Time Series Analysis

3. Multivariate time series models

Andrew Lesniewski

Baruch College
New York

Fall 2018

Outline

Multivariate time series analysis

- Multivariate time series analysis seeks to analyze several time series simultaneously.
- The rational behind this is the presence of interdependences between the different time series.
- These interdependencies, when quantified appropriately, should lead to improved reliability of forecasts.
- For example, it is reasonable to expect various dependencies when considered jointly the following time series:
  - (i) (a) US GDP data, (b) S&P 500 data, and (c) oil prices.
  - (ii) (a) Mortgage applications, (b) interest rate data, and (c) unemployment rates.
  - (iii) (a) order flow data, (b) price levels, and (c) volatilities
  - (iv) ...

Stationarity

- Consider a time series X_t, $-\infty < t < \infty$ of an n-dimensional state variable $X_t \in \mathbb{R}^n$. We represent X_t as a column vector.
- For modeling purposes, we assume that each X_t is a random variable on a probability space, which is measurable with respect to the information set generated by all X_s with $s \leq t$.
- A time series is called covariance-stationary if:
  - (i) The mean $\mathrm{E}(X_t) = \mu \in \mathbb{R}^n$ is a constant vector.
  - (i) The autocovariance $\mathrm{Cov}(X_t, X_s) = \mathrm{E}\big((X_t - \mu)(X_s - \mu)^{\mathrm{T}}\big) \in \mathrm{Mat}_n(\mathbb{R})$ is time translation-invariant,

$$\mathrm{Cov}(X_{t+\tau}, X_{s+\tau}) = \mathrm{Cov}(X_t, X_s), \tag{1}$$

  for all τ. In particular, $\mathrm{Var}(X_t)$ is a constant positive semidefinite matrix.
- In the following, we will use the notation $\Gamma_{t-s} = \mathrm{Cov}(X_t, X_s)$. By R_{t-s} we will denote the autocorrelation matrix.

Stationarity

- We claim that, as a consequence of stationarity, the following symmetry relation holds:

$$\Gamma_t^{\mathrm{T}} = \Gamma_{-t}, \tag{2}$$

  which is a multivariate extension of the univariate property $\Gamma_t = \Gamma_{-t}$.
- The proof goes as follows:

$$\begin{aligned}
\Gamma_{t,ab}^{\mathrm{T}} &= \Gamma_{t,ba} \\
&= \mathrm{Cov}(X_{t,b}, X_{0,a}) \\
&= \mathrm{Cov}(X_{0,b}, X_{-t,a}) \qquad \text{(time shift by } -t\text{)} \\
&= \mathrm{Cov}(X_{-t,a}, X_{0,b}) \\
&= \Gamma_{-t,ab}.
\end{aligned}$$

- Equation (2) implies the following symmetry property of the autocorrelation matrix:

$$R_t^{\mathrm{T}} = R_{-t}. \tag{3}$$

VAR(1) model

- A *vector autoregressive model* *VAR*(1) is a linear time series model given by

$$X_t = a + BX_{t-1} + \varepsilon_t, \tag{4}$$

where $a \in \mathbb{R}^n$ is a constant vector, and $B \in \mathrm{Mat}_n(\mathbb{R})$ is a constant matrix. The disturbances $\varepsilon \in \mathbb{R}^n$ are i.i.d. random vectors with $\varepsilon \sim N(0, \Omega)$, where $\Omega \in \mathrm{Mat}_n(\mathbb{R})$ is a fixed (positive definite) covariance matrix,

$$\Omega = \mathrm{E}(\varepsilon_t \varepsilon_t^{\mathrm{T}}). \tag{5}$$

- We will now find the conditions under which the process (4) is covariance stationary.
- Taking expected value of (4) yields

$$(I - B)\mu = a,$$

where I is the $n \times n$ identity matrix.

$VAR(1)$ model

- This has a solution (for any a), if and only if 1 is not an eigenvalue of B, or

$$\det(I - B) \neq 0. \tag{6}$$

- Then

$$\mu = (I - B)^{-1} a. \tag{7}$$

- This is the first condition for covariance stationarity. In order to formulate the second condition, we will study the autocovariance of X_t.

$VAR(1)$ model

- To this end, we use (7) to write (4) as

$$X_t - \mu = B(X_{t-1} - \mu) + \varepsilon_t. \tag{8}$$

- Next, multiply (8) on the right by $(X_t - \mu)^{\mathrm{T}}$ and take expectation. This yields the following equation for Γ_0:

$$\Gamma_0 = B\Gamma_0 B^{\mathrm{T}} + \Omega. \tag{9}$$

- Similarly, multiplying (8) on the right by $(X_{t-k} - \mu)^{\mathrm{T}}$ and taking expectation, we obtain the equation:

$$\Gamma_k = B\Gamma_{k-1}. \tag{10}$$

- For a covariance stationary process, equation (9) has to have a solution. Unlike the univariate case, we cannot in general solve it in a closed form (the exception is when the matrices Γ_0 and B commute, $\Gamma_0 B = B\Gamma_0$).

VAR(1) model

- We can, nevertheless, draw from it some useful conclusions about covariance stationarity of the process (4). Let λ be an eigenvalue of B^{T},

$$B^{\mathrm{T}} v = \lambda v, \text{ with } v \neq 0.$$

- Multiplying (9) on the left by v^{T} and on the right by v, we find that

$$v^{\mathrm{T}} \Gamma_0 v = \lambda^2 v^{\mathrm{T}} \Gamma_0 v + v^{\mathrm{T}} \Omega v,$$

or

$$(1 - \lambda^2) v^{\mathrm{T}} \Gamma_0 v = v^{\mathrm{T}} \Omega v.$$

- Since both matrices Ω and Γ_0 are positive definite, this is possible only if $|\lambda| < 1$. As a consequence, a necessary condition for the stationarity of the process is that all eigenvalues of B^{T} have absolute values less than 1.

