1. Vector autoregression and cointegration

1.1 See slides 9, and 10-11

1.2 See slides 14

1.3 See slides 17-18, 19-21

Express $X_{2,t}$ as a transfer function noise (TFN) model of $X_{1,t}$:

$$X_{2,t} = \sum_{i=1\ to\ p} \phi_{21}^{(i)} X_{1,t-i} + \sum_{i=1\ to\ p} \phi_{22}^{(i)} X_{2,t-i} + a_{2,t}$$

$$X_{2,t} - \sum_{i=1\ to\ p} \phi_{22}^{(i)} X_{2,t-i} = \sum_{i=1\ to\ p} \phi_{21}^{(i)} X_{1,t-i} + a_{2,t}$$

$$\phi_{22}(B) X_{2,t} = \phi_{21}(B) X_{1,t} + a_{2,t}$$

$$X_{2,t} = \frac{\phi_{21}(B)}{\phi_{22}(B)} X_{1,t} + \frac{1}{\phi_{22}(B)} a_{2,t}$$

This is

$$X_{2,t} = \nu(B) X_{1,t} + N_t$$

1.4 Derive implied model for $X_{2,t}$ for VAR(1) model

$$\begin{pmatrix} X_{1,t} \\ X_{2,t} \end{pmatrix} = \begin{bmatrix} \phi_{11}^{(1)} & \phi_{12}^{(1)} \\ \phi_{21}^{(1)} & \phi_{22}^{(1)} \end{bmatrix} \begin{pmatrix} X_{1,t-1} \\ X_{2,t-1} \end{pmatrix} + \begin{pmatrix} a_{1,t} \\ a_{2,t} \end{pmatrix}$$

$$\begin{bmatrix} 1-\phi_{11}^{(1)}B & -\phi_{12}^{(1)}B \\ -\phi_{21}^{(1)}B & 1-\phi_{22}^{(1)}B \end{bmatrix} \begin{pmatrix} X_{1,t} \\ X_{2,t} \end{pmatrix} = \begin{pmatrix} a_{1,t} \\ a_{2,t} \end{pmatrix}$$

$$\left(adj \begin{bmatrix} 1-\phi_{11}^{(1)}B & -\phi_{12}^{(1)}B \\ -\phi_{21}^{(1)}B & 1-\phi_{22}^{(1)}B \end{bmatrix} \right) \begin{bmatrix} 1-\phi_{11}^{(1)}B & -\phi_{12}^{(1)}B \\ -\phi_{21}^{(1)}B & 1-\phi_{22}^{(1)}B \end{bmatrix} = det \begin{bmatrix} 1-\phi_{11}^{(1)}B & -\phi_{12}^{(1)}B \\ -\phi_{21}^{(1)}B & 1-\phi_{22}^{(1)}B \end{bmatrix} \times I_2$$

$$adj \begin{bmatrix} 1-\phi_{11}^{(1)}B & -\phi_{12}^{(1)}B \\ -\phi_{21}^{(1)}B & 1-\phi_{22}^{(1)}B \end{bmatrix} = \begin{bmatrix} 1-\phi_{22}^{(1)}B & \phi_{12}^{(1)}B \\ \phi_{21}^{(1)}B & 1-\phi_{11}^{(1)}B \end{bmatrix}$$

$$det \begin{bmatrix} 1-\phi_{11}^{(1)}B & -\phi_{12}^{(1)}B \\ -\phi_{21}^{(1)}B & 1-\phi_{22}^{(1)}B \end{bmatrix} = \left(1-\phi_{11}^{(1)}B\right)\left(1-\phi_{22}^{(1)}B\right) - \phi_{12}^{(1)}\phi_{21}^{(1)}B^2$$

$$det \begin{bmatrix} 1-\phi_{11}^{(1)}B & -\phi_{12}^{(1)}B \\ -\phi_{21}^{(1)}B & 1-\phi_{22}^{(1)}B \end{bmatrix} \begin{pmatrix} X_{1,t} \\ X_{2,t} \end{pmatrix} = \begin{bmatrix} 1-\phi_{22}^{(1)}B & \phi_{12}^{(1)}B \\ \phi_{21}^{(1)}B & 1-\phi_{11}^{(1)}B \end{bmatrix} \begin{pmatrix} a_{1,t} \\ a_{2,t} \end{pmatrix}$$

