CQF Lecture. Cointegration: Modelling Long-Term Relationships

Workings

Autoregressive System

A system of endogenous variables that depend only on their past (lagged) values. It is also an essential structural equation model of 'seemingly unrelated regressors', can estimated *by row*.

$$\begin{aligned} y_{1,t} &= \beta_{1,0} + \beta_{11} y_{1,t-1} + \beta_{12} y_{2,t-1} + \ldots \beta_{1n} y_{n,t-1} + \ldots_{t-2} \ldots + \epsilon_{1,t} \\ y_{2,t} &= \beta_{2,0} + \beta_{21} y_{1,t-1} + \beta_{22} y_{2,t-1} + \ldots \beta_{2n} y_{n,t-1} + \ldots_{t-2} \ldots + \epsilon_{2,t} \\ \ldots & \quad \ldots \\ y_{n,t} &= \beta_{n,0} + \beta_{n1} y_{1,t-1} + \beta_{nn} y_{2,t-1} + \ldots \beta_{nn} y_{n,t-1} + \ldots_{t-2} \ldots + \epsilon_{n,t} \end{aligned}$$

Matrix Form

The same system of equations can be represented *in matrix form*. The advantage is in the simpler formulation of analytical conditions and solutions.

$$\boldsymbol{Y_t = C + \beta_1 Y_{t-1} + \ldots + \beta_p Y_{t-p} + \epsilon_t}$$

where $\boldsymbol{Y_t} = (y_{1,t}, \ldots, y_{n,t})'$ is $n \times 1$ column **vector**, and $\boldsymbol{\beta_p}$ is $n \times n$ matrix of regression coefficients. Each lag $\boldsymbol{Y_{t-1} \ldots Y_{t-p}}$ will have its own matrix of coefficients $\boldsymbol{\beta_p}$.

$$\boldsymbol{\beta_p} = \begin{bmatrix} \beta_{11}^p & \cdots & \beta_{1n}^p \\ \vdots & \ddots & \vdots \\ \beta_{n1}^p & \cdots & \beta_{nn}^p \end{bmatrix}$$

For example, it is possible to estimate the VAR system of regression equations *in one go*, using the following matrix form solution where $\boldsymbol{Z}$ formed from descending $\boldsymbol{Y}$ with ones in the top row:

$$\begin{aligned} \boldsymbol{Y} &= \boldsymbol{B\,Z + \epsilon} \\ \boldsymbol{\hat{B}} &= \boldsymbol{Y Z' (Z Z')^{-1}} \\ \boldsymbol{\hat{\epsilon}} &= \boldsymbol{Y - \hat{B}\,Z} \end{aligned}$$

The document prepared by CQF Faculty, Dr Richard Diamond at r.diamond@cqf.com

VAR Model Specification

Stability condition for VAR requires for the eigenvalues of *each* relationship matrix $\boldsymbol{\beta_p}$ to be inside the unit circle (< 1).

$$|\lambda \boldsymbol{I} - \boldsymbol{\beta}| = 0$$

Optimal Lag p is determined by the lowest values of AIC, BIC statistics constructed using penalised likelihood. In practice, the choice of optimal lag p is evident from Partial Autocorrelation Function (PACF) diagram: the significant serial autocorrelations will be above confidence limits.

By keeping more lags in Vector Autoregression you condition a model on them, thus making it prone to overfitting. That will transpire with in attempts to use VAR for forecasting, for example, forecasting of market index returns gives an average Mean Absolute Percentage Error (MAPE) $O(200\%)$.

Residuals

- VAR inference is particularly *sensitive to*:
  parameter non-constancy, **serially correlated residuals**, and skewness in residuals.
- The other breaches, such as excess kurtosis and heteroscedasticity in residuals do not invalidate the MLE properties of the linear regression.

The most important part of a regression is... **residuals** (eg, innovations, shocks, disturbances, moving averages). Residuals are not supposed to 'cluster' when observations are extreme, ie, 'from the tail'. Generally, residuals are not supposed to be autocorelated (serially correlated) – a breach in this requirement invalidates the regression model.

$$\epsilon_t \sim IN(0, \sigma_t^2)$$

Since residuals are an *iid* invariant of the model,

- We know their properties: stationary, constant mean and variance, empirically insignificant autocorrelation.
- We can explore a model by stress-testing its residuals (e.g., introducing shocks). The common method to do so is constructing the Impulse Response Function by Cholesky Decomposition of residual covariance matrix $\hat{\boldsymbol{\Sigma}} = \frac{1}{T}\sum_{t=1}^{T} \hat{\epsilon}_t \hat{\epsilon}_t'$.

$$\begin{aligned} \hat{\boldsymbol{\Sigma}} &= \boldsymbol{A}\boldsymbol{A}' \\ \boldsymbol{\epsilon}_t' &= \boldsymbol{A}\boldsymbol{v} \qquad \boldsymbol{v}' = (1, 1, \ldots, 1) \end{aligned}$$

- Significance testing is done on residuals. t-statistic and F-statistic are essentially ratios of RSS.

We study residuals with

- Location-dispersion ellipsoid, Q-Q plot and K-S test as checks for being *iid*.
- Testing for autocorrelation (serial correlation) by LM test.
  Exploring ACF and PACF
- **Testing for stationarity** (absence of the unit root) by ADF and Phillips-Perron tests.

Infinite Divisibility We can represent each residual ϵ_t as an *infinite* sum of the Standard Normal random variables.

$$\epsilon_t = Y_t - Y_{t-1} \stackrel{D}{=} Z_1 + \cdots + Z_k$$

Residuals variance is the sum of squared errors $Z = (\epsilon - \mu)/\sigma, \quad \mu = 0,\ \sigma^2 = 1$

$$\sum \epsilon_t^2 = \sum_{i=1}^{k} Z_i^2 \quad \sim \quad \chi_k^2$$

For a sample, variance is itself a random variable, Chi-square distributed.

Stationary Process

The first-order stationary random process has probability density that is **independent of time**. That means its first-order statistics (moments) are invariant *wrt* time shift: the mean, variance and autocovariance are **constant**.

- $\mathbb{E}[Y_t] = \frac{\beta_0}{1-\beta_1}$.
- $\mathbb{V}\text{ar}[Y_t] = \frac{\sigma^2}{1-\beta_1^2}$. This is unconditional variance σ_0.
- $\text{Cov}[Y_t, Y_s] = \beta_1^{t-s} \times \sigma_0$. Autocorrelation $\times$ Variance.

  where **autocorrelation decays exponentially** with time

  $\text{Corr}[Y_t, Y_s] = \beta_1^{t-s} = e^{(t-s)\ln\beta_1}$

Asymptotics of autoregressive process

- **Stationary**

  $$Y_t \stackrel{D}{\approx} N\left(0, \frac{\sigma^2}{1-\beta_1^2}\right)$$

  A shock to the innovations of stationary time series dies out. The process mean-reverts.

  Example: Think of returns for a global market index. 3σ move can happen but investing in the index is clearly safer.

- **Non-stationary** with stochastic trend

  $$\Delta Y_t \stackrel{D}{=} N\left(0, \sigma^2 t\right)$$

  Non-stationary series accumulate shocks and experience
  permanent shifts in $Y_t = \sum \epsilon_t$. Robust estimation of σ_t poses a further problem.

ARMA Representation

To understand a construction of time series, we test how many AR and MA terms it contains. Financial time series are most likely to have an autoregressive term of *at least* order one.

$$\begin{aligned} Y_t &= \beta_0 + \beta_1 Y_{t-1} + \epsilon_t \\ Y_t - \beta_1 Y_{t-1} &= \beta_0 + \epsilon_t \\ (1 - \beta_1 L) Y_t &= \ldots \qquad \text{where } LY_t \equiv Y_{t-1} \end{aligned}$$

$$\begin{aligned} Y_t &= \frac{\beta_0}{1-\beta_1} + \frac{\epsilon_t}{(1-\beta_1 L)} \\ Y_t &= \frac{\beta_0}{1-\beta_1} + \epsilon_t + \beta_1 \epsilon_{t-1} + \beta_1^2 \epsilon_{t-2} + \cdots \end{aligned}$$

We decomposed an AR(1) process into MA(∞) process. The same decomposition can be represented with the lag operator L and the following expansion rule for a stationary Y_t

$$\begin{aligned} \frac{1}{1-\beta L} &\approx 1 + \beta L + \beta^2 L^2 + \cdots \\ &\approx \sum_{i=0}^{\infty} (\beta L)^i \end{aligned}$$

$$\text{where} \quad |\beta| < 1 \text{ is stationarity condition}$$

This result is known as the Wold's theorem.

Unit Root Maths

Without an intercept, $AR(1)$ process can be re-arranged as a polynomial using lag operator L,

$$Y_t = \beta Y_{t-1} + \epsilon_t$$

$$\begin{aligned} \epsilon_t &= Y_t - \beta Y_{t-1} \\ &= Y_t - \beta L Y_t \\ &= (1 - \beta L) Y_t \end{aligned}$$

where $(1 - \beta L)$ is *a characteristic polynomial* with a root of β^{-1}.

$$\begin{aligned} \epsilon_t &= \Delta Y_t \\ \epsilon_t &\sim \text{IN}(0, \sigma_t^2) \end{aligned}$$

The special unit root case is $\beta = 1$. Then, even if Y_t is non-stationary, $\Delta Y_t = \epsilon_t$ is a stationary process and an *iid* invariant.

Unit Root (Non-Stationarity) Testing

As a first item, one should test time series for stationarity. The most popular test is *Augmented Dickey-Fuller* that relies on p lags of the dependent variable y_t

$$\Delta y_t = \phi y_{t-1} + \sum_{k=1}^{p} \phi_i \Delta y_{t-k} + u_t.$$

Feel free to add a constant and time trend in above specification, but notice that the critical value would have changed. You should associate rejection of the null hypothesis of unit root in favour of alternative hypothesis with the implication that the time series y_t is stationary. The critical value of ADF test are available from most time series textbooks.

If y_t is stationary then autoregression applies, otherwise cointegration analysis should be the preferred method.

Equilibrium Correction Mechanism

A regression model that includes lagged values is called **dynamic**

$$y_t = \alpha y_{t-1} + \beta_0 + \beta_1 x_t + \beta_2 x_{t-1} + \epsilon_t \tag{1}$$

We would like to re-specify into the error correction form (ECM)

$$\Delta y_t = \underbrace{\beta_1 \Delta x_t} - (1-\alpha) \underbrace{(y_{t-1} - a - b x_{t-1})} + \epsilon_t \tag{2}$$

Equations (1) and (2) are equivalent. To show this, we introduce terms y_{t-1}, $\beta_1 x_{t-1}$ and re-arrange:

$$\begin{aligned}
y_t - \underline{y_{t-1}} &= \alpha y_{t-1} - \underline{y_{t-1}} + \beta_0 + \beta_1 x_t + \beta_2 x_{t-1} - \underline{\beta_1 x_{t-1}} + \underline{\beta_1 x_{t-1}} + \epsilon_t \\
\Delta y_t &= -(1-\alpha) y_{t-1} + \beta_0 + \beta_1 \Delta x_t + (\beta_1 + \beta_2) x_{t-1} + \epsilon_t \\
&= \beta_1 \Delta x_t - (1-\alpha) \left[y_{t-1} - \frac{\beta_0}{1-\alpha} - \frac{\beta_1 + \beta_2}{1-\alpha} x_{t-1} \right] + \epsilon_t \\
&= \beta_1 \Delta x_t - (1-\alpha) \left[y_{t-1} - a_e - b_e x_{t-1} \right] + \epsilon_t \qquad (3) \\
&= \beta_1 \Delta x_t - (1-\alpha) \, e_{t-1} + \epsilon_t
\end{aligned}$$

Vector Error Correction (Cointegrated VAR)

If all the data series are non-stationary you might choose to proceed with cointegration analysis instead of VAR. Cointegration analysis is a methodology of structural equation modelling based on Vector Autoregressions. It is used to understand a set of relationships as a whole. This and other issues limit the use of cointegration in forecasting – please see *Learning and Trusting Cointegration in Statistical Arbitrage* that discusses the limits and uncertainty of direct forecasting of Y_t. For example, if that is returns, the forecasting errors can reach several hundred per cent.

Vector Error Correction Model is estimated on differences

$$\Delta Y_t = \Pi Y_{t-p} + \sum_{k=1}^{p-1} \Gamma_k \Delta Y_{t-k} + \epsilon_t$$

where

$$\Pi = I - \sum_{k=1}^{p} A_k, \quad \Gamma_k = \sum_{i=1}^{k} A_i - I.$$

'Optimisation' of matrix Π to the reduced-rank form (a cointegration rank restriction) that reveals the true rank and separates cointegrating relationships from the noise relationships is carried out using the Johansen Procedure. The bivariate cointegration can be estimated without matrix operations, using the two-step Engle-Granger procedure.

Johansen Procedure

Johansen Procedure: Step One

To obtain invariants, regress ΔY_t and Y_{t-1} on the lagged first differences as follows: (p is a number of lags in $\boldsymbol{Y_{t-1} \ldots Y_{t-p}}$)

$$\begin{aligned} \Delta Y_t &= \sum_{k=1}^{p-1} \Gamma_k \Delta Y_{t-k} + R_{0t} \\ Y_{t-1} &= \sum_{k=1}^{p-1} \Gamma_k \Delta Y_{t-k} + R_{kt} \end{aligned}$$

These residuals keep information about cointegration, while other variables being removed from the concentrated model:

$$\begin{aligned} R_{0t} &= \boldsymbol{\alpha\beta'} R_{kt} + u_t \\ \Delta\widetilde{Y}_t &= \boldsymbol{\alpha\beta'} \widetilde{Y}_t + u_t \end{aligned}$$

where u_t is a residual outside of the cointegrating relationships.

