11 Cointegration and the Error Correction Models

When we model multiple time series simultaneously, we may encounter the situation that more than one component series are non-stationary, and they could be drive by some common stochastic trends. For example, let $\{\varepsilon_t\}$ and $\{u_t\}$ be independent white noise, $\Delta x_t = \varepsilon_t$ and $y_t = 2x_t + u_t$. Then $\{x_t\}$ is a random walk, and $\{y_t\}$ is a random walk plus some stationary disturbances. Both series are non-stationary, and they are driven by the same random walk. This system could also be written as

$$\begin{bmatrix} x_t \\ y_t \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 2 & 0 \end{bmatrix} \begin{bmatrix} x_{t-1} \\ y_{t-1} \end{bmatrix} + \begin{bmatrix} 1 & 0 \\ 2 & 1 \end{bmatrix} \begin{bmatrix} \varepsilon_t \\ u_t \end{bmatrix},$$

or more compactly,

$$r_t = Ar_{t-1} + v_t,$$

where $r_t = (x_t, y_t)'$,

$$A = \begin{bmatrix} 1 & 0 \\ 2 & 0 \end{bmatrix},$$

and v_t is white noise with variance

$$\begin{bmatrix} 1 & 0 \\ 2 & 1 \end{bmatrix} \mathrm{Var}\left(\begin{bmatrix} \varepsilon_t \\ u_t \end{bmatrix}\right) \begin{bmatrix} 1 & 2 \\ 0 & 1 \end{bmatrix}.$$

11.1 Cointegration

If we know that two non-stationary time series $\{x_t\}$ and $\{y_t\}$ are related through $y_t = \beta x_t + u_t$ where u_t are some weakly stationary time series, we may want to estimate β by OLS. The corresponding estimator is given by

$$\hat{\beta} = \left(\frac{1}{T}\sum_{t=1}^{T} x_t^2\right)^{-1} \left(\frac{1}{T}\sum_{t=1}^{T} x_t y_t\right).$$

As oppose to the case in which $\{x_t\}$ and $\{y_t\}$ are weakly stationary, in the non-stationary case $\sqrt{T}(\hat{\beta}-\beta)$ does not converge in distribution to some normal random variable any more. In particular, in the case when $\Delta x_t = \varepsilon_t$ for some white noise $\{\varepsilon_t\}$, we have

$$\begin{aligned} T\left(\hat{\beta}-\beta\right) &= \left(\frac{1}{T^2}\sum_{t=1}^{T} x_t^2\right)^{-1}\left(\frac{1}{T}\sum_{t=1}^{T} x_t u_t\right) \\ &\to_d \left(\int_0^1 W_1(r)^2 \mathrm{d}r\right)^{-1}\left(\int_0^1 W_1(r)\mathrm{d}W_2(r)+\delta\right) \end{aligned}$$

where δ is a term that depends on the correlation between u_t and ε_t.

The essence of the cointegration is that the non-stationarity of $\{y_t\}$ comes from the non-stationary of $\{x_t\}$ through some linear relationship. After the cointegration, the non-stationarity vanishes. That is, the remaining part $\{u_t\}$ is weakly stationary. This could be generalized to the case where x_t is a non-stationary vector process which is weakly stationary after first difference, $y_t = x_t'\beta + u_t$ and u_t is weakly stationary.

There could be the case that the two time series $\{x_t\}$ and $\{y_t\}$ are non-stationary, but they do not completely share the same source of non-stationarity. That is,

$$y_t = x_t'\beta + e_t,$$

but e_t is non-stationary for all choice of β.

Now if we estimate β by OLS, we will get into the trouble of the so-called spurious regressions. In particular, if $\{e_t\}$ is weakly stationary after first difference, then

$$\hat{\beta} \to_d \left(\int_0^1 W_1(r)W_1(r)'\mathrm{d}r\right)^{-1}\left(\int_0^1 W_1(r)\mathrm{d}W_2(r)\right),$$

$$\frac{1}{T}W(\hat{\beta}) \to_d \frac{\int_0^1 W_2W_1'\left(\int_0^1 W_1W_1'\right)^{-1}\int_0^1 W_1W_2}{\int_0^1 W_2^2 - W_2W_1'\left(\int_0^1 W_1W_1'\right)^{-1}\int_0^1 W_1W_2},$$

