Try to play around with ~~R code some times.~~

Time series analysis review

LECTURE 1

Time series: a collection of r.v.'s indexed according to the order they obtained in time.

E.g's of t.s. models:

① IID noises with finite second moment.

$\{X_t\}$ iid, mean -0, $\sigma^2 < \infty$.

$X_t \sim IID(0, \sigma^2)$

② White noise

$E(X_t) = 0$, $\sigma^2 < \infty$,

$$\gamma_X(r,s) = \begin{cases} \sigma^2, & r=s \\ 0, & o.w. \end{cases}$$

$X_t \sim WN(0, \sigma^2)$

③ random walk

$X_t = X_{t-1} + a_t$, $a_t \sim NID(0, \sigma^2)$

④ first-order autoregressive progress.

$X_t - \phi X_{t-1} = a_t$, $a_t \sim WN(0, \sigma^2)$

a_t is not correlated with X_s, $\forall s<t$.

⑤ Moving averaging

$X_t = \frac{1}{3}(a_t + a_{t-1} + a_{t-2})$, $a_t \sim WN(0, \sigma^2)$

T.S. = Trend + seasonal / cyclic variation + random noise.

$X_t = m_t + S_t + Y_t$.

$EY_t = 0$

$S_{t+d} = S_t$

$\sum_{j=1}^{d} S_j = 0$

aim: extra m_t, S_t, hope $\{Y_t\}$ be (weakly) stationary

normal distribution
strict = weak

- Two types of stationarity:
  - $\boxed{\text{strictly:}}$

    (defined by the distributional property of a stochastic process (time series).)

    One for which the probabilistic behavior of every collection of values $(X_{t+1}, \cdots, X_{t+n})$ is identical to that of the time shifted set $(X_{t+h+1}, \cdots, X_{t+h+n})$.

    i.e.

    $$P\{X_{t+1} \leq C_1, \cdots, X_{t+n} \leq C_n\}$$
    $$= P\{X_{t+1+h} \leq C_1, \cdots, X_{t+n+h} \leq C_n\}$$

    $\forall t, n \in \mathbb{Z}, n > 0, C_1, \cdots, C_n \in \mathbb{R}$.

    and all time shifts $h = 0, \pm 1, \pm 2, \cdots$

  - properties not affected by a change in the time origin.

- def. $X_t = (x_t, x_{t+1}, \cdots, \cancel{x})^T$

  $Y_t = (y_t, y_{t+1}, \cdots)^T$

  $$\left.\begin{array}{l} \mu_t^x = E(\cancel{\ } x_t) \\ \mu_t^y = E(y_t) \end{array}\right\} \forall t \in \mathbb{Z}$$

  $\gamma(t,s) = E[(x_t - \mu_t)(\overset{X_t}{\cancel{x_s}} - \mu_s)] \quad \forall t, s \in \mathbb{Z}$

autocorrelation: $\rho(t,s) = \gamma(t,s) / \sqrt{\gamma(t,t)\gamma(s,s)}$

$\gamma_{xy}(t,s) = E[(x_t - \mu_t^x)(y_s - \mu_s^y)]$

crosscorrelation: $\rho_{xy}(t,s) = \gamma_{xy}(t,s) / \sqrt{\gamma_x(t,t) \cdot \gamma_y(s,s)}$

For stationary process, denote $\gamma(\cancel{s}\, t, s) = \gamma(t, t+h)$ where $s = t+h$.

· positive autocorrelation → "persistance" tendency to remain in the same state.

$\boxed{\text{weakly:}}$

$E(X_t) = \mu$, $Var(X_t) = \gamma(0) = \sigma^2$ are constant & does not depend on time t for all $t \in \mathbb{Z}$.

The autocovariance function $\gamma(t, t+h)$ independent of time
"covariance-stationary / stationary in the wide sense".

$\gamma(t, t+h) = \gamma(h)$

$\rho(t, t+h) := \rho(h)$

$\gamma(0) = var(X_t)$ ✗

So, it's true that:

strictly: $f(x_1, x_2) = f(x_2, x_3) = \cdots = f(x_t, x_{t+1}) = \cdots$

weakly: $cov(x_1, x_2) = cov(x_2, x_3) = \cdots = cov(X_t, X_{t+1}) = \cdots$

SACF (Sample autocorrelation functions)

plotting $\hat{\rho}(h)$ against lag $h = 1, 2, \cdots, M$ ($M \ll$ series length)

· $\hat{\gamma}(h) = \frac{1}{n} \sum_{t=1}^{n-h} (X_{t+h} - \bar{x})(X_t - \bar{x})$ with $\hat{\gamma}(-h) = \hat{\gamma}(h)$ for $h = 0, 1, \cdots, n-1$

· $\hat{\rho}(h) = \frac{\hat{\gamma}(h)}{\hat{\gamma}(0)}$

SACF of T.S. with linear time trends:

$\hat{\rho}_k$ will not $\to 0$ except lag h is $\uparrow$.

seasonal trend ⇒ SACF exhibits oscillation at the same frequency

Steps to time-series modeling.