Use difference equation residuals R_{0t} and level equation residuals R_{kt} to form the residual product-moment matrices.

$$S_{ij} = \frac{1}{T} \sum_{t=1}^{T} R_{it} R'_{jt}, \quad i, j = [0, k]$$

In econometric notation, this is fully equivalent to

$$\Sigma_{ij} = \frac{1}{T}\sum_{t=1}^{T} \epsilon_{it}\epsilon'_{jt}, \quad i,j = [0,k]$$

$$\mathbf{S_{00}} = \frac{1}{T}\mathbf{R_{0t}}\mathbf{R'_{0t}} \quad \mathbf{S_{kk}} = \frac{1}{T}\mathbf{R_{kt}}\mathbf{R'_{kt}} \quad \mathbf{S_{k0}} = \frac{1}{T}\mathbf{R_{kt}}\mathbf{R'_{0t}} \quad \mathbf{S_{0k}} = \frac{1}{T}\mathbf{R_{0t}}\mathbf{R'_{kt}}$$

Each residual matrix R will have dimensions of $n \times T$ where $T \equiv N_{obs}$. Each product-moment matrix will have dimensions of $n \times n$.

Johansen Procedure: Step Two

Maximisation of the concentrated likelihood $\max_{\boldsymbol{\beta}} L(\boldsymbol{\beta}) = |\hat{\boldsymbol{\Sigma}}|^{-\frac{T}{2}}$ becomes an eigenvalue problem:

$$\begin{aligned}
(S_{k0}S_{00}^{-1}S_{0k} - \lambda S_{kk})\beta &= 0 \qquad \text{see Guide to Johansen MLE for a proof} \\
|\lambda S_{kk} - S_{k0}S_{00}^{-1}S_{0k}| &= 0 \qquad \text{another form} \\
\underbrace{S_{kk}^{-1}S_{k0}S_{00}^{-1}S_{0k}}\,\beta_i &= \lambda_i\beta_i
\end{aligned}$$

To solve the problem, we decompose the underlined matrix term that should have $n \times n$ dimensions. Eigenvalues $\lambda_1 > \ldots > \lambda_n$ provide squared canonical correlations for ΔY_t and Y_t.

Eigenvectors $\beta = [\beta_1, \beta_2, \ldots, \beta_n]$ might have cointegrating combinations as determined by tests carried out at Step 3. It is sensible to normalise eigenvectors around the covariance matrix:

$$\widetilde{\beta}' S_{kk} \widetilde{\beta} = I$$

which is achieved by decomposing and post-multiplying

$$\beta' S_{kk} \beta = A'A \qquad \widetilde{\beta} = \beta A^{-1}.$$

Johansen Procedure: Step Three

Since eigenvalues $\propto$ correlations, it make sense to select cointegrating vector(s) that correspond to the largest eigenvalue(s), which are above certain critical levels.

The reduced-rank form of $\boldsymbol{\beta}$ contains selected eigenvectors for r largest eigenvalues as determined by the either of the rank tests:

- Trace statistic (LR test)

$$\lambda_{trace} = -T\sum_{i=r+1}^{n} \log(1-\lambda_i)$$

- Maximum eigenvalue statistic

$$\lambda_{max} = -T\log(1-\lambda_{r+1})$$

While $H_0 = r$ is the same for both tests, the alternative hypothesis is different for each test: $H_1 = n$ and $H_1 = r+1$ respectively. The critical values are available in research literature.

Johansen Procedure: Step Four (Final)

MLE value for the adjustment coefficient (error correction speed) can be estimated as

$$\hat{\boldsymbol{\alpha}} = \widetilde{\boldsymbol{\beta}}'_{\boldsymbol{Coint}} \boldsymbol{S_{k0}}$$

.

Estimate the remaining parameters $\mathbf{\Gamma}$ and $\mathbf{\Psi D_t}$ by OLS

$$\Delta \boldsymbol{Y}_t = \hat{\boldsymbol{\alpha}} \widetilde{\boldsymbol{\beta}}' \boldsymbol{Y_{t-1}} + \sum_{k=1}^{p-1} \mathbf{\Gamma}_k \Delta \boldsymbol{Y}_{t-k} + \boldsymbol{\epsilon}_t$$

We can construct *a cointegrated VAR* forecast for h periods using the full rank of matrix $\widehat{\mathbf{\Pi}}$

$$\widehat{\boldsymbol{Y}}_{\boldsymbol{T+h}} = \widehat{\mathbf{\Pi}}^{\boldsymbol{h}} \boldsymbol{Y_T} + \sum_{\boldsymbol{i=0}}^{\boldsymbol{h-1}} \widehat{\mathbf{\Pi}}^{\boldsymbol{i}} \boldsymbol{C}$$

The reduced rank (with cointegrating relationships only) can be used to amend the usual VAR forecast with the $\widehat{\mathbf{\Pi}}_{\boldsymbol{MLE}} \boldsymbol{Y_T}$ term.