Then the implied model for $X_{2,t}$ is

$$\left[\left(1-\phi_{11}^{(1)}B\right)\left(1-\phi_{22}^{(1)}B\right)-\phi_{12}^{(1)}\phi_{21}^{(1)}B^2\right]X_{2,t}=\phi_{21}^{(1)}a_{1,t-1}+a_{2,t}-\phi_{11}^{(1)}a_{2,t-1}$$

1.5 See slides 19-21

1.6 Answer:

If $\left(X_{1,t},\ X_{2,t}\right)$ are not cointegrated, we can first difference the two series individually untile we have stationary processes, then we can use VAR(p) model to fit the two stationary processes and test Granger causality; If $\left(X_{1,t},\ X_{2,t}\right)$ are cointegrated, then we need to use error correction model (ECM), i.e. to include lagged disequilibrium terms as explanatory variables.

1.7 Answer:

If cointegration exists, the VAR(p) model directly fitted to the differenced stationary processes will be misspecified.

1.8 See slides 34

1.9 Answer:

Granger representation theorem implies that we can and should use ECM model to model cointegrated non-stationary time series.

2. Bootstrap time series

Consider AR(2) model

$$y_t=\mu+\phi_1y_{t-1}+\phi_2y_{t-2}+a_t$$

$$a_t\sim NID(0,1)$$

2.1 Steps of (unconditional) parametric bootstrap

Step1: Estimate $\hat{\mu}$, $\widehat{\phi_1}$, $\widehat{\phi_2}$ (maximum likelihood estimation);

Step2: Compute unconditional distribution of y_t

$$\hat{E}[y_t]=\frac{\hat{\mu}}{1-\widehat{\phi_1}-\widehat{\phi_2}}$$

$$\widehat{Var}[y_t] = \frac{(1-\widehat{\phi_2})}{(1+\widehat{\phi_2})(1-\widehat{\phi_1}-\widehat{\phi_2})(1+\widehat{\phi_1}-\widehat{\phi_2})}$$

$$\hat{\rho}(1) = \widehat{corr}(y_t, y_{t-1}) = \frac{\widehat{\phi_1}}{1-\widehat{\phi_2}}$$

$$\widehat{cov}(y_t, y_{t-1}) = \hat{\rho}(1) \times \widehat{Var}[y_t]$$

Then

$$\begin{pmatrix} y_t \\ y_{t-1} \end{pmatrix} \sim N(\begin{pmatrix} \hat{E}[y_t] \\ \hat{E}[y_{t-1}] \end{pmatrix}, \begin{bmatrix} \widehat{Var}[y_t] & \widehat{cov}(y_t, y_{t-1}) \\ \widehat{cov}(y_t, y_{t-1}) & \widehat{Var}[y_{t-1}] \end{bmatrix})$$

Step3: Simulate $\tilde{y}_0$ and $\tilde{y}_1$ jointly from above bivariate normal distribution, simulate $\tilde{a}_2$ from $N(0,1)$, then

$$\tilde{y}_2 = \hat{\mu} + \widehat{\phi_1}\tilde{y}_1 + \widehat{\phi_2}\tilde{y}_0 + \tilde{a}_2$$

Step4: With $\tilde{y}_{t-1}$ and $\tilde{y}_{t-2}$, simulate $\tilde{a}_t$, and

$$\tilde{y}_t = \hat{\mu} + \widehat{\phi_1}\tilde{y}_{t-1} + \widehat{\phi_2}\tilde{y}_{t-2} + \tilde{a}_t$$

2.2 Steps of Sieve bootstrap

Step1: Estimate $\hat{\mu}$, $\widehat{\phi_1}$, $\widehat{\phi_2}$, $\hat{a}_t$ (least square estimation);

Step2: (optional) Rescale estimated residual

$$\ddot{a}_t = \left(\frac{n}{n-3}\right)^{1/2} \hat{a}_t$$

Step3: Randomly sample $\tilde{a}_k$ from $\{\ddot{a}_t\}$, then

$$\tilde{y}_2 = \hat{\mu} + \widehat{\phi_1}y_1 + \widehat{\phi_2}y_0 + \tilde{a}_2$$

$$\tilde{y}_3 = \hat{\mu} + \widehat{\phi_1}\tilde{y}_2 + \widehat{\phi_2}y_1 + \tilde{a}_3$$

$$\tilde{y}_k = \hat{\mu} + \widehat{\phi_1}\tilde{y}_{k-1} + \widehat{\phi_2}\tilde{y}_{k-2} + \tilde{a}_2, \;\; k > 3$$

2.3 Steps of Block bootstrap

Step1: Determine block size b, and construct n-b+2 overlapping blocks as

$$B_0 = (y_0 \quad \cdots \quad y_{b-1})$$

$$B_1 = (y_1 \quad \cdots \quad y_b)$$

$$B_2 = (y_2 \quad \cdots \quad y_{b+1})$$

…

$$B_{n-b+1} = (y_{n-b+1} \quad \cdots \quad y_n)$$

Step2: Randomly sample from B_0 , …, B_{n-b+1} .

2.4 Pros and Cons of above method

Parametric bootstrap:

Pros: If the assumption of model distribution is correct, then bootstrap data is accurate in doing inference;

Cons: If the assumption of model distribution is invalid, then bootstrap data leads to asymptotically invalid inference.

Sieve bootstrap:

Pros: Minimum assumptions on model distribution, i.e., only require innovations (errors) to be I.I.D.

Cons: Can't handle heteroskedasticity in the data.

Block bootstrap:

Pros: Can handle heteroskedasticity in the data, can provide higher order accuracy in doing inference;

Cons: Choice of block size b is critical.

3. Modeling seasonality

3.1 Define the seasonal autoregressive integrated moving average (SARIMA) model;

Answer:

SARIMA(p,d,q)(P,D,Q)$_S$ model is

$$\Phi_P(B^S)\phi_p(B)(1-B)^d(1-B^S)^D y_t = \Theta_Q(B^S)\theta_q(B)\epsilon_t$$

Where $\{\epsilon_t\}$ is a white noise process with mean 0 and variance σ^2, and

$$\phi_p(z) = 1 - \phi_1 z - \cdots - \phi_p z^p$$

$$\theta_q(z) = 1 + \theta_1 z + \cdots + \theta_q z^q$$

$$\Phi_P(z^S) = 1 - \Phi_1 z^S - \cdots - \Phi_P z^{SP}$$

$$\Theta_q(z^S) = 1 + \Theta_1 z^S + \cdots + \Theta_Q z^{SQ}$$

(To ensure stationarity, we need all roots of $\phi_p(z) = 0$ and $\Phi_P(z^S) = 0$ are not on unit circle, i.e., $|z| \neq 1$; To ensure causality, we need all roots of $\phi_p(z) = 0$ and $\Phi_P(z^S) = 0$ are not on unit circle, i.e., $|z| > 1$; To ensure invertibility, we need all roots of $\theta_q(z) = 0$ and $\Theta_Q(z^S) = 0$ are not on unit circle, i.e., $|z| > 1$.)

3.2 Define the periodic autoregressive (PAR) model;

Answer:

PAR(p) model is

$$y_{s,n} = \mu_s + \phi_{s,1} y_{s-1,n} + \cdots + \phi_{s,p} y_{s-p,n} + \epsilon_{s,n}$$

Where $s = 1, \dots, S$, denotes season; $n = 1, \dots, N$ denotes the year; $\{\epsilon_{s,n}\} \sim NID(0, \sigma^2)$; $y_{i,n} = y_{S+i,n-1}$ when $i \leq 0$; p is the maximum order of all $s = 1, \dots, S$.

(To ensure stationarity, for all $s = 1, \dots, S$, we need all roots of $\phi_{s,p}(z) = 0$ are not on unit circle, i.e., $|z| \neq 1$; To ensure causality, for all $s = 1, \dots, S$, we need all roots of $\phi_{s,p}(z) = 0$ are not on unit circle, i.e., $|z| > 1$.)

3.3 Define the periodic moving average (PMA) model.

PMA(q) model is

$$y_t = \delta_s + \theta_{s,1}\epsilon_{t-1} + \cdots + \theta_{s,q}\epsilon_{t-q}$$

Where $t = 1, \dots, S \times 1, S \times 1 + 1, \dots S \times 2, \dots, S \times N$, S is number of seasons and N is number of years; $\{\epsilon_t\} \sim NID(0, \sigma^2)$; q is the maximum order of all $s = 1, \dots, S$.

4. State Space Model (SSM)

Observation equation (Measurement equation):

$$Y_t = G_t X_t + W_t$$

Where $\{Y_t\}$ is n-dimensional observation time series, $\{X_t\}$ is m-dimensional state variable time series, $\{W_t\}$ is n-dimensional white noise process, e.g., $\{W_t\} \sim NID_n(\vec{0}, \Sigma_W)$.

State equation:

$$X_t = F_{t-1} X_{t-1} + V_t$$

Where $\{V_t\}$ is m-dimensional white noise process, e.g., $\{V_t\} \sim NID_n(\vec{0}, \Sigma_V)$.

4.1 Express

$$y_t - \phi_1 y_{t-1} - \phi_2 y_{t-2} = a_t + \theta_1 a_{t-1} + \theta_2 a_{t-2}$$

$$a_t \sim NID(0,1)$$

as a state space model.

Answer:

Observation equation (Measurement equation):

$$y_t = [1 \quad 0 \quad 0 \quad 0]\begin{pmatrix} y_t \\ y_{t-1} \\ a_t \\ a_{t-1} \end{pmatrix}$$

State equation:

$$\begin{pmatrix} y_t \\ y_{t-1} \\ a_t \\ a_{t-1} \end{pmatrix} = \begin{bmatrix} \phi_1 & \phi_2 & \theta_1 & \theta_2 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}\begin{pmatrix} y_{t-1} \\ y_{t-2} \\ a_{t-1} \\ a_{t-2} \end{pmatrix} + \begin{pmatrix} a_t \\ 0 \\ a_t \\ 0 \end{pmatrix}$$

4.2 Express

$$y_t = \alpha + \sum_{i=1\ to\ p} \beta f_{it} + a_t$$

$$a_t \sim NID(0,1)$$

as a state space model.

Answer:

Observation equation (Measurement equation):

$$y_t = [1 \quad f_{1t} \quad \cdots \quad f_{pt}]\begin{pmatrix} \alpha \\ \beta \\ \vdots \\ \beta \end{pmatrix} + a_t$$

State equation:

$$\begin{pmatrix} \alpha \\ \beta \\ \vdots \\ \beta \end{pmatrix} = x_t$$

$$x_t = x_{t-1}$$

4.3 Review the structural time series model in the course note.

Please review slides 9 to 18 of TSMwithGoogleData(29July2017).pdf.

5. ARCH/GARCH process

5.1 Define the autoregressive conditional heteroskedasticity (ARCH) process;

Answer:

ARCH(p) model assumes that

$$X_t = \sigma_t Z_t$$

$$\sigma_t^2 = w_0 + w_1 X_{t-1}^2 + \cdots + w_p X_{t-p}^2$$

Where $\{Z_t\}$ is a sequence of independent and identically distributed random variables with mean zero and variance 1, $w_0 > 0$, and $w_i \geq 0$ for $i = 1, \dots, p$.

5.2 Define the generalized autoregressive conditional heteroskedasticity (GARCH) process.

Answer:

GARCH(p,q) model assumes that

$$X_t = \sigma_t Z_t$$

$$\sigma_t^2 = w_0 + \sum_{i=1 \ to \ p} w_i X_{t-i}^2 + \sum_{j=1 \ to \ q} \eta_j \sigma_{t-j}^2$$

Where $\{Z_t\}$ is a sequence of independent and identically distributed random variables with mean zero and variance 1, $w_0 > 0$, $w_i \geq 0$ for $i = 1, \dots, p$, and $\eta_j \geq 0$ for $j = 1, \dots, q$.