① Plot the time series & check for
- trend, seasonal & other cyclic components, any apparent sharp changes in behaviour, as well as any outlying observations.

② Remove trend & seasonal components to get residuals.

③ Chose a model to fit the residuals

④ Forecasting can be carried out by forecasting residual & then inverting the transformation carried out ~~to~~ in step ②

LECTURE 2

Supplement
- Partial autocorrelation functions (PACF)
- Yule-Walker equations
- Solving Yule-walker equations — Cramer's rule & Durbin-Levinson algorithm.

BOX-JEKINS
- Autoregressive & moving average model. (ARMA)
- 3 stages of the Box-Jenkins approach
- Model identification
- Model adequacy
- Model selection

By "classical decomposition": T.S. = ○ + ○ + ○

First two can be filtered out.
noise cannot be.

We use ARMA to model the noise.

So we gonna learn how to model the irregular component as a linear difference equation ~~between~~ with a stochastic forcing process. i.e autogressive moving average model. (ARMA)

A process $\{X_t\}$ is said to be an ARMA(p.q) process if $\{X_t\}$ is stationary & $\forall t$

$$X_t - \phi_1 X_{t-1} - \cdots - \phi_p X_{t-p} = a_t + \theta_1 a_{t-1} + \cdots + \theta_q a_{t-q}$$

where $a_t \sim WN(0, \sigma^2)$.

- $\{X_t\}$ is ARMA(p.q) with mean μ if $\{X_t - \mu\}$ is ARMA(p.q)

compact notation:

$$\Phi(B)(X_t - \mu) = \Theta(B) a_t,$$

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

$\Theta(B) = 1 + \theta_1 z + \cdots + \theta_q z^q$ B: backward Shift operator

- Theoretical support of ARMA model

$\boxed{\text{Wold decomposition}}$

Any zero-mean process $\{X_t\}$ which is not $\boxed{\text{deterministic}}$ can be expressed as a sum $X_t = U_t + V_t$ of an $\boxed{\text{MA}(\infty)\text{ process}}$ $\{U_t\}$ & a $\boxed{\text{deterministic}}$ $\{V_t\}$ which is uncorrelated with $\{U_t\}$.

$\boxed{\text{Deterministic process}}$

~~If the process~~ values X_{n+j}, $j \geq 1$ of the process $\{X_t, t = 0, \pm 1, \pm 2, \cdots\}$ were perfectly predictable in terms of $\mathcal{M}_n = \overline{sp}\{X_t, -\infty < t \leq n\}$. Such processes are called deterministic.

- If X_n is from a deterministic process it can be predicted by its past observations of the process.

e.g. $\{X_t\}$ is nondeterministic stationary process, mean is zero.

$$X_t = \underbrace{\sum_{j=0}^{\infty} \psi_j a_{t-j}}_{\text{Moving average process: linear combination of past innovations } \{a_t\}} + \underbrace{V_t}_{\mathcal{M}_n = \overline{sp}\{X_t, -\infty < t \leq n\} \text{ Autoregressive process}} = \psi(L) a_t + V_t$$

- 1. $\psi_0 = 1$ & $\sum_{j=0}^{\infty} \psi_j^2 < \infty$
- 2. $\{a_t\} \sim WN(0, \sigma^2)$, $\sigma^2 > 0$
- 3. $cov(a_s, V_t) = 0$, $\forall s, t = 0, \pm 1, \pm 2, \ldots$
- 4. V_t is deterministic

- Special forms of ARMA models.

