INTRODUCTION TO MULTIVARIATE TIME SERIES

JEN-WEN LIN, PHD, CFA

29 October, 2018

VECTOR AUTOREGRESSION

Introduction

- Vector autoregressions (VAR) is a generalization of the autoregressive and moving average (ARMA) process.
- Christopher A. Sims (1980) introduced VAR to econometricians and won the Nobel prize in 2011.
- *The academy said he had developed a method based on "vector autoregression" to analyse how the economy is affected by temporary changes in economic policy and other factors - for instance, the effects of an increase in the interest rate set by a central bank.--BBC News Website.*

Why use VAR

- A concise way of summarizing interrelationship among data.
- Good forecast results.
- Testing for Granger causality among time series
- Finance applications: strategic asset allocation (Campbell et al 2003), variance decomposition of excess stock returns (Campbell 1991) and long-horizon stock return predictability (Campbell and Shiller, 1989)

Vector autoregression of order one

- The VAR(1)-process of K endogenous variables is defined as

$$\boldsymbol{y}_t = \boldsymbol{A}\boldsymbol{y}_{t-1} + \boldsymbol{u}_t,$$

where $\boldsymbol{y}_t = (y_{1t}, \ldots, y_{jt}, \ldots, y_{Kt})'$ for $j = 1, \ldots K$, $\boldsymbol{A}$ are (K × K) coefficient matrix, and $\boldsymbol{u}_t$ is a K-dimensional white noise process with time-invariant positive definite covariance matrix $E(\boldsymbol{u}_t\boldsymbol{u}'_t) = \Sigma_u$.

- By repeated substitution, the above VAR(1) model becomes

$$\boldsymbol{y}_t = \boldsymbol{u}_t + \boldsymbol{A}\boldsymbol{u}_{t-1} + \boldsymbol{A}^2\boldsymbol{u}_{t-2} + \boldsymbol{A}^3\boldsymbol{u}_{t-3} + \cdots$$

- For the above VMA(infinity) process to be stationary, $\boldsymbol{A}^j$ must converge to zero as j goes to infinity. Mathematically, we require that all K eigenvalues of $\boldsymbol{A}$ be less than one in modulus.

Vector autoregression of order p

- The VAR(p)-process of K endogenous variables is defined as

$$\boldsymbol{y}_t = \boldsymbol{A}_0 + \boldsymbol{A}_1 \boldsymbol{y}_{t-1} + \ldots + \boldsymbol{A}_p \boldsymbol{y}_{t-p} + \boldsymbol{u}_t, (1)$$

where $\boldsymbol{y}_t = (y_{1t}, \ldots, y_{kt}, \ldots, y_{Kt})$ for k = 1, . . .K, $\boldsymbol{A}_0$ stands for a (K x 1) mean vector, $\boldsymbol{A}_i$ are (K × K) coefficient matrices for i = 1, . . . , p and $\boldsymbol{u}_t$ is a K-dimensional white noise process with time-invariant positive definite covariance matrix $E(\boldsymbol{u}_t \boldsymbol{u}'_t) = \Sigma_u$.

– *We could include deterministic regressors, such as a constant, trend, and seasonal dummy variables, in eqn. (1)*

$$\begin{bmatrix} y_1 \\ y_2 \end{bmatrix}_t = \begin{bmatrix} 5.0 \\ 10.0 \end{bmatrix} + \begin{bmatrix} 0.5 & 0.2 \\ -0.2 & -0.5 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \end{bmatrix}_{t-1} + \begin{bmatrix} -0.3 & -0.7 \\ -0.1 & 0.3 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \end{bmatrix}_{t-2} + \begin{bmatrix} u_1 \\ u_2 \end{bmatrix}_t .$$

Vector autoregression of order p

- Equation (1) may be written in the compact form

$$A(B)\boldsymbol{y}_t = \boldsymbol{u}_t, (2)$$

  where $A(B) = (\boldsymbol{I}_K - \boldsymbol{A}_1 B - \ldots - \boldsymbol{A}_p B^p)$.