VAR(1) model

- This condition can be rephrased in a more familiar form as follows. Consider the matrix-valued (linear) polynomial $\psi(z) = I - Bz$, where I is the n-dimensional identity matrix.
- Then the process is stationary if all the roots $z_1, \ldots, z_n$ of the determinant $\det(\psi(z))$, or
$$\det(I - Bz) = 0,$$
lie outside of the unit circle.
- This is clear, as the roots of this equation are given by $1/\lambda_1, \ldots, 1/\lambda_n$, where λ_i are the (nonzero) eigenvalues of B.
- Equations (9) and (10) form the *Yule-Walker equations* for a *VAR*(1) time series model.

$VAR(1)$ model

- As an example, consider the following $VAR(1)$ model:

$$\begin{aligned} X_{t,1} &= 1.1 + 0.6X_{t,1} + 1.5X_{t,2} + \varepsilon_{t,1} \\ X_{t,2} &= 2.0 + 0.1X_{t,1} + 0.4X_{t,2} + \varepsilon_{t,2} \end{aligned} \tag{11}$$

  with

$$\Omega = \begin{pmatrix} 1.0 & 0.4 \\ 0.4 & 1.0 \end{pmatrix}.$$

- Then

$$B = \begin{pmatrix} 0.6 & 1.5 \\ 0.1 & 0.4 \end{pmatrix}.$$

  and the eigenvalues are 0.9 and 0.1, and the process (11) is covariance stationary.
- Notice that this is true (the eigenvalues have absolute values less than 1), even though the coefficient $B_{12} = 1.5$ is greater than 1!.

$VAR(1)$ model

- Let us go back to equation (9). Its solution can be written as an infinite series.
- Namely, iterating the equation, we find that

$$\begin{aligned}\Gamma_0 &= B\Gamma_0 B^{\mathrm{T}} + \Omega \\ &= B^2\Gamma_0(B^2)^{\mathrm{T}} + B\Omega B^{\mathrm{T}} + \Omega \\ &\dots \\ &= \sum_{j=0}^{\infty} B^j \Omega (B^j)^{\mathrm{T}}.\end{aligned} \tag{12}$$

- One can show that, if all the eigenvalues have absolute values less 1, this series converges.
- This and equation (10) also imply that, for all $k \geq 0$,

$$\Gamma_k = \sum_{j=0}^{\infty} B^{j+k} \Omega (B^j)^{\mathrm{T}}. \tag{13}$$

Forecasting with $VAR(1)$

- In order to forecast a multivariate time series X_t we minimize the expected MSE

$$\mathrm{E}_t\left((X_{t+k} - X^*_{t+k|1:t})^{\mathrm{T}}(X_{t+k} - X^*_{t+k|1:t})\right) = \sum_{a=1}^{n} \mathrm{E}_t\left((X_{t+k,a} - X^*_{t+k|1:t,a})^2\right). \quad (14)$$

- Extending the reasoning of Lecture Notes #1, we find the optimal prediction is the conditional expectation

$$X^*_{t+k|1:t} = E_t(X_{t+k}). \quad (15)$$

- This formula, applied to the one-period forecast in a $VAR(1)$ yields

$$\begin{aligned} X^*_{t+1|1:t} &= \mathrm{E}_t(X_{t+1}) \\ &= \mathrm{E}_t(a + BX_t + \varepsilon_{t+1}) \\ &= a + BX_t. \end{aligned} \quad (16)$$

- The forecast error is ε_{t+1}, and so the variance of the forecast error is Ω.

Forecasting with $VAR(1)$

- In order to produce a k-period forecast with a $VAR(1)$ model, we iterate:

$$\begin{aligned} X_{t+k} &= a + BX_{t+k-1} + \varepsilon_{t+k} \\ &= (I+B)a + B^2 X_{t+k-2} + \varepsilon_{t+k} + B\varepsilon_{t+k-1} \\ &\dots \\ &= (I + B + \ldots + B^{k-1})a + B^k X_t + \varepsilon_{t+k} + B\varepsilon_{t+k-1} + \ldots + B^{k-1}\varepsilon_{t+1}. \end{aligned}$$

- This yields the optimal prediction:

$$X^*_{t+k|1:t} = (I + B + \ldots + B^{k-1})a + B^k X_t. \tag{17}$$

- The forecast error is

$$\varepsilon_{t+k} + B\varepsilon_{t+k-1} + \ldots + B^{k-1}\varepsilon_{t+1},$$

and so the forecast variance is

$$\Omega + B\Omega B^{\mathrm{T}} + \ldots + B^{k-1}\Omega(B^{k-1})^{\mathrm{T}}.$$

$VAR(p)$ model

- A $VAR(p)$ model is an extension of the $VAR(1)$ model to include p lags. It is specified by:
$$X_t = a + B_1 X_{t-1} + \ldots + B_p X_{t-p} + \varepsilon_t, \tag{18}$$
where $a \in \mathbb{R}^n$ is a constant vector, and $B_1, \ldots, B_p \in \mathrm{Mat}_n(\mathbb{R})$ are constant matrices.
- The disturbances $\varepsilon \in \mathbb{R}^n$ are i.i.d. random vectors with $\varepsilon \sim N(0, \Omega)$.
- Using the lag operator, we can write this process in the form
$$\psi(L) X_t = a + \varepsilon_t, \tag{19}$$
where
$$\psi(z) = I - B_1 z - \ldots - B_p z^p \tag{20}$$
is a polynomial in z taking on matrix values.

$VAR(p)$ model

- The requirement that X_t is covariance stationary leads to the the requirement that $\mathrm{E}_t(X_t) = \mu$ is a constant, i.e.

$$(I - B_1 - \ldots - B_p)\mu = a. \tag{21}$$

- This is possible, if the matrix $\psi(1) = I - B_1 - \ldots - B_p$ is invertible, or

$$\det(\psi(1)) \neq 0. \tag{22}$$

- Then $\mu = \psi(L)^{-1}a$, and we can write

$$X_t = \mu + \psi(L)^{-1}\varepsilon_t. \tag{23}$$

$VAR(p)$ model

- In order to find further conditions on the parameters of the process, we rewrite (18) in the form:

$$\tilde{X}_t = B_1 \tilde{X}_{t-1} + \ldots + B_p \tilde{X}_{t-p} + \varepsilon_t, \tag{24}$$

where

$$\tilde{X}_t = X_t - \mu \tag{25}$$

is the "de-meaned" process.

- Using the same technique as in the case of $VAR(1)$ we verify that

$$\begin{aligned} \Gamma_0 &= B_1 \Gamma_0 B_1^{\mathrm{T}} + \ldots + B_p \Gamma_0 B_p^{\mathrm{T}} + \Omega, \\ \Gamma_k &= B_1 \Gamma_{k-1} + \ldots + B_p \Gamma_{k-p}, \end{aligned} \tag{26}$$

which is the $VAR(p)$ version of the Yule-Walker equation.

$VAR(p)$ model

- In order to state the second necessary condition for the covariance stationarity of the process (18), we note first that (18) can be written as a $VAR(1)$ of a higher dimension. Namely, we define a pn-dimensional time series

$$\mathbf{X}_t = \begin{pmatrix} \tilde{X}_t \\ \tilde{X}_{t-1} \\ \vdots \\ \tilde{X}_{t-p+1} \end{pmatrix}$$

- We also define

$$\mathbf{B} = \begin{pmatrix} B_1 & B_2 & \dots & B_{p-1} & B_p \\ I & 0 & \dots & 0 & 0 \\ \vdots & \vdots & \dots & \vdots & \vdots \\ 0 & 0 & \dots & I & 0 \end{pmatrix}, \text{ and } \eta_t = \begin{pmatrix} \varepsilon_t \\ 0 \\ \vdots \\ 0 \end{pmatrix},$$

where 0 denotes here the $n \times n$ zero matrix.

$VAR(p)$ model

- With these definitions, the process (18) can be written as (verify!):

$$\mathbf{X}_t = \mathbf{B}\mathbf{X}_{t-1} + \eta_t. \tag{27}$$

- One can show, by extending the arguments presented in the case of $VAR(1)$, that the is the requirement that all the roots of the equation

$$\det(I - \mathbf{B}z) = 0 \tag{28}$$

  lie outside of the unit circle.
- As in the case of $VAR(1)$, we can find an expression for Γ_k in terms of an infinite series in the coefficients B_j. This series converges under the stationarity condition formulated above.
- All of this is fairly technical, and we will not study it in detail.

MLE for $VAR(p)$

- $VAR(p)$ models can be estimated using MLE very much the same way as in the case of the the univariate $AR(p)$ model. Let $x_0, \ldots, x_T \in \mathbb{R}^n$ be a series of observations of an n-variate time series.
- Denote by $\theta = (a, B_1, \ldots, B_p, \Omega)$ the parameters of the model. The conditional PDF of the data can be written as:

$$\begin{aligned} p(x_{p:T}|x_{0:p-1}, \theta) &= \prod_{t=p}^{T} p(x_t|x_{0:t-1}, \theta) \\ &= \prod_{t=p+1}^{T} p(\widehat{\varepsilon}_t|x_{0:p-1}, \theta) \\ &= \prod_{t=p+1}^{T} \frac{1}{(2\pi)^{n/2}(\det(\Omega))^{1/2}} \exp\Big(-\frac{1}{2}\widehat{\varepsilon}_t^{\mathbb{T}}\Omega^{-1}\widehat{\varepsilon}_t\Big). \end{aligned}$$

MLE for *VAR*(p)

- Hence the LLF is

$$-\log \mathcal{L}(\theta|x_{0:T}) = \frac{1}{2}(T-p)\log\det(\Omega) + \frac{1}{2}\sum_{t=p+1}^{T} \varepsilon_t^{\mathrm{T}}\Omega^{-1}\widehat{\varepsilon}_t + const. \qquad (29)$$

- Notice that, if v_i are vectors and M is a matrix, then

$$\sum_i v_i^{\mathrm{T}} M v_i = \mathrm{tr}\Big(M\sum_i v_i v_i^{\mathrm{T}}\Big),$$

  where tr denotes the trace of a matrix (verify this!).
- This allow us to rewrite the LLF above as:

$$-\log \mathcal{L}(\theta|x_{0:T}) = \frac{1}{2}(T-p)\log\det(\Omega) + \frac{1}{2}\mathrm{tr}\Big(\Omega^{-1}\sum_{t=p+1}^{T}\widehat{\varepsilon}_t\widehat{\varepsilon}_t^{\mathrm{T}}\Big) + const. \qquad (30)$$

MLE for *VAR*(p)

- With a bit of multi-linear algebra, one can calculate the derivatives of the LLF with respect to the components of θ and find the critical points (the preferred values of the parameters $\widehat{\theta}$). In particular, we find that the estimated covariance matrix $\widehat{\Omega}$ is given by

$$\widehat{\Omega} = \frac{1}{T-p} \sum_{t=p+1}^{T} \widehat{\varepsilon}_t \widehat{\varepsilon}_t^{\mathrm{T}}, \tag{31}$$

  where the $\widehat{\varepsilon}_t$'s are calculated using the optimized values of $\widehat{a}$ and $\widehat{B}_j$.
- We can thus write the maximized LLF in the following way:

$$-\log \mathcal{L}(\widehat{\theta}|x_{0:T}) = \frac{1}{2}(T-p)\log\det(\widehat{\Omega}) + \frac{1}{2}(T-p)n + const. \tag{32}$$

  We will use this expression later when discussing cointegration.

$VAR(p)$ model selection and checking

- As in the case of $AR(p)$, a useful guide in the selection of the lag parameter p may be one of the information criteria, which in the case of $VAR(p)$ take the form:
  - (i) Akaike information criterion,

$$\mathrm{AIC} = 2pn^2 - 2\log \mathcal{L}(\widehat{\theta} \mid x_{0:T}). \tag{33}$$

  - (ii) Bayesian information criterion,

$$\mathrm{BIC} = \log(T)pn^2 - 2\log \mathcal{L}(\widehat{\theta} \mid x_{0:T}). \tag{34}$$

- It is always a good idea to check that the residuals $\widehat{\varepsilon}_t$ calculate with the optimized values of the parameters have the desired properties: independence for different t's and stationarity. This can be done by running the appropriate statistics.

Structural versus reduced form $VAR(p)$ models

- Sometimes, especially in the economic literature, one finds the distinction between two forms of the $VAR(p)$ model: *structural* and *reduced form*.
- In the structural form, the covariance matrix Ω of the shocks is diagonal (i.e. structural shocks are uncorrelated). The model is specified as follows:

$$B_0 X_t = a + B_1 X_{t-1} + \ldots + B_p X_{t-p} + \varepsilon_t, \tag{35}$$

  where B_0 has 1's on its diagonal.
- In particular, this means that the current value of the variable $X_{t,c}$ can be influenced by the lagged values of all the variables as well as the contemporaneous values of the variables $X_{t,b}$ with $b \neq c$.
- From an economic point of view, a relationship represented by a structural $VAR(p)$ model reflects an underlying "structural" economic reality.
- This is particularly valuable when:
  (i) the drivers of various components of X_t are independent of each other;
  (i) different components of X_t can exert contemporaneous influence on each other.

Structural versus reduced form $VAR(p)$ models

- Multiplying equation (35) on the left by B_0^{-1} and renaming $B_0^{-1}a$ by a, $B_0^{-1}B_1$ by B_1, etc, we arrive at the reduced, i.e. standard, form (4) of the $VAR(p)$ model specification.
- In the reduced form, the variables on the right hand side are predetermined at time t: no contemporaneous impact of one variable on another one is possible.
- On the the other hand, the covariance matrix $B_0^{-1}\Omega(B_0^{-1})^{\intercal}$ of the reduced form shocks is no longer diagonal, and thus the shocks are not independent. Different variables impact each other not through contemporaneous influences but through correlations among shocks.
- For an in depth discussion of structural $VAR(p)$ models, see Lütkepohl's book [2].

VARMA(p, q) model

- A *VARMA*(p, q) model is an extension of the *VAR*(1) model to include p lags. It is specified by:

$$\psi(L)X_t = a + \varphi(L)\varepsilon_t, \tag{36}$$

where $a \in \mathbb{R}^n$ is a constant vector, and $\psi(z)$ and $\varphi(z)$ are matrix-valued polynomials of an argument z:

$$\begin{aligned} \psi(z) &= 1 - B_1 z - \ldots - B_p z^p, \\ \varphi(z) &= 1 + \Theta_1 z + \ldots + \Theta_q z^q. \end{aligned} \tag{37}$$

- Assuming that $\psi(z)$ has no unit roots, we can write this model in the moving average form:

$$X_t = \mu + \gamma(L)\varepsilon_t, \tag{38}$$

where $\gamma(L) = \psi(L)^{-1}\varphi(L)$.

Granger causality

- Multivariate time series analysis allows for a concept of causality that opens up new ways to determine whether one time series in useful in forecasting another one.
- This concept, the *Granger causality*, is a statistical test based on a $VAR(p)$ forecasting and is formulated as follows.
- Consider first a bivariate time series, which is assumed to follow an $AR(1)$ process:

$$\begin{aligned} X_{t,1} &= a_1 + B_{11}X_{t-1,1} + B_{12}X_{t-1,2} + \varepsilon_{t,1}, \\ X_{t,2} &= a_2 + B_{21}X_{t-1,1} + B_{22}X_{t-1,2} + \varepsilon_{t,2}. \end{aligned} \tag{39}$$

- If $B_{12} = 0$, there is a directional asymmetry between the two time series: $X_{t,1}$ serves as an input variable, while $X_{t,2}$ is an output variable. In other words, $X_{t,2}$ does not influence $X_{t,1}$, while $X_{t,1}$ influences $X_{t,2}$. The two variables are said to form a *transfer function* relationship.

Granger causality

- Continuing with the bivariate example (39), consider now two k-period forecasts of X_t:

$$X^*_{t+k|f} = \mathrm{E}(X_{t+k}|X_{1:t,1}, X_{1:t,2}),$$
$$X^*_{t+k|p} = \mathrm{E}(X_{t+k}|X_{1:t,1}).$$

  In other words, $X^*_{t+k|f}$ is the forecast based on full information up time t, while $X^*_{t+k|p}$ is based only on the information generated by the first series.

- Let $\varepsilon_{k|f}$ and $\varepsilon_{k|p}$ denote the corresponding forecasting errors. Then, we say that $X_{t,2}$ Granger-causes $X_{t,1}$, if

$$\mathrm{Var}(\varepsilon_{k|f}) < \mathrm{Var}(\varepsilon_{k|p}). \tag{40}$$

- In other words, $X_{t,2}$ Granger-causes $X_{t,1}$, if the forecast of $X_{t,1}$ is improved by including the information about $X_{t,2}$.

Granger causality

- The null hypothesis in the Granger test is no Granger causality, i.e. $H_0 : B_{12} = 0$. This is essentially an F-test on the regression coefficient (having no explanatory power).
- One can generalize the discussion above to the $VAR(p)$ setup with $p \geq 2$. In this case, the null hypothesis cannot be rejected, if no lagged values of $X_{t,2}$ are retained in the regression.

Granger causality

- Note that Granger causality is not necessarily identical with physical causality. Granger causality is observed only indirectly, through a time series.
- If both processes $X_{t,1}$ and $X_{t,2}$ are driven by a common (but not explicitly included in the model) third process, one might still be likely not to reject the alternative hypothesis of Granger causality.
- The Granger test is designed to handle two variables, and may result in incorrect conclusions when the underlying relationship involves such *confounding variables* that causally affect both variables $X_{t,1}$ and $X_{t,2}$.

Cointegration

- Cointegration is an expression of a long term equilibrium relationship between the components of X_t.
- In Lecture Notes #2 we discussed the concept of cointegration in the context of two time series.
- Using the concepts of multivariate time series analysis discussed above, we can now extend this concept to multiple cointegrating relations.
- Let X_t be a multivariate time series of dimension n, such that
  - (i) each component of X_t is $I(1)$,
  - (ii) a *cointegrating vector* $a \in \mathbb{R}^n$, i.e. a non-zero vector such that the univariate time series $a^{\mathrm{T}} X_t$ is $I(0)$ (i.e. stationary) exists.
- We say then that X_t is cointegrated of order 1. If each component of X_t is $I(d)$ and $a^{\mathrm{T}} X_t$ is $I(d')$ with $d' < d$, we say that X_t is cointegrated.

Cointegration

- More generally, suppose that there exists a matrix $\alpha \in \mathrm{Mat}_{n,r}(\mathbb{R})$, with $r < n$, such that
  - (i) α has (full) rank r (i.e. there are no linear dependencies among the columns of the matrix α),
  - (ii) the r-dimensional time series $Y_t = \alpha^{\mathrm{T}} X_t$ is $I(0)$.

  Then X_t is called cointegrated of order 1 with r *cointegrating vectors*, namely the columns of α.
- From linear algebra we know that there exists a matrix $\tilde{\alpha} \in \mathrm{Mat}_{n,n-r}(\mathbb{R})$ of rank $n - r$, such that $\alpha^{\mathrm{T}} \tilde{\alpha} = 0$.
- As a consequence, if X_t is cointegrated of rank r, then the time series $Z_t = \tilde{\alpha}^{\mathrm{T}} X_t$ is an $(n - r)$-dimensional unit root non-stationary time series. Its components are called the *common trends* of X_t.
- We shall now develop an MLE framework for testing for cointegration.

Cointegration

- We assume that X_t follows a $VAR(p)$ specification given by (18). We wish to study it for possible cointegration.
- The method used to derive the ECM form of the $AR(p)$ process in Lecture Notes #2 can be applied to (18). Namely, we define

$$\begin{aligned} B &= B_1 + \ldots + B_p, \\ \Gamma_j &= -(B_{j+1} + \ldots + B_p), \qquad \text{for } j = 1, \ldots, p-1. \end{aligned} \tag{41}$$

- After some algebra we can then write (18) in the form:

$$\left(1 - BL - (\Gamma_1 L + \ldots + \Gamma_{p-1} L^{p-1})(1-L)\right) X_t = a + \varepsilon_t. \tag{42}$$

- Equivalently, this can be written as

$$X_t = a + BX_{t-1} + \Gamma_1 \Delta X_{t-1} + \ldots + \Gamma_{p-1} \Delta X_{t-p+1} + \varepsilon_t. \tag{43}$$

This form of the $AR(p)$ process is referred to as the *vector error correcting model* (VECM).

Cointegration

- For the purpose of testing for cointegration, we shall write the equation above in the following form:

$$\Delta X_t = a + \Pi X_{t-1} + \Gamma_1 \Delta X_{t-1} + \ldots + \Gamma_{p-1} \Delta X_{t-p+1} + \varepsilon_t, \tag{44}$$

  where the matrix Π is defined by

$$\Pi = B - I.$$

- This form of the equation makes it explicit that key to understanding the cointegration structure of X_t is the rank of the matrix Π.

Cointegration

- For example, the process (43) of Lecture Notes #2 can be written in the *VECM* form as follows:

$$\begin{pmatrix} \Delta X_t \\ \Delta Y_t \end{pmatrix} = \begin{pmatrix} \alpha_1 + \gamma\alpha_2 \\ \alpha_2 \end{pmatrix} + \begin{pmatrix} -1 & \gamma \\ 0 & 0 \end{pmatrix} \begin{pmatrix} X_{t-1} \\ Y_{t-1} \end{pmatrix} + \begin{pmatrix} \varepsilon_{t,1} + \gamma\varepsilon_{t,2} \\ \varepsilon_{t,2} \end{pmatrix}$$

- Notice that the 2×2-matrix

$$\Pi = \begin{pmatrix} -1 & \gamma \\ 0 & 0 \end{pmatrix}$$

is of rank 1.

- Define

$$\alpha = \begin{pmatrix} 1 \\ -\gamma \end{pmatrix}, \ \beta = \begin{pmatrix} -1 \\ 0 \end{pmatrix}, \ \tilde{\alpha} = \begin{pmatrix} \gamma \\ 1 \end{pmatrix}.$$

Then $\alpha^{\mathrm{T}}\tilde{\alpha} = 0$, and $\Pi = \beta\alpha^{\mathrm{T}}$.

Cointegration

- General methodology for the testing for cointegration was developed by Johansen [1]. For a textbook presentation, I recommend Tsay's book [3].
- A cointegration test is essentially a test on the rank r of the matrix Π in (44).
- If $r = 0$, then $\Pi = 0$, and clearly there are no cointegrating vectors.
- If, on the other hand, $r > 0$, then X_t has r cointegrating vectors and $n - r$ unit roots. In this case there is a cointegrating matrix $\alpha \in \mathrm{Mat}_{n,r}$ such that the r-dimensional process $\alpha^{\mathrm{T}} X_t$ is $I(0)$.
- The matrix Π can be represented as $\Pi = \beta\alpha^{\mathrm{T}}$, where $\beta \in \mathrm{Mat}_{n,r}$ has rank r. In other words, in the presence of r cointegrating relations, only r linear combinations of X_{t-1} appear in (44).
- The $(n - r)$-dimensional unit root process $\tilde{\alpha}^{\mathrm{T}} X_t$, where $\tilde{\alpha} \in \mathrm{Mat}_{n,n-r}(\mathbb{R})$, $\alpha^{\mathrm{T}}\tilde{\alpha} = 0$, is the process of common trends (as discussed earlier).

Cointegration

- We reformulate the model (44) in terms of two regressions:

$$\begin{aligned} \Delta X_t &= a + Q_1 \Delta X_{t-1} + \ldots + Q_p \Delta X_{t-p+1} + u_t, \\ X_{t-1} &= a + R_1 \Delta X_{t-1} + \ldots + R_p \Delta X_{t-p+1} + v_t, \end{aligned} \tag{45}$$

  where u_t and v_t are the errors. These regressions are estimated using the the least squares (MLE).

- Given the estimated coefficients $\widehat{Q}_i$ and $\widehat{R}_i$ in (45), we calculate the residuals $\widehat{u}_t$ and $\widehat{v}_t$, and run the third regression:

$$\widehat{u}_t = \Pi \widehat{v}_t + \varepsilon_t, \tag{46}$$

  with errors ε_t. Its purpose is to estimate the matrix Π.

- Not coincidentally, the residulas in (46) and (44) are identical, hence the notation.

Likelihood ratio test

- Before we move on, we take a bit of a break to review the basic ideas behind the *likelihood ratio test* under multivariate Gaussian distribution.
- Consider an n-dimensional Gaussian random variable $X \sim N(0, \Sigma)$. We decompose it into two components U and V,

$$X = \begin{pmatrix} U \\ V \end{pmatrix}, \tag{47}$$

where U has dimension p, $U \sim N(0, \Sigma_U)$, and V has dimension $q = n - p$, $V \sim N(0, \Sigma_V)$.

- We write accordingly

$$\Sigma = \begin{pmatrix} \Sigma_U & \Sigma_{UV} \\ \Sigma_{VU} & \Sigma_V \end{pmatrix}, \tag{48}$$

where $\Sigma_{UV} = \Sigma_{VU}^{\mathrm{T}}$ is the covariance matrix of U and V.

Likelihood ratio test

- Our goal is to test the null hypothesis $H_0 : \Sigma_{UV} = 0$, i.e. $\mathrm{Cov}(U, V) = 0$, versus the alternative $H_a : \Sigma_{UV} \neq 0$.
- Let x_i, $i = 1, \ldots, N$, be a sample of observations, and let

$$\begin{aligned} \widehat{\Sigma}_U &= \frac{1}{N} \sum_{i=1}^{N} u_i u_i^{\mathrm{T}}, \\ \widehat{\Sigma}_U &= \frac{1}{N} \sum_{i=1}^{N} v_i v_i^{\mathrm{T}} \end{aligned} \tag{49}$$

  denote the MLE estimates of Σ_U and Σ_V, respectively.
- Under the null hypothesis H_0:

$$\widehat{\Sigma}_0 = \begin{pmatrix} \widehat{\Sigma}_U & 0 \\ 0 & \widehat{\Sigma}_V \end{pmatrix}. \tag{50}$$

Likelihood ratio test

- As we have already seen, the maximized likelihood function under the multivariate Gaussian distribution is of the form

$$\begin{aligned}\mathcal{L}(\widehat{\Sigma}|x) &= const \times \det(\widehat{\Sigma})^{-N/2} \\ &\propto \det(\widehat{\Sigma})^{-N/2},\end{aligned}$$

- Therefore, under H_0,

$$\begin{aligned}\mathcal{L}(\widehat{\Sigma}_0|x) &\propto \det(\widehat{\Sigma}_0)^{-N/2} \\ &= \det(\widehat{\Sigma}_U)^{-N/2} \det(\widehat{\Sigma}_V)^{-N/2},\end{aligned} \tag{51}$$

- On the other hand, under the alternative hypothesis H_a,

$$\widehat{\Sigma}_a = \begin{pmatrix} \widehat{\Sigma}_U & \widehat{\Sigma}_{UV} \\ \widehat{\Sigma}_{VU} & \widehat{\Sigma}_V \end{pmatrix}. \tag{52}$$

Likelihood ratio test

- We shall now use a property of the determinant of a matrix M familiar from linear algebra. Namely, its value remains unchanged, if to a row of M we add any linear combination of other rows of M.
- In particular,

$$\det \begin{pmatrix} \widehat{\Sigma}_U & \widehat{\Sigma}_{UV} \\ \widehat{\Sigma}_{VU} & \widehat{\Sigma}_V \end{pmatrix} = \det \begin{pmatrix} \widehat{\Sigma}_U & \widehat{\Sigma}_{UV} \\ 0 & \widehat{\Sigma}_V - \widehat{\Sigma}_{VU}\widehat{\Sigma}_U^{-1}\widehat{\Sigma}_{UV} \end{pmatrix},$$

  where we have performed linear operations on $\widehat{\Sigma}_a$ in such a way that $\widehat{\Sigma}_{VU}$ is replaced by the zero matrix.
- The resulting matrix has a block upper triangular structure, and its determinant is the product of the determinants of the diagonal blocks. Hence,

$$\begin{aligned} \mathcal{L}(\widehat{\Sigma}_a|x) &\propto \det(\widehat{\Sigma}_a)^{-N/2} \\ &= \det(\widehat{\Sigma}_U)^{-N/2} \det(\widehat{\Sigma}_V - \widehat{\Sigma}_{VU}\widehat{\Sigma}_U^{-1}\widehat{\Sigma}_{UV})^{-N/2}. \end{aligned} \tag{53}$$

Likelihood ratio test

- The ratio statistics $\mathcal{R}$ is defined as

$$\mathcal{R} = \frac{\mathcal{L}(\widehat{\Sigma}_0|x)}{\mathcal{L}(\widehat{\Sigma}_a|x)}\,. \tag{54}$$

  The null hypothesis is rejected if $\mathcal{R}$ is small.

- Let λ_i, $i = 1, \ldots, q$, denote the eigenvalues of the matrix $\widehat{\Sigma}_V^{-1}\widehat{\Sigma}_{VU}\widehat{\Sigma}_U^{-1}\widehat{\Sigma}_{UV}$. Then

$$\begin{aligned}\mathcal{R} &= \frac{\det(\widehat{\Sigma}_V)^{-N/2}}{\det(\widehat{\Sigma}_V - \widehat{\Sigma}_{VU}\widehat{\Sigma}_U^{-1}\widehat{\Sigma}_{UV})^{-N/2}} \\ &= \det(1 - \widehat{\Sigma}_V^{-1}\widehat{\Sigma}_{VU}\widehat{\Sigma}_U^{-1}\widehat{\Sigma}_{UV})^{N/2} \\ &= \prod_{i=1}^{q}(1-\lambda_i)^{N/2}\,.\end{aligned}$$

Likelihood ratio test

- It is customary to work with the log likelihood statistics normalized as follows:

$$\mathcal{LR} = -2 \log \mathcal{R}. \tag{55}$$

Explicitly,

$$\begin{aligned} \mathcal{LR} &= -N \log \prod_{i=1}^{q} (1 - \lambda_i) \\ &= -N \sum_{i=1}^{q} \log(1 - \lambda_i). \end{aligned} \tag{56}$$

- The null hypothesis is rejected if $\mathcal{LR}$ is large.
- We can now go back to the analysis of cointegrating time series.

Cointegration tests

- Going back to (46), we now wish to test the following hypothesis:

$$H_0 : \operatorname{rank}(\Pi) = r, \qquad \text{against } H_1 : r < \operatorname{rank}(\Pi) \leq r_1. \tag{57}$$

- All residulas and estimated covariance matrices are computed according to the expressions discussed earlier (see (49)).
- The statistic used for testing the hypothesis (57) is given by the log likelihood ratio $\mathcal{LR}$ introduced above (with N replaced by T, the length of the observed series).
- We use it as follows. Define

$$\begin{aligned}\mathcal{LR}(r, r_1) &= \mathcal{LR}(r_1) - \mathcal{LR}(r) \\ &= T\Big(-\sum_{j=1}^{r_1} \log(1-\lambda_j) + \sum_{j=1}^{r} \log(1-\lambda_j)\Big) \\ &= -T \sum_{j=r+1}^{r_1} \log(1-\lambda_j).\end{aligned} \tag{58}$$

Cointegration tests

- Calculations, similar to the calculations we did in Lecture Notes #2, show that the asymptotic distributions of these statistics under the null hypothesis are as follows.
- Let $\mathcal{X}$ be the random matrix:

$$\mathcal{X} = \Big(\int_0^1 W(s) dW(s)^{\mathrm{T}} \Big)^{\mathrm{T}} \Big(\int_0^1 W(s) W(s)^{\mathrm{T}} ds \Big)^{-1} \Big(\int_0^1 W(s) dW(s)^{\mathrm{T}} \Big). \quad (59)$$

Here, $W(s)$ is the standard Brownian motion of dimension $n - r$. Then:

(i)

$$\mathcal{LR}(r, n) \longrightarrow \mathrm{tr}(\mathcal{X}), \quad (60)$$

(ii)

$$\mathcal{LR}(r, r+1) \longrightarrow \lambda_{max}, \quad (61)$$

where λ_{max} denotes the largest eigenvalue of $\mathcal{X}$.

Cointegration

- In other words, the limit distributions of both statistics (under null hypothesis) are functionals of a multi-factor Brownian motion.
- Samples from these distributions can be easily simulated by standard Monte Carlo methods, and the critical values for the log likelihood ratio tests can be deduced.
- The critical values are tabulated in software packages. An R implementation is in the package `urca`.
- For these tests, the number of lags p has to be known. This can be chosen with the help of the AIC or BIC criteria discussed above.

Cointegration tests

- There are two versions of the Johansen cointegration test:
  - (i) The *trace test*: in the trace test we work with the log likelihood ratio stat $\mathcal{LR}(r, n)$ for testing the null hypothesis (57).
  - (ii) The *maximum eigenvalue test*: in the maximum eigenvalue test we work with the log likelihood ratio stat $\mathcal{LR}(r, r+1)$ for testing the null hypothesis (57).
- Generally, the trace test is regarded as stronger.

Cointegration

- In order to determine the cointegrating rank we run a sequence of tests:

$$\begin{aligned} H_0 &: r = 0, \\ H_0 &: r = 1, \\ &\dots, \\ H_0 &: r = n-1, \end{aligned} \tag{62}$$

  and reject the null hypothesis the first time it is possible.
- Both the trace test and maximum eigenvalue test can be used.
- Example: take $n = 3$.
  - (i) We first test whether $r = 0$. If this cannot be rejected, we further analyze the corresponding VECM model.
  - (ii) If we can reject $r = 0$, we proceed to testing $r = 1$. If $r = 1$ cannot be rejected, we continue with analyzing the VECM model with $r = 1$.
  - (iii) Otherwise, we test $r = 2$. If this hypothesis cannot be rejected, we take $r = 2$ and analyze the VECM model.
  - (iv) Finally, if $r = 2$ cannot be rejected, it may mean that the VAR model is stationary.

Volatility in multivariate time series

- When analyzing jointly the volatility of a number of financial time series, it may be important to take into account the dependencies between the shocks contributing to the volatilities of each of the individual time series
- Modeling each individual volatility series as a univariate *GARCH* (or other) model may be then inappropriate, as this approach does not take into account dependencies between different time series.
- There exist a number of more or less sophisticated approaches to modeling joint stochastic volatility of a multivariate time series. Here we provide a very brief introduction to some of them.

EWMA model

- The simplest and most parsimonous approach is the *exponentially weighted moving average* (EWMA) model which is a natural extension of the analogous univariate model discussed in Lecture Notes #2.
- According to the general principle, we decompose a (multivariate) time series into the seasonality, trend, and disturbance parts (see formula (6) in Lecture Notes # 1).
- The disturbance model is specified as follows:

$$\begin{aligned} \varepsilon_t &= \Omega_t^{1/2} z_t, \\ \Omega_t &= \lambda \Omega_{t-1} + (1-\lambda)\varepsilon_t \varepsilon_t^{\mathrm{T}}, \end{aligned} \tag{63}$$

where z_t is a sequence of i.i.d. random vectors $z_t \sim N(0, I_n)$, and where I_n denotes the n-dimensional identity matrix.

- Here, $\Omega^{1/2}$ denotes the *square root* of the positive definite matrix Ω, which is defined as the *unique positive define* matrix such that $(\Omega^{1/2})^2 = \Omega$.
- Note that

$$\begin{aligned} \text{Cov}(\varepsilon_t, \varepsilon_t) &= \text{E}(\varepsilon_t \varepsilon_t^{\text{T}}) \\ &= \text{E}(\Omega_t^{1/2} z_t z_t^{\text{T}} \Omega_t^{1/2}) \\ &= \Omega_t^{1/2} I_n \Omega_t^{1/2} \\ &= \Omega_t, \end{aligned} \tag{64}$$

i.e. the process Ω_t is the covariance process of ε_t.

BEKK model

- The $BEKK(p, q)$ model (after Baba, Engle, Kraft, and Kroner) is a multivariate extension of the GARCH.
- Its simplest version, $BEKK(1, 1)$ is specified as follows:

$$\begin{aligned} \varepsilon_t &= \Omega_t^{1/2} z_t, \\ \Omega_t &= AA^{\mathrm{T}} + B\varepsilon_t\varepsilon_t B^{\mathrm{T}} + C\Omega_{t-1}C^{\mathrm{T}}. \end{aligned} \tag{65}$$

  Here, $A, B, C \in \mathrm{Mat}_n(\mathbb{R})$ are square matrices. It is usually additionally assumed that A is lower triangular.
- The model has a vast number of parameters (even with the restriction on A): $n(n+1)/2 + 2n^2$. For example, without further restrictions, this model requires 11 parameters for a bivariate time series.

Dynamic conditional correlation model

- Another simple approach to modeling multivariate volatility is the *dynamic conditional correlation* (DCC) model, which is specified as follows.
- Let Ω_t denote the covariance matrix of the shocks ε_t, and let σ_t^2 be its diagonal. The nonzero elements of σ_t^2 are the variances of the individual time series at time t, and so

$$\rho_t = \sigma_t^{-1} \Omega_t \sigma_t^{-1} \tag{66}$$

  is the correlation matrix at time t.
- The model is specified as follows:

$$\begin{aligned} q_t &= (1 - \vartheta_1 - \vartheta_2)\overline{q} + \vartheta_1 q_{t-1} + \vartheta_2 \eta_{t-1} \eta_{t-1}^{\mathrm{T}}, \\ \rho_t &= d_t^{-1} q_t d_t^{-1}. \end{aligned} \tag{67}$$

Dynamic conditional correlation model

- Here:
  - (i) ϑ_1, ϑ_2 are real parameters such that $0 < \vartheta_1 + \vartheta_2 < 1$,
  - (ii) $\eta_{t,a} = \varepsilon_{t,a}/\sigma_{t,a}$ are the standardized disturbances,
  - (iii) d_t is the positive square root of the diagonal of q_t and is simply a normalizing factor.
  - (iv) $\overline{q}$ is the unconditional covariance matrix of the η's.
- Note that the model is specified so that q_t is guaranteed to be positive definite.
- In practice, the volatilities $\sigma_{t,a}$ can be estimated from univariate *GARCH* models.
- The DCC model is thus very parsimonious, as it has two free parameters only. The impact of initial choices can be decayed by priming the model as explained in Lecture Notes #2.

References


Johansen, S.: *Likelihood-Based Inference in Cointegrated Vector Autoregressive Models*, Oxford University Press (1995).


Lütkepohl, H.: *New Introduction to Multiple Time Series Analysis*, Springer (2006).


Tsay, R. S.: *Multivariate Time Series Analysis*, Wiley (2014).