① $AR(p)$

$$X_t - \phi_1 X_{t-1} - \cdots - \phi_p X_{t-p} = a_t \ , \ a_t \sim WN(0, \sigma^2)$$

② $MA(q)$

$$X_t = a_t + \theta_1 a_{t-1} + \cdots + \theta_q a_{t-q} \ , \ a_t \sim WN(0, \sigma^2)$$

$\boxed{MA(q)}$

$$X_t = a_t + \theta_1 a_{t-1} + \cdots + \theta_q a_{t-q}$$
$$= \Theta(B) a_t$$

$$\boxed{\begin{array}{l} \Theta(B) = 1 + \theta_1 B + \cdots + \theta_q B^q \\ B^h X_t = X_{t-h} \end{array}}$$

$$a_t \sim WN(0, \sigma^2)$$

Check stationarity: e.g. $MA(1)$

① check mean constant

$$E(X_t) = E(a_t + \cancel{\theta_q a_{t-q}}\ \theta_1 a_{t-1}) = 0 \quad ✓$$

② Check variance constant & exists.

$$Var(X_t) = \gamma(0) = Cov(a_t + \theta_1 a_{t-1}, a_t + \theta_1 a_{t-1})$$
$$= E(a_t^2 + 2\theta a_t a_{t-1} + \theta_1^2 a_{t-1}^2) - E(a_t + \theta_1 a_{t-1}) E(a_t + \theta a_{t-1})$$
$$= (\sigma^2 + 0 + \theta^2 \sigma^2) - 0$$
$$= (1 + \theta^2)\sigma^2 < \infty \quad ✓$$

③ $\gamma(1) = Cov(a_t + \theta a_{t-1}, a_{t-1} + \theta a_{t-2})$

$$= E(a_t a_{t-1} + (a_t a_{t-2} + a_{t-1}^2)\theta + a_{t-2} a_{t-1} \theta^2) - E(a_t + \theta a_{t-1}) E(a_{t-1} + \theta a_{t-2})$$
$$= E(0 + (0 + \sigma^2)\theta + 0) - 0 \cdot 0$$
$$= \sigma^2 \theta \quad ✓$$

④ $\gamma(2) = 0$, $\ldots$ For $h > q$, $\gamma(h) = 0$. ✓

So stationary.

MA(∞) process

Thm The MA(∞) process is stationary with mean zero & autocovariance function $\gamma(k) = \sigma^2 \sum_{j=0}^{\infty} \psi_j \psi_{j+|k|}$

- Conclusion:

ACF of MA(q) processes

$$X_t = a_t + \cdots + \theta_q a_{t-q}, \quad a_t \sim WN(0, \sigma^2)$$

$$\psi_j = \begin{cases} \theta_j & j = 0, 1, \ldots, q \\ 0 & , j > q \end{cases}$$

$$\gamma(k) = Cov(X_t, X_{t+|k|})$$

$$= \begin{cases} \sigma^2 \sum_{i=0}^{q-k} \theta_i \theta_{i+|k|} & , k = 0, \pm 1, \ldots, \pm q \\ 0 & \text{o.w.} \end{cases}$$

$\boxed{AR(P)}$ Autoregressive model of ~~a~~ order p.

$$X_t - \phi_1 X_{t-1} - \cdots - \phi_p X_{t-p} = a_t = \Phi(B) X_t$$

$a_t \sim WN(0, \sigma^2)$

$B^h X_t = X_{t-h}, \quad h \in \mathbb{Z}$

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

AR(1)

$$X_t = \phi X_{t-1} + a_t$$
$$= \phi(\phi X_{t-2} + a_{t-1}) + a_t$$
$$= \phi^2 X_{t-2} + \phi a_{t-1} + a_t$$
$$= \sum_{j=0}^{\infty} \phi^j a_{t-j} \implies \text{MA}(\infty) \text{ process}$$

$$\gamma(k) = cov(X_t, X_{t+k}) = Cov(\sum_{l=0}^{\infty} \phi^l a_{t-l}, \sum_{j=0}^{\infty} \phi^j a_{t+k-j})$$
$$= \cancel{\phi^k \gamma(0) + Cov(X_t}$$
$$= Cov(X_t, \phi^k X_t + \sum_{j=0}^{k} \phi^j a_{t+k-j}$$
$$= \phi^k \gamma(0) + \underbrace{Cov(X_t, \sum_{j=0}^{k} \phi^j a_{t+k-j})}_{0}$$
$$= \phi^k \gamma(0)$$

$k=0$

$\Rightarrow \gamma(0) = Var(X_t) = E(\sum_{j=0}^{\infty} \phi^j a_{t-j}, \sum_{j=0}^{\infty} \phi^j a_{t-j})$

$$= \sigma^2 \sum_{j=0}^{\infty} (\phi^j)^2$$

$$= \frac{\sigma^2}{1-\phi^2} \quad ☆$$

finite if $|\phi| < 1$

~~Remarks~~

Another:

$$X_t = \phi^{-1} X_{t+1} - \phi^{-1} a_{t+1}$$

$$= \phi^{-1}(\phi^{-1} X_{t+2} - \phi^{-1} a_{t+2}) - \phi^{-1} a_{t+1}$$

$$= \cdots$$

$$= -\sum_{j=1}^{\infty} \phi^{-j} a_{t+j}$$

$$Var(X_t) = \sigma^2(\phi^{-2} + \phi^{-4} + \cdots)$$

$$= \phi^{-2} \frac{\sigma^2}{1-\phi^{-2}} < \infty \text{ if } |\phi| > 1$$

- Remark:

For AR(1), we check whether it is stationary via finding its MA(∞) representation & checking its ACF.