- One important question to ask for a VAR(p)-process is how to check its stationarity (stability).

- The necessary and sufficient condition for the stationarity of $\boldsymbol{y}_t$ is that all solutions (roots) of $\det(\boldsymbol{I}_K - \boldsymbol{A}_1 B - \ldots - \boldsymbol{A}_p B^p) = 0$ are greater than one in absolute value.

The companion form of VAR(p) process

- In practice, the stability of a VAR(p) model is analyzed via its companion form.
- The companion form of a VAR(p)-process is given by

$$\boldsymbol{\xi}_t = \boldsymbol{A}\boldsymbol{\xi}_{t-1} + \boldsymbol{v}_t, \qquad (3)$$

with

$$\xi_t = \begin{bmatrix} \mathbf{y}_t \\ \vdots \\ \mathbf{y}_{t-p+1} \end{bmatrix}, \quad A = \begin{bmatrix} A_1 & A_2 & \cdots & A_{p-1} & A_p \\ I & 0 & \cdots & 0 & 0 \\ 0 & I & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & I & 0 \end{bmatrix}, \quad \mathbf{v}_t = \begin{bmatrix} \mathbf{u}_t \\ \mathbf{0} \\ \vdots \\ \mathbf{0} \end{bmatrix},$$

where the dimension of the stacked vectors $\boldsymbol{\xi}_t$ and $\boldsymbol{v}_t$ is ($Kp \times 1$) and that of the matrix $\boldsymbol{A}$ is ($Kp \times Kp$). The companion form of a VAR(p) process is also a VAR(1) process.

- If the moduli of the *eigenvalues* of $\boldsymbol{A}$ are less than one, then the VAR(p)-process is stable/stationary.

Example

$$\begin{bmatrix} y_1 \\ y_2 \end{bmatrix}_t = \begin{bmatrix} 5.0 \\ 10.0 \end{bmatrix} + \begin{bmatrix} 0.5 & 0.2 \\ -0.2 & -0.5 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \end{bmatrix}_{t-1} + \begin{bmatrix} -0.3 & -0.7 \\ -0.1 & 0.3 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \end{bmatrix}_{t-2} + \begin{bmatrix} u_1 \\ u_2 \end{bmatrix}_t .$$

$$A = \begin{bmatrix} A_1 & A_2 & \cdots & A_{p-1} & A_p \\ I & 0 & \cdots & 0 & 0 \\ 0 & I & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & I & 0 \end{bmatrix}, \quad A = \begin{bmatrix} 0.5 & 0.2 & -0.3 & -0.7 \\ -0.2 & -0.5 & -0.1 & 0.3 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix}$$

```
> A<-matrix(c(0.5,0.2,-0.3,-0.7,-0.2,-0.5,-0.1 , 0.3 , 1,0 ,0 ,0 ,0 ,1,0,0),nrow=4, byrow = TRUE)
> A
     [,1] [,2] [,3] [,4]
[1,]  0.5  0.2 -0.3 -0.7
[2,] -0.2 -0.5 -0.1  0.3
[3,]  1.0  0.0  0.0  0.0
[4,]  0.0  1.0  0.0  0.0
```

```
> ev<-eigen(A,only.value=TRUE)$values
> ev[1]
[1] -0.8180175+0i
> ev[2]
[1] 0.5974589+0i
> sqrt(Re(ev[3])^2+Im(ev[3])^2)
[1] 0.5721695
> sqrt(Re(ev[4])^2+Im(ev[4])^2)
[1] 0.5721695
```

Estimating VAR(p) process

- For a given sample of the endogenous variables $\boldsymbol{y}_1, \ldots \boldsymbol{y}_T$ and sufficient presample values $\boldsymbol{y}_{-\boldsymbol{p}+\boldsymbol{1}}, \ldots, \boldsymbol{y}_0$, the coefficients of a VAR(p)-process can be estimated efficiently by least squares applied separately to each of the equations.
- If the error process $\boldsymbol{u}_t$ is normally distributed, then this estimator is equal to the maximum likelihood estimator conditional on the initial values.

Test model adequacy
Portmanteau tests

Test statistics:

$$Q_{BP} = T\sum_{j=1}^{m} \mathrm{tr}(\hat{C}_j^T \hat{C}_0^{-1} \hat{C}_j \hat{C}_0^{-1}) \sim \chi^2_{k^2 m - n^*}$$

$$Q_{LB} = T^2\sum_{j=1}^{m} \frac{1}{T\text{-}j} \mathrm{tr}(\hat{C}_j^T \hat{C}_0^{-1} \hat{C}_j \hat{C}_0^{-1}) \sim \chi^2_{k^2 m - n^*}$$

where n^* is the number of coefficients excluding deterministic terms of a $VAR(p)$ model and

$$\hat{C}_i = \frac{1}{T}\sum_{t=i+1}^{T} \hat{\mathbf{a}}_t \hat{\mathbf{a}}_{t-i}^T$$

Order selection and curse of dimensionality

- Order selection
  - *Sequential likelihood ratio tests*
    - Based on the likelihood ratio test for testing VAR(p) versus VAR(p-1)
  - *Information criteria*
- BigVAR: Dimension Reduction Methods for Multivariate Time Series

GRANGER CAUSALITY

- *Causality tests are useful to infer whether a variable helps predict another one.*
- *An operational definition of causality between two time series can be defined in terms of predictability (Granger, 1969).*

Granger causality

- Ideally, causality may be defined through the concept of the conditional distribution.
  - y_{2t} does not cause y_{1t} if the distribution of y_{1t}, conditional on past values of both y_{1t} and y_{2t}, is the same as the distribution of y_{1t} conditional on its own past values.
- In practice, it would be very difficult to test whether the entire distribution y_{1t} depends on past values of y_{2t}.
- Therefore, we consider an alternative by asking whether the conditional mean of y_{1t} depends on past values of y_{2t}. If this is the case, we can test causality by imposing restrictions on a VAR model.

Granger causality

- Consider a VAR(p) model as follows:

$$\begin{bmatrix} y_{1t} \\ y_{2t} \end{bmatrix} = \begin{bmatrix} \alpha_1 \\ \alpha_2 \end{bmatrix} + \sum_{j=1}^{p} \begin{bmatrix} \phi_{j,11} & \phi_{j,12} \\ \phi_{j,21} & \phi_{j,22} \end{bmatrix} \begin{bmatrix} y_{1,t-j} \\ y_{2,t-j} \end{bmatrix} + \begin{bmatrix} a_{1t} \\ a_{2t} \end{bmatrix}$$

- If y_{2t} does not Granger cause y_{1t}, then all of the $\phi_{j,12}$'s must be zero. Note that $\phi_{j,12}$'s only appear in the equation for y_{1t}.

- Similarly, if y_{1t} does not Granger cause y_{2t}, then all of the $\phi_{j,21}$'s must be zero.

Likelihood Ratio Test

- Obtain ML (or OLS) estimates of the following equations.

$$y_{1t} = \alpha_1 + \sum_{j=1}^{p} \phi_{j,11} y_{1,t-j} + e_{1t}, \qquad (2)$$

$$y_{1t} = \alpha_1 + \sum_{j=1}^{p} \phi_{j,11} y_{1,t-j} + \sum_{j=1}^{p} \phi_{j,12} y_{2,t-j} + \varepsilon_{1t}, (3)$$

- Calculate the values of the log likelihood functions in eqn. (2) and (3). And, the LR statistic is given by

$$n(\log|\tilde{\Sigma}| - \log|\hat{\Sigma}|) \sim \chi_p^2$$

where $\tilde{\Sigma}$ and $\hat{\Sigma}$ denote the residual covariance matrix estimated from eqn. (2) and (3), respectively.

Portmanteau test For Granger Causality

- Pierce and Haugh (1977) expanded up the work of Granger (1969) and gave a comprehensive survey regarding research on causality in temporal systems.
- For simplicity, in what follows, we consider the case of two time series $\{X_t\}$ and $\{Y_t\}$.
- Let $\{X_t\}$ and $\{Y_t\}$ be causal and invertible univariate $ARMA$ processes and be given by

$$\phi_X(B)(X_t - \mu_X) = \theta_X(B)u_t, \quad u_t \sim \mathrm{WN}(0, \sigma_u^2)$$
$$\phi_Y(B)(Y_t - \mu_Y) = \theta_Y(B)v_t, \quad v_t \sim \mathrm{WN}(0, \sigma_v^2)$$

Portmanteau test For Granger Causality

The cross-correlation function at lag k between u_t and v_t series is given by

$$\rho_{uv}(k) = \frac{E(u_t, v_{t+k})}{\sqrt{E(u_t^2)E(v_t^2)}}$$

Pierce and Haugh (1977) explained that there are many possible types of causal interpretation between $\{X_t\}$ and $\{Y_t\}$ which can be characterized by the properties of $\rho_{uv}(k)$.

Portmanteau test For Granger Causality

RELATIONSHIPS	RESTRICTIONS ON $\rho_{uv}(k)$
X causes Y	$\rho_{uv}(k) \neq 0$ for *some* $k > 0$
Y causes X	$\rho_{uv}(k) \neq 0$ for *some* $k < 0$
Instantaneous Causality	$\rho_{uv}(0) \neq 0$
Feedback	$\rho_{uv}(k) \neq 0$ for *some* $k > 0$ and for some $k < 0$
X causes Y but not instantaneously	$\rho_{uv}(k) \neq 0$ for *some* $k > 0$ and $\rho_{uv}(0) = 0$
Y does not cause X	$\rho_{uv}(k) = 0$ for *all* $k < 0$
Y does not cause X at all	$\rho_{uv}(k) = 0$ for *all* $k \leq 0$
Unidirectional causality from X to Y	$\rho_{uv}(k) \neq 0$ for some $k > 0$ and $\rho_{uv}(k) = 0$ for either (a) all $k < 0$ or (b) all $k \leq 0$
X and Y are only related instantaneously	$\rho_{uv}(0) \neq 0$ and $\rho_{uv}(k) = 0$ for all $k \neq 0$
X and Y are independent	$\rho_{uv}(k) = 0$ for *all* k

Portmanteau tests for Granger causality

- $H_0 : X$ does not cause Y
- $Q_L = n^2 \sum_{k=0}^{L} (n-k)^{-1} r_{uv}^2(k) \sim \chi_{L+1}^2$

COINTEGRATION

Models for multiple integrated process

Review I(d) process

- In econometrics, a time series z_t is said to be an integrated process of order one, that is, an $I(1)$ process, if $(1-B)z_t$ is stationary and invertible.
  - *A stationary and invertible time series is said to be an $I(0)$ process.*
- In general, a univariate time series z_t is an $I(d)$ process if $(1-B)^d z_t$ is stationary and invertible, where $d > 0$ and order d is referred to as the order of integration or the multiplicity of a unit root.

Motivation of cointegration

- It is incorrect to analyze nonstationary time series using standard statistical inference techniques.
- We've learned that the Box-Jenkins approach uses differencing to solve the problem.

– *Cointegration is another technique to model nonstationary (multivariate) time series.*

– *What is the intuition behind cointegration?*

1. *Balance of the (linear) regression equation*
2. *If time series share the same source of the $I(1)$'ness, or time series move together in the long-run.*

Cointegration

– *Consider a multivariate time series $\boldsymbol{z}_t$. If z_{it} $\forall i$ are $I(1)$ processes but a nontrivial linear combination $\boldsymbol{\beta}'\boldsymbol{z}_t$ is $I(0)$, then $\boldsymbol{z}_t$ is said to be cointegrated of order one.*

■ The linear combination vector $\boldsymbol{\beta}$ is called a cointegrating vector.

– *In general, if z_{it} are $I(d)$ nonstationary and $\boldsymbol{\beta}'\boldsymbol{z}_t$ is $I(h)$ with $h < d$, then $\boldsymbol{z}_t$ is cointegrated. In practice, the case of $d = 1$ and $h = 0$ is of major interest.*

– *Thus, cointegration often means that a linear combination of individually unit-root nonstationary time series becomes a stationary and invertible series.*

USEFUL RESULTS for the linear combination of stochastic process

Linear combinations of $I(0)$ and $I(1)$ processes

1. $X_t \to I(0) \Rightarrow a + bX_t \to I(0)$

   $X_t \to I(1) \Rightarrow a + bX_t \to I(1)$

2. $X_t, Y_t \to I(0) \Rightarrow aX_t + bY_t \to I(0)$

3. $X_t \to I(0), Y_t \to I(1) \Rightarrow aX_t + bY_t \to I(1)$

4. $X_t, Y_t \to I(1) \Rightarrow aX_t + bY_t \to I(1),\ \text{in general}$

COMMON TRENDS

- The idea of Stock and Watson (1988) provides a very useful way to understand cointegration relationships.
  - *Cointegrated variables sharing common stochastic trends*
- A naive example: X_t and Y_t are $I(1)$ processes and satisfy:

$$X_t \equiv \alpha \cdot W_t + \widetilde{X}_t$$

$$Y_t \equiv W_t + \widetilde{Y}_t$$

Stationary ARMA(p,q) process, or I(0) process

X_t and Y_t share the same nonstationary sources W_t--an ARIMA(p,1,q) process, or I(1) process

COMMON TRENDS

- X_t and Y_t have a common $I(1)$ trend, W_t.
- Consider a linear combination Z_t as follows:

$$Z_t \equiv X_t - \alpha \cdot Y_t = \cancel{\alpha \cdot W_t} \boxed{+ \tilde{X}_t} \cancel{- \alpha \cdot W_t} \boxed{- \alpha \cdot \tilde{Y}_t}$$

$$Z_t \equiv \tilde{X}_t - \alpha \cdot \tilde{Y}_t \rightarrow I(0) \quad \text{(rule 2)}$$

If two $I(1)$ process have a common $I(1)$ trend (factor) and $I(0)$ idiosyncratic components, then they are cointegrated.

In the case, we say that $(1,-\alpha)$ as the cointegrating vector.

MORE COMMON TRENDS

Example 1.

$$\left.\begin{aligned} Y_t &\equiv W_t + u_t \\ X_t &\equiv W_t + v_t \\ Z_t &\equiv W_t + s_t \end{aligned}\right\} W_t \to I(1) \quad u_t, v_t, s_t \to I(0)$$

1 common stochastic trend $\to W_t$

2 cointegrating vectors : $(1 \quad -1 \quad 0)' \quad (0 \quad 1 \quad -1)'$

Example 2.

$$\left.\begin{aligned} Y_t &\equiv W_t + u_t \\ X_t &\equiv W_t + R_t + v_t \\ Z_t &\equiv R_t + s_t \end{aligned}\right\} W_t, R_t \to I(1) \quad u_t, v_t, s_t \to I(0)$$

2 common stochastic trends $\to W_t, R_t$

1 cointegrating vector : $(1 \quad -1 \quad 1)'$

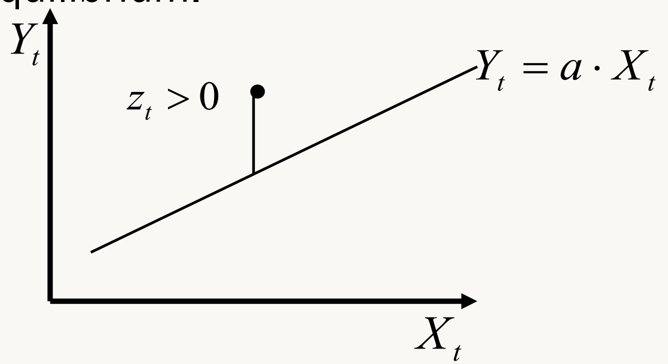
What conclusion can we draw from the above examples

Error correction model

- Let $z_t = Y_t - aX_t$ denote the deviation from the long-run equilibrium.

- If the system is going to return to long-run equilibrium, the short-run movements of the variables (at least some of them) must be respond to the magnitude of disequilibrium.

- Hence, the path of a cointegrated system is influenced by the extend of deviation from the long-run equilibrium.



Error Correction Model

- Consider the example in "Applied Econometric Time Series"

$$\Delta r_{S,t} = a_{10} + \alpha_s (r_{L,t-1} - \beta \cdot r_{S,t-1}) + \sum a_{11}(i) \Delta r_{s,t-i} + \sum a_{12}(i) \Delta r_{L,t-i} + \varepsilon_{S,t}$$
$$\Delta r_{L,t} = a_{20} - \alpha_L (r_{L,t-1} - \beta \cdot r_{S,t-1}) + \sum a_{21}(i) \Delta r_{s,t-i} + \sum a_{22}(i) \Delta r_{L,t-i} + \varepsilon_{L,t}$$
$$\alpha_s, \alpha_L > 0, \quad \varepsilon_{i,t} \sim \mathrm{WN}(0, \sigma_i^2), \quad i = s, L$$

- This two variable error-correction model is a bivariate VAR in first differences augmented by the error-correction terms. Need to understand the following concepts
  1. *Speed of adjustment parameters*
  2. *Granger representation theorem*
  3. *Co-integration coefficient restrictions in a VAR model*

Granger representation theorem

Granger Representation Theorem: If X_t and Y_t are co-integrated, then there exists an ECM representation. Co-integration is a necessary condition for ECM and vice versa.

1. Vector autoregressions on differenced $I(1)$ processes will be a misspecification if the component series are cointegrated.

2. Engle and Granger (1987) showed that an equilibrium specification is missing from a VAR representation.

3. However, when lagged disequilibrium terms are included as explanatory variables, the model becomes well specified.

4. Such a model is called an error correction model (ECM) because the model is structured so that short-run deviation from the long-run equilibrium will be corrected.

The procedure of Engle and Granger (1987)

1) Test wherer X_t and Y_t are $I(1)$ using a unit root test.
2) If both series are $I(1)$, regress one series against the other using least squares.
3) Run a unit root test on regression residuals. If residuals are stationary, these two series are cointegrated.

- The regression line indicates the long-run equilibrium relationship between two variables. The disequilibrium term is simply the regression residuals.

4) Finally, we consider the following ECM

$$\Delta X_t = c_1 + \rho_1(Y_{t-1} - \hat{\alpha} X_{t-1}) + \beta_{x1}\Delta X_{t-1} + \ldots.. + \beta_{y1}\Delta Y_{t-1} + \ldots. + \varepsilon_{xt}$$
$$\Delta Y_t = c_2 + \rho_2(Y_{t-1} - \hat{\alpha} X_{t-1}) + \gamma_{x1}\Delta X_{t-1} + \ldots.. + \gamma_{y1}\Delta Y_{t-1} + \ldots. + \varepsilon_{yt}$$

WHY ENGLE-GRANGER METHOD

- It is very straightforward to implement and to interpret the Engle-Granger procedure.
- From the risk management point of view, the Engle-Granger criterion that minimizes variance is usually more important than the Johansen criterion that maximizes stationarity.
- Sometimes there is a natural choice of dependent variables in the cointegrating regressions, for example, in equity index tracking.

REMARKS

- What's the assumption implicitly imposed in this approach?
  - *The Engle-Granger procedure is only applicable to systems with more than two variables in a very special circumstances.--Carol Alexander (2001)*
- Question: Is there another way to test (model) co-integration?
  - *The Johansen procedure (1988) seeks the linear combination which is most stationary whereas the Engle-Granger tests seek the linear combination having minimum variance.*
  - *The Johansen tests are a multivariate generalization of the unit root tests.*
- The presence of change points will affect the effectiveness of cointegration analysis.