consumers react to last period's disequilibrium by revising their consumption. Substituting in, we have

$$\Delta c_t = \theta_0 + \theta_1 \Delta y_t + \theta_2 \left(c_t - y_t\right)_{-1} + \epsilon_t$$

where the parenthesized quantity is the error correction term. Consumption will change if income changes, or if there was a disequilibrium in the relationship last period. (Note that the proportionality factor k has been subsumed in the constant term of this relationship; we could, instead, leave it in the error correction term as a coefficient on y). The coefficient θ_2 has limits $-1 \leq \theta_2 < 0$, since for stability one should not overadjust to the disequilibrium (which would correspond to closing the entire gap this period) nor to fail to adjust at all (a coefficient of zero), let alone a positive coefficient (which would drive c away from its equilibrium relationship). The ECM contains both the short–run mechanism by which consumption will adjust to current changes in income, as well as the long–run adjustment to equilibrium. The relative importance of short-run and long-run fluctuations in consumption will be governed by the relative magnitudes of the θ_1 and θ_2 coefficients. Incorporating the equilibrium coefficient in the relationship, we have

$$\Delta c_t = \theta_0 + \theta_1 \Delta y_t + \theta_2 (c_t - \varphi_2 y_t)_{-1} + \epsilon_t \tag{1}$$

corresponding to the "levels" or static relationship

$$c_t = \varphi_1 + \varphi_2 y_t + \xi_t \tag{2}$$

If these variables are both $I(1)$, this relationship may be (super)consistently estimated by OLS, and the residuals tested for noncointegration. If the null of nonCI is rejected, then the relationship (1) may be estimated, replacing the unknown ζ_{t-1} with the lagged residuals from (2), by OLS.

If c_t and y_t are both $I(1)$ and cointegrated $CI(1,1)$, and neither has a trend in the mean, then by the Granger Representation Theorem (Granger and Weiss, 1983) there will always exist an error correction representation of the form

$$\begin{aligned} \Delta c_t &= lagged(\Delta c_t, \Delta y_t) + \theta_{2c}\xi_{t-1} + \upsilon_{1t} \\ \Delta y_t &= lagged(\Delta c_t, \Delta y_t) + \theta_{2y}\xi_{t-1} + \upsilon_{2t} \end{aligned} \tag{3}$$

where $\xi_t = (c_t - \varphi_1 + \varphi_2 y_t)$, $\upsilon_{1t} = \omega(L)\varepsilon_{1t}$ and $\upsilon_{2t} = \omega(L)\varepsilon_{2t}$, with the ε sequences white noise. It must be so that $|\theta_{2c}| + |\theta_{2y}| \neq 0$; that is, the lagged disequilibrium term must appear in at least one of the equations. In general, it will appear in both. Note that this model is essentially a VAR augmented by the error correction term. The transformation of the dynamic model into this form illustrates that a regular VAR in the differences of these variables would be misspecified, in that it would omit the error correction term. That term is required to fully specify the dynamics of the model. In its absence, the reversion of these variables to a long–run equilibrium is not modeled.

These equations are “balanced” in that if the levels variables are $I(1)$, their differences are $I(0)$. If the variables are $CI(1,1)$, then the error correction term will be $I(0)$, and all terms in (3) are $I(0)$. If these two variables are $I(1)$ and $CI(1,1)$, then knowledge of the one variable helps to forecast the other—at least in one direction. Having established cointegration as a long-run property of the data, it is natural to think of an ECM as an appropriate way of capturing the dynamic adjustments of these variables to the long run.

Multiple cointegrating relationships

Consider a set of time series variables. If they are each (trend–)stationary, a VAR may be employed to estimate their joint evolution and interdependence. If one or more of the variables in the VAR are nonstationary, it would be inappropriate to estimate the VAR in levels. However, it might also be inappropriate to estimate a VAR in differences, even in the case where all variables in the VAR possess unit roots. If there are cointegrating relationships among the level variables, the proper representation will be the error correction model (ECM), and the VAR in differences may be seen to be a misspecified version of that model (excluding, as it does, the error correction term). The VAR in differences is also uninformative about the long–run behavior of the series, since it only expresses their short–run paths of adjustment, without any link to the long–run equilibrium relationships among the variables. The error correction model explicitly provides that link, capturing the short–run adjustment toward the long–run equilibria.

When there are more than two variables in the set, there may be multiple cointegrating relationships among them. In a two–variable system, the variables either form a cointegrating combination or they do not. In a three–variable system, there may be zero, one, or two cointegrating vectors. If zero, then these are three independent random walks. If there are two CI vectors, the Engle–Granger procedure will locate one of them, but it is incapable of identifying the multiplicity, or of estimating a second relationship. Likewise, for higher–order systems of order k, there may be up to $k-1$ CI vectors defining long–run equilbria among the variables.

The most common methodology employed to evaluate multiple cointegrating relationships is that of Johansen (1988) and Johansen and Juselius (1990), which is based on the estimation of a p^{th}–order VAR in the k variables. The VAR in the k–vector y is:

$$y_t = \Pi_1 y_{t-1} + \Pi_2 y_{t-2} + \ldots + \Pi_p y_{t-p} + \Psi D_t + \epsilon_t \qquad (4)$$

where D_t is a d–vector of deterministic terms, such as a constant, trend and seasonal dummies. The VAR may be reparameterized into an ECM:

$$\Delta y_t = \Pi y_{t-1} + \Gamma_1 \Delta y_{t-1} + \Gamma_2 \Delta y_{t-2} + \ldots + \Gamma_{p-1} \Delta y_{t-(p-1)} + \Psi D_t + \epsilon_t \qquad (5)$$

No assumption is made about the rank of Π. In the decomposition $\Pi = \alpha\beta'$, α and β are k x k matrices. We seek to determine whether any columns of β (that is, rows of β') are statistically indistinguishable from zero vectors. The existence of r cointegrating vectors reduces the rank of Π by $k - r$: that is, if there are r cointegrating relationships among the variables, then there will be r nonzero eigenvalues in the dynamic system, and $k - r$ zero eigenvalues. The decomposition will then relate $\Pi = \alpha\beta'$ where α and β are both k x r matrices. If the CI rank is full, that is, $r = k$, then the VAR is stationary in the levels. If the rank is zero, then there is no implied long run, and the VAR may be safely reformulated in first differences.

The Johansen methodology provides inference on the number of nonzero eigenvalues, or CI relationships, by setting up an eigenvalue problem derived from the levels and differences of the k variables. The eigenvalues are ordered, from largest to smallest. The space spanned by the r largest eigenvalues is the r–dimensional cointegrating space. If $r = 1$, β is k x 1, and is the eigenvector corresponding to the largest eigenvalue. If $r = 2$, β is k x 2; the first column is as before, and the second column is the eigenvector corresponding to the second largest eigenvalue.

Two statistics are defined in Johansen's work to determine the CI rank: first, the trace statistic,

$$trace = -T \sum_{i=r+1}^{k} ln(1 - \lambda_i) \tag{6}$$

which allows for the test of $H(r)$: the rank of Π is r, against the alternative that the rank of Π is k. A large value of the trace statistic is evidence against $H(r)$: that is, with $r = 1$, a value of the trace statistic greater than the appropriate critical value allows us to reject $r = 1$ in favor of $r > 1$. The test may then be repeated for $r = 2$, and so on.

Alternatively, the λ_{max} statistic may be used:

$$\lambda_{max} = -Tln(1 - \lambda_{r+1}) \tag{7}$$

This test allows for the comparison of a CI rank of r against the alternative of a CI rank of $r + 1$. This test also may then be repeated for larger values of r until one fails to reject the null hypothesis.

The distribution of both statistics is nonstandard, and depends on nuisance parameters in D_t. Critical values have been tabulated by Johansen and Osterwald–Lenum, and are reproduced in the textbook. Research by Reimers and Cheung and Lai (1993) have identified small–sample biases in the tabulated values of these test statistics, and they recommend applying a small–sample adjustment.

Extensions of the Johansen methodology include tests of various restrictions on the CI vectors: either zero (exclusion) restrictions, indicating that certain variables should not appear in certain of the equilibrium relationships, or restrictions on parameters' values, such as those forthcoming from theory (e.g. purchasing power parity not only specifies a long run relationship, but indicates that the coefficients in the CI combination should be (1, 1, -1)).

References

[1] Cheung, Y.-W., and K. S. Lai. 1993. Finite–sample sizes of Johansen's likelihoood ratio tests for cointegration. Oxford Bulletin of Economics and Statistics, 55, 313–328.

[2] Engle, R., and C.W.J. Granger. 1987. Cointegration and error correction: Representation, estimation and testing. Econometrica, 55, 251–276.

[3] Granger, C.W.J., and P. Newbold. 1974. Spurious regressions in econometrics. Journal of Econometrics, 2, 111–120.

[4] Granger, C.W.J., and A. Weiss. 1983. Time–series analysis of error correction models, in S. Karlin, T. Amemiya and L. Goodman, eds., Studies in Econometrics, Time Series and Multivariate Statistics. Academic Press.

[5] Johansen, S. 1988. Statistical analysis of cointegration vectors. Journal of Economic Dynamics and Control, 12, 231–254.

[6] Johansen, S. and K. Juselius. 1990. Maximum likelihood estimation and inference on cointegration with applications to the demand for money. Oxford Bulletin of Economics and Statistics, 522, 169–210.

[7] Phillips, P.C.B. 1986. Understanding spurious regressions in econometrics. Jorunal of Econometrics, 33, 311–340.

[8] Stock, James. 1987. Asymptotic properties of least squares estimators of co–integrating vectors. Econometrica, 55, 1035–1056.