and

$$R^2 \to_d \frac{W_2 W_1' \left(\int_0^1 W_1 W_1'\right)^{-1} \int_0^1 W_1 W_2}{\int_0^1 W_2^2},$$

where $\hat{\beta}$ is the OLS estimator of β, $W(\hat{\beta})$ is the Wald test statistic for the null hypothesis that $\beta = 0$, and R^2 is the R-square of the regression. We see from the above results that the OLS estimator of β is random and does not converge to the true value of β even if the sample size T is large. The Wald statistic diverges at the rate T, which will lead us to mistakenly reject the null hypothesis $\beta = 0$ even if y_t is not quite related to x_t. The R^2 is also random as the sample size gets very large, and usually the values are close to one (once again even if x_t does not have any explanation power for y_t).

11.2 Cointegrated VAR Models

To better understand what happens in a non-stationary VAR model where there may be cointegration relationships among components of the time series, we look at two examples. Both examples are VAR(1) models in the form of

$$x_t = Ax_{t-1} + \varepsilon_t$$

where $\{x_t\}$ is a two-dimensional vector process, A is the 2×2 AR coefficient matrix, and ε_t is two-dimensional white noise. In the first example,

$$A = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}.$$

The model becomes

$$\begin{bmatrix} x_{1t} \\ x_{2t} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x_{1,t-1} \\ x_{2,t-1} \end{bmatrix} + \begin{bmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{bmatrix}.$$

Note that we have

$$x_{1t} = x_{1,t-1} + \varepsilon_{1t}$$

and

$$x_{2t} = x_{2,t-1} + \varepsilon_{2t}.$$

Both $\{x_{1t}\}$ and $\{x_{2t}\}$ are non-stationary, and we cannot find any β such that $x_{1t} + \beta x_{2t}$ is weakly stationary, as long as ε_{1t} and ε_{2t} are not perfectly correlated. In this case, we may simply difference the series and conduct analysis using the differenced series. Note that in this case, the AR coefficient A has two unit eigenvalues.

Now consider the second example where

$$A = \begin{bmatrix} \frac{1}{2} & -1 \\ -\frac{1}{4} & \frac{1}{2} \end{bmatrix}.$$

The model is given by

$$\begin{bmatrix} x_{1t} \\ x_{2t} \end{bmatrix} = \begin{bmatrix} \frac{1}{2} & -1 \\ -\frac{1}{4} & \frac{1}{2} \end{bmatrix} \begin{bmatrix} x_{1,t-1} \\ x_{2,t-1} \end{bmatrix} + \begin{bmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{bmatrix}.$$

It is easy to verify that

$$x_{1t} + 2x_{2t} = \varepsilon_{1t} + 2\varepsilon_{2t},$$

and

$$x_{1t} - 2x_{2t} = x_{1,t-1} - 2x_{2,t-1} + (\varepsilon_{1t} - 2\varepsilon_{2t}).$$

Then we have found a linear combination of x_{1t} and x_{2t}, namely $x_{1t} + 2x_{2t}$, such that the combined series is weakly-stationary, and another linear combination, namely $x_{1t} - 2x_{2t}$, such that the combined series is a random walk, and therefore, non-stationary. Any other linear combination $ax_{2t} + bx_{2t}$ could be written as a linear combination of $x_{1t} + 2x_{2t}$ and $x_{1t} - 2x_{2t}$, i.e., a linear combination of a weakly stationary process and a non-stationary process. Therefore, these combined series must be non-stationary. In particular, the series x_{1t} and x_{2t} are non-stationary. This argument shows that there is only one cointegration relationship between x_{1t} and x_{2t}, given by $x_{1t} + 2x_{2t} = \varepsilon_{1t} + 2\varepsilon_{2t}$.

The above results could be obtained by left multiply both sides of the model in the second example by

$$\begin{bmatrix} 1 & 2 \\ 1 & -2 \end{bmatrix}.$$

This operation could be viewed as a change of the coordinate system. What happens above is that for the process $\{x_t\}$, if we decompose x_t according to the usual orthogonal coordinate system, we get two coordinate processes, namely x_{1t} and x_{2t}, both of which are non-stationary. If we only stick to this decomposition, we may likely to conclude that there is two dimensional non-stationarity in the process. However, if we decompose x_t according to the coordinate system given by the matrix above, we get two coordinate processes of which one is non-stationary and the other is stationary. So in fact, the process of x_t has only one-dimensional non-stationarity! In this case, differencing the series will difference both the stationary part and the non-stationary part. The stationary part is the part that does not require differencing. And if we difference a stationary series, all the information in the series will be lost!

We can connect the above results with the eigenvalues of the AR coefficient matrix. The AR coefficient matrix has two eigenvalues, namely 1 and 0. So as opposed to the first example, in this example there are only one unit root, and the dimension of the non-stationary component of the resulting process is one.

How to obtain the above two particular linear combinations? We may use the error correction representation of the VAR model. We write

$$\Delta x_t = x_t - x_{t-1} = (A - I)x_{t-1} + \varepsilon_t,$$

where

$$A - I = \begin{bmatrix} -\frac{1}{2} & -1 \\ -\frac{1}{4} & -\frac{1}{2}. \end{bmatrix}$$

Note that this matrix is not of full rank. Therefore, we may write $A - I = \alpha\beta'$ where both α and β are some $N \times m$ matrices, and m is the rank of $A - I$. Note that α and β are not uniquely determined, but the space spanned by their column vectors are unique. In our example, we could let $\alpha = (-1, -1/2)'$ and $\beta = (1/2, 1)'$. Then the cointegrating relationship is given by $\beta' x_t$, and the "random walk" relationship is given by $\alpha'_\perp x_t$ where $\alpha_\perp$ is any vector that is orthogonal to α.

11.3 The Error Correction Models

In the previous section, we encountered the simplest error correction model given by

$$\Delta x_t = \alpha\beta' x_{t-1} + \varepsilon_t.$$

We may consider a more general error correction model of the form

$$\Delta x_t = \mu + \alpha\beta' x_{t-1} + \Phi_1 \Delta x_{t-1} + \cdots + \Phi_p \Delta x_{t-p} + \varepsilon_t$$

for $\varepsilon_t \sim \text{WN}(0, \Sigma)$ and some $N \times m$ matrices α, β where $m < N$. The column vectors in β give the cointegrating relationships, i.e., $\beta' x_t$ is weakly stationary.

To estimate the model, we consider two linear regressions

$$\Delta x_t = \mu_0 + \Omega_1 \Delta x_{t-1} + \cdots + \Omega_p \Delta x_{t-p} + u_t$$

and

$$x_t = \mu_1 + \Theta_1 \Delta x_{t-1} + \cdots + \Theta_p \Delta x_{t-p} + v_t.$$

We denote the residuals from the two regressions by $\hat{u}_t$ and $\hat{v}_t$ respectively. Then we construct

$$S_{00} = \frac{1}{T} \sum_{t=p+1} \hat{u}_t \hat{u}_t',$$

$$S_{01} = \frac{1}{T} \sum_{t=p+1} \hat{u}_t \hat{v}_t',$$

$$S_{10} = \frac{1}{T} \sum_{t=p+1} \hat{v}_t \hat{u}_t',$$

$$S_{10} = \frac{1}{T} \sum_{t=p+1} \hat{v}_t \hat{v}_t',$$

and solve the eigenvalue problem

$$S_{10} S_{00}^{-1} S_{01} e = \lambda S_{11} e$$

where (λ, e) is the eigen-pair we would like to obtain. We then estimate β by

$$\hat{\beta} = [\hat{e}_1, \hat{e}_2, \cdots, \hat{e}_m]$$

where the eigen-pairs are ordered so that $\hat{\lambda}_1 > \hat{\lambda}_2 > \cdots$. We may estimate the other parameters by running the regression

$$\Delta x_t = \mu + \alpha \hat{\beta}' x_{t-1} + \Phi_1 \Delta x_{t-1} + \cdots + \Phi_p \Delta x_{t-p} + \varepsilon_t.$$

To determine m, we consider the following hypothesis

$$H_0 : m = r$$

versus

$$H_1 : m > r$$

and the likelihood ratio statistic

$$-2 \ln Q = -T \sum_{i=r+1}^{N} \ln(1 - \hat{\lambda}_i).$$

Johansen (1988) has shown that under the null,

$$-2 \ln Q \to_d \operatorname{tr} \left\{ \int_0^1 \mathrm{d}W W' \left(\int_0^1 W W' \mathrm{d}r \right)^{-1} \int_0^1 W \mathrm{d}W \right\}$$

where W is an $(N-r)$-dimensional Brownian motion with covariance matrix I. The critical values could be obtained by simulations.