☆ **General approach to check stationarity**

A general way:

is that the roots of $\Phi(B) = 1 - \phi_1 B - \cdots - \phi_p B^p = 0$ lie outside the unit circle.

For AR(2)

$$1 - \phi_1 B - \phi_2 B^2 = 0$$

$$\phi_2 B^2 + \phi_1 B - 1 = 0$$

$$B = \frac{-\phi_1 \pm \sqrt{\phi_1^2 + 4\phi_2}}{2\phi_2} \qquad |B| = \left| \frac{-\phi_1 \pm \sqrt{\phi_1^2 + 4\phi_2}}{2\phi_2} \right| > 1$$

Causal & Invertible process

Causal process

(previous terms)

$\{X_t\}$ can be expressed ~~by~~ in terms of $\{a_s, s \le t\}$

Such processes ~~is~~ are called causal or future independent autoregressive process.

- able to use ACF/PACF b/c it's stationary.

Invertible process

☆ In general, no restrictions on $\{\theta_i\}$ are required for a finite order MA process to be stationary.

The imposition of the invertibility condition ensures that $\exists$ a unique MA process for a given set of ACF.

Causal/stationary $AR(p) \rightarrow MA(\infty)$

Invertible ~~AR~~ $MA(q) \rightarrow AR(\infty)$

Causal ARMA process

An ARMA(p,q) process, defined by $\Phi(B)X_t = \Theta(B)a_t$ is causal if $\exists$ a sequence of constants $\{\psi_j\}$ such that $\sum_{j=0}^{\infty}|\psi_j| < \infty$ and $X_t = \sum_{j=0}^{\infty}\psi_j a_{t-j}$, $t = 0, \pm1, \pm2, \cdots$

☆ $$\Psi(z) = \sum_{j=0}^{\infty}\psi_j z^j = \frac{\Theta(z)}{\Phi(z)}, \quad |z| \le 1$$

$$\Psi(B)\Phi(B) = \Theta(B)$$

$$\Leftrightarrow \left(\sum_{j=0}^{\infty}\psi_j B^j\right)\left(\sum_{k=0}^{p}\phi_p B^p\right) = \sum_{l=0}^{q}\theta_l B^l$$

$$(\psi_0 + \psi_1 B + \psi_2 B^2 + \cdots)(1 - \phi_1 B - \cdots \phi_p B^p) = 1 + \theta_1 B + \cdots + \theta_q B^q$$

待定係數 $B^0: \psi_0 = 1$

$B^1: \cdots\cdots$

Invertible ARMA ~~model~~ process:

An ARMA(p, q) process is said to be invtble if $\exists$ a sequence of constants $\{\pi_j\}$ s.t. $\sum_{j=0}^{\infty}|\pi_j|<\infty$, $a_t=\sum_{j=0}^{\infty}\pi_j X_{t-j}$, $t=0,\pm1,\pm2,\cdots$

similarly we do 待定系数:

$$\Pi(z)=\sum_{j=0}^{\infty}\pi_j\cdot z^j=\frac{\Phi(z)}{\Theta(z)}\quad,|z|\leq 1$$

$$\Pi(B)\cdot\Theta(B)=\Phi(B)$$

$$\Leftrightarrow (\sum_{j=0}^{\infty}\pi_j B^j)(\sum_{l=0}^{q}\theta_l B^l)=\sum_{k=0}^{p}\phi_k B^k \text{ actually 负的}$$

$$(\pi_0+\pi_1 B+\cdots)(1+\theta_1 B+\cdots+\theta_q B^q)=(1-\phi_1 B-\ldots-\phi_p B^p)$$

B^0: ...

B^1: ...

...

Supplement Material

- PACF
- Yule-Walker equations
- Solve YW eq'tn. : Cramer's rule & Durbin-Levinson algorithm.

- The maximum lag of the non-zero sample autocorrelation is a good indicator of the MA(q) processes.

$$Y_t = e_t + \theta_1 e_{t-1} + \theta_2 e_{t-2} + \cdots + \theta_q e_{t-q}$$, lag q ~~cutt~~ cut off

$$\rho_k = \begin{cases} \dfrac{\theta_k + \theta_1\theta_{k+1} + \cdots + \theta_{q-k}\theta_q}{1 + \theta_1^2 + \cdots + \theta_q^2} = \dfrac{\gamma(k)}{\gamma(0)}, & k = 1, \cdots, q \\ 0, & k > q. \end{cases}$$

- The conditional correlation

$$\phi_{kk} = corr(X_t, X_{t+k} \mid X_{t+1}, X_{t+2}, \cdots, X_{t+k-1})$$

is usually referred to as the PACF in t.s. analysis.

- PACF can also be defined as the correlation between 2 prediction errors at lag k

$$\phi_{kk} = Corr(X_t - \beta_1 X_{t-1} - \cdots - \beta_{k-1} X_{t-2}, X_{t+k} - \beta_1 X_{t+k+2} - \cdots \beta_{k-1} X_{t+1})$$

β's are minimizer of MSE of prediction.

Yule-Walker equations.

A general method to find PACF for any stationary process with autocorrelation function ρ_k is :

For a given lag k,

ϕ_{kk} satisfies the YW equation:

$$\rho_j = \phi_{k1}\rho_{j-1} + \phi_{k2}\rho_{j-2} + \cdots + \phi_{kk}\rho_{j-k} \quad , j = 1,2,\ldots,k$$

so we can solve ϕ_{kk} by knowing $\rho_1, \ldots, \rho_k$.

If AR(p), then $\phi_{pp} = \phi_p$

and $\phi_{kk} = 0$ for $k > p$.

Derivation:

$$X_t = \phi_{k1}X_{t-1} + \phi_{k2}X_{t-2} + \cdots + \phi_{kk}X_{t-k} + e_t$$

$$X_t^2 = \phi_{k1}X_tX_{t-1} + \phi_{k2}X_tX_{t-2} + \cdots + \phi_{kk}X_tX_{t-k} + e_tX_t$$

$\vdots$

$$X_{t-k}X_t = \phi_{k1}X_{t-1}X_{t-k} + \cdots + \phi_{kk}X_{t-k}X_{t-k} + e_tX_{t-k}$$

take E( )

$$\gamma(0) = \phi_{k1}\gamma(1) + \cdots + \phi_{kk}\gamma(k) + \sigma_e^2$$

$$\gamma(1) = \phi_{k1}\gamma(0) + \phi_{k2}(1) + \cdots + \gamma(k-1) + 0$$

$$\gamma(k) = \phi_{k1}\gamma(k-1) + \cdots + \gamma(0) + 0$$

have k+1 equations with k unknowns. order is 1.

Solve to get ϕ_{kk}.

Matrix Form : YW

$$\begin{pmatrix} \gamma(1) \\ \gamma(2) \\ \vdots \\ \gamma(k) \end{pmatrix} = \begin{pmatrix} \gamma(0) & \gamma(1) & \cdots & \gamma(k-1) \\ \gamma(1) & \gamma(0) & \cdots & \gamma(k-2) \\ \vdots & \vdots & \ddots & \vdots \\ \gamma(k-1) & \gamma(k-2) & \cdots & \gamma(0) \end{pmatrix} \begin{pmatrix} \phi_{k1} \\ \phi_{k2} \\ \vdots \\ \phi_{kk} \end{pmatrix}$$

similar b/c $\rho(k) = \frac{\gamma(k)}{\gamma(0)}$

$$\begin{pmatrix} \rho(1) \\ \rho(2) \\ \vdots \\ \rho(k) \end{pmatrix} = \begin{pmatrix} 1 & \rho(1) & \cdots \\ \rho(1) & 1 & \\ \vdots & & \ddots \\ \rho(k-1) & & & 1 \end{pmatrix} \begin{pmatrix} \phi_{k1} \\ \phi_{k2} \\ \vdots \\ \phi_{kk} \end{pmatrix}$$

Cramer's Rule: $$\phi_{kk} = \frac{\det\begin{pmatrix} 1 & \rho_1 & \cdots & \rho_{k-2} & \rho_1 \\ \rho_1 & 1 & \cdots & \rho_{k-3} & \rho_2 \\ \vdots & & & & \vdots \\ \rho_{k-1} & \rho_{k-2} & \cdots & \rho_1 & \rho_k \end{pmatrix}}{\det\begin{pmatrix} 1 & \rho_1 & \cdots & \rho_{k-2} & \rho_{k-1} \\ \rho_1 & 1 & \cdots & \rho_{k-3} & \rho_{k-2} \\ \vdots & & & & \vdots \\ \rho_{k-1} & \cdots & & \rho_1 & 1 \end{pmatrix}} = \cdots$$

$\rho(l) = \rho_l, \forall l$

e.g. AR(2)

$\rho_k = \phi_1 \rho_{k-1} + \phi_2 \rho_{k-2} \quad \forall k \geq 1$

lag 1, k=1

$X_t = \phi_{11} X_{t-1} + a_t$

$X_{t-1} X_t = \phi_{11} X_{t-1}^2 + a_t X_{t-1}$

$\gamma(1) = E(X_{t-1} X_t) = \phi_{11} \gamma(0) + 0$

$\phi_{11} = \frac{\gamma(1)}{\gamma(0)} = \rho_1$

$\rho_1 = \phi_1 \rho(0) + \phi_2 \rho_1$

$(1-\phi_2)\rho_1 = \phi_1$

$\rho_1 = \frac{\phi_1}{1-\phi_2}$

lag 2, k=2

$X_t = \phi_{21} X_{t-1} + \phi_{22} X_{t-2} + a_t$

$\gamma(1) = \phi_{21}\gamma(0) + \phi_{22}\gamma(1) + 0$

$\gamma(2) = \phi_{21}\gamma(1) + \phi_{22}\gamma(0) + 0$

$\rho_1 = \phi_{21} + \phi_{22}\rho_1$

$\rho_2 = \phi_{21}\rho_1 + \phi_{22}$

$\phi_{22} = \frac{\rho_2 - \rho_1^2}{1-\rho_1^2} = \phi_2$

$\phi_{33} = 0$, $\phi_{kk} = 0$ for $k > 2$.

General Idea $\boxed{\text{Solving ACF of autoregressive process using YW eqn's.}}$

$$X_t = \phi X_{t-1} + a_t \qquad (*)$$

times

X_{t-1}, where the lag $k=1$.

$$X_t X_{t-1} = \phi X_{t-1} X_{t-1} + a_t X_t$$

take $E(\ )$

$$\gamma(1) = \phi \gamma(0) + 0$$

$k=2$.

times X_{t-2}

$$\gamma(2) = \phi \cdot \gamma(1) + 0$$

$$= \phi \, \phi \, \gamma(0)$$

$$= \phi^2 \gamma(0)$$

So $\gamma(k) = \phi^k \gamma(0)$

cut off : out of range $\pm$

tails off : goes to 0

Back to Lecture 2.

Important Table.

Process	ACF	PACF
AR(p)	Tails off as exp decay or damped sine wave	Cuts off after lag p
MA(q)	Cuts off after lag q	Tails off as exp/damped
ARMA(p,q)	Tails off after lag (q−p)	tails off after lag (p−q)

MODEL ADEQUACY

(Step 3 in B-J framework)

Check if residuals are approximately uncorrelated.

Goodness of fit tests: "portmanteau tests"

Box & Pierce:

$$Q_{BP} = n \cdot \sum_1^m \hat{\rho}_k^2 \sim \chi^2_{m-(p+q)}$$

Ljung & Box

$$Q_{LB} = \sum_1^m \frac{n(n+2)}{(n-k)} \hat{\rho}_k^2 \sim \chi^2_{m-(p+q)}$$

same for $n \nearrow$

MODEL SELECTION

method
- AIC $AIC = -2\log ML + 2k$
- BIC $BIC = -2\log ML + k\log(n)$

selects the smaller model with smaller AIC/BIC

ML : maximum likelihood.

LECTURE 3

$\boxed{\text{Nonstationary time series \& Forecasting}}$

Plot data

$\boxed{ARIMA(p, d, q)}$

A t.s. $\{X_t\}$ is said to follow an integrated autoregressive moving average model if the dth difference

$$W_t = (1-B)^d X_t \text{ is a stationary ARMA model.}$$

if $\{W_t\}$ follows $ARMA(p,q)$ model

$\Rightarrow$ $\{X_t\}$ is an $ARIMA(p, d, q)$ model.

$$(1-B)^d \Phi(B) X_t = \Theta(B) a_t$$

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

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

$$a_t \sim WN(0, \sigma^2)$$

write

$$\Phi(B) W_t = \Theta(B) a_t$$

where $W_t = (1-B)^d X_t$

$\boxed{\text{Differencing to remove } \text{~~time/seasonal~~ time Trend}}$

e.g. $Y_t = \underbrace{a + bt + ct^2}_{\text{time trend}} + X_t$, X_t is stationary t.s.

- $B \;:\Rightarrow BY_t = Y_{t-1}$, $Bt = t-1$, $Bc = c$
- $\nabla^d = (1-B)^d$, $\nabla^2 = (1-B)(1-B)$

~~(1-B) remove~~ test $\left.\begin{matrix} a \\ bt \\ \& ct^2 \end{matrix}\right\}$ all removed.

Differencing at lag d to remove seasonal component

~~this~~ is different from ∇^d

define $\nabla_d X_t = X_t - X_{t-d} = (1-B^d)X_t$

- applying ∇_d to classical decompoition model

$X_t = m_t + s_t + Y_t$

(s_t has period d)

$\nabla_d X_t = m_t - m_{t-d} + Y_t - Y_{t-d}$ NO SEASONAL TREND!

Nonstationarity in Variance

- Differencing (trend-removing) can be used to reduce a homogeneous nonstationary t.s. to a stationary (trend stationary) t.s.

Transformation for nonstationarity in var. (skipped)

I(d).

? skipped, but in final..

Dickey-Fuller unit root test

: ① remove deterministic time trend
② conduct statistical inference using the test.

- Consider a regression test on $X_t = \phi X_{t-1} + a_t$, $a_t \sim NID(0, \sigma^2)$
- $\Delta X_t = (\phi - 1) X_{t-1} + a_t = \pi X_{t-1} + a_t$
- $H_0 : \pi = 0$
  H_a : a trend strationary process

...

Augmented Dickey-Fuller test

: replace AR(1) process for $\{X_t\}$ with ARMA(p,q) or AR(p)

conclusion: the serial correlation in error terms is removed.

$\boxed{\text{FORECASTING}}$ A large topic.

ARIMA (p, d, q)

$$\varphi(B) X_t = \theta(B) a_t \qquad (1)$$
$$\varphi(B) = \phi(B)(1-B)^d$$
$$\phi(B) = 1 - \phi_1 B - \cdots - \phi_p B^p$$
$$\theta(B) = 1 + \theta_1 B + \cdots + \theta_q B^q$$

- Forecast X_{t+l}, $l \geq 1$

③ explicit forms of ARIMA model

① Difference equation form:
directly in terms of difference equation

every (1) has $\Rightarrow$ $$X_{t+l} = \varphi_1 X_{t+l-1} + \cdots + \varphi_{p+d} X_{t+l-p-d} + a_{t+l} + \theta_1 a_{t+l-1} + \cdots + \theta_q a_{t+l-q}$$

② Integrated form:
as an infinite $\underset{\text{weighted}}{\text{sum}}$ of current & previous shocks:

$$X_{t+l} = \sum_0^\infty \psi_j a_{t+l-j} \qquad (2)$$

where $\{\psi\}$ may be calculated.

rewrite (2)

$$X_{t+l} = a_{t+l} + \psi_1 a_{t+l-1} + \cdots + \psi_{l-1} a_{t+1} + C_t(l)$$

well $C_t(l) = \sum_l^\infty \psi_j a_{t+l-j}$

③ Weighted average of previous observations

The forecast error for lead time l is
$e_t(l) = a_{t+l} + \psi_1 a_{t+l-1} + \cdots + \psi_{l-1} a_{t+1}$
$E_t(e_t(l)) = 0$ unbiased
$Var(e_t(l)) = (1+\psi_1^2 + \cdots + \psi_{l-1}^2)\sigma^2$

$\boxed{\text{Min MSE Forecast}}$

Sps the best forecast at t

$\hat{X}_t(l) = \psi_l^* a_t + \psi_{l+1}^* a_{t-1} + \psi_{l+2}^* a_{t-2} + \cdots$
where $*$ terms are to be determined

MSE is

$E(X_{t+l} - \hat{X}_t(l))^2$
$= (1+\psi_1^2 + \cdots + \psi_{l-1}^2)\sigma^2 + \sum_0^\infty (\psi_{l+j} - \psi_{l+j}^*)^2 \sigma^2$

fixed

$\psi_{l+j}^* = \psi_{l+j}$

Rules for Calculating the conditional expectations
Use any 3 explicit form to write down
then apply the following rules:

X_{t-j} $(j=0,1,2,\cdots)$ which have already happened at t, are left unchanged.
X_{t+j} $(j=1,2,\cdots)$ are replaced by $\hat{X}_t(j)$
a_{t-j} already happened, replaced by $X_{t-j} - \hat{X}_{t-j-1}(1)$
a_{t+j} not happened, zeros

$\boxed{\text{Probability limits of the forecast}}$

The variance of l-step-ahead forecast error for any origin t is the expected value of $e_t^2(l) = \{X_{t+l} - \hat{X}_t(l)\}^2$
& is given by $(1+\sum_{j=1}^{l-1} \psi_j^2)\sigma^2$

2014 Practice Questions

Q1. Random Walk

$Y_t = Y_{t-1} + e_t$, $e_t \sim NID(0, \sigma^2)$.

① Let $Y_1 = e_1$

$Y_0 = 0$

then $Y_2 = Y_1 + e_2 = e_1 + e_2$

so ... $Y_t = e_1 + \cdots + e_t$

$E(Y_t) = E(e_1 + \cdots + e_t) = 0$.

② $var(Y_t) = \cancel{E(Y_t^2) - (E(Y_t))^2}$

$= Var(e_1 + \cdots + e_t) = t\sigma^2$

③ $\gamma(t,s) = Cov(Y_t, Y_s) = t\sigma^2$

④ $\rho(t,s) = \dfrac{\gamma(t,s)}{\cancel{\gamma(t,s)}} = \dfrac{t\sigma^2}{\sqrt{t\sigma^2 s\sigma^2}} = \sqrt{\dfrac{t}{s}}$

$\sqrt{\gamma(t,t)\gamma(s,s)}$

⑤ $\lim_{t\to\infty} \rho(h) = 1$

$\rho(h) = \dfrac{\gamma(h)}{\gamma(0)} = \rho(t, t+h) = \sqrt{\dfrac{t}{t+h}}$

as $t \to \infty$, $\sqrt{\dfrac{t}{t+h}} \to 1$

Q2 Moving Average of order 2

$Y_t = 0.5e_t + 0.5e_{t-1}$, $e_t \sim NID(0, \sigma_e^2)$

(1). $E(Y_t) = 0$?

$\mu_t = E(Y_t) = E\left(\frac{e_t + e_{t-1}}{2}\right) = 0 + 0 = 0$

(2). $Var(Y_t) = \frac{Var(e_t) + Var(e_{t-1})}{4} = \frac{1}{2}\sigma_e^2$

(3). $\gamma(t,s) = \{$ 3 situations

$Cov(Y_t, Y_{t+k}) = Var(Y_t) = \frac{1}{2}\sigma_e^2$ when $k=0$

$Cov(Y_t, Y_{t+1}) = Cov(0.5e_t + 0.5e_{t-1}, 0.5e_{t-1} + 0.5e_{t-2})$

$= E(0.25e_t e_{t-1} + 0.25e_{t-1}^2 + 0.25e_t e_{t-2} + 0.25e_{t-1}e_{t-2})$

$= \frac{1}{4}\sigma_e^2$ when $k=1$

$Cov(Y_t, Y_{t+k}) = 0$ when $k>1$

(4). $\rho(t,s) = \begin{cases} \frac{\gamma(t,s)}{\gamma(0)} = 1 & t=0 \\ 0.5 & |t-s|=1 \\ 0 & |t-s|>1 \end{cases}$

Q3. General linear process:

MA(∞).

$$Y_t = a_t + \psi_1 a_{t-1} + \psi_2 a_{t-2} + \cdots = \sum_{j=0}^{\infty} \psi_j a_{t-j}$$

where $\psi_0 = 1$

$\sum_{j=0}^{\infty} |\psi_j| < \infty$, $a_t \sim NID(0, \sigma^2)$

(1). $E(Y_t) = E(a_t + \psi_1 a_{t-1} + \cdots) = 0 + 0 + \cdots = 0$

(2). $\gamma(h) = Cov(Y_t, Y_{t-h})$

$= Cov(a_t + \psi_1 a_{t-1} + \cdots, a_{t-h} + \psi_1 a_{t-h-1} + \cdots)$

$= E(Y_t \cdot Y_{t-h}) - E(Y_t) E(Y_{t-h})$

$= \sigma^2 \sum_{i=0}^{\infty} \psi_i \psi_{i+|h|}$

as the ith term in Y_t and $i+|h|$ term in Y_{t-h} have the same a_{t-i} term

Q4. MA(1).

$X_t = a_t + \theta a_{t-1}$, $a_t \sim NID(0, \sigma^2)$

$X_1 = a_1 + \theta a_0$

$X_2 = a_2 + \theta a_1$

$X_3 = a_3 + \theta a_2$

$Var(X_1 + X_2 + X_3) = E[(X_1 + X_2 + X_3)^2] - [E(X_1 + X_2 + X_3)]^2$

$= E[(\theta a_0 + (1+\theta) a_1 + (1+\theta) a_2 + a_3)^2] - 0$

$= [\theta^2 + (1+\theta)^2 + (1+\theta)^2 + 1] \sigma^2$

$= [\theta^2 + 1 + 2\theta + \theta^2 + 1 + 2\theta + \theta^2 + 1] \sigma^2$

$= (3\theta^2 + 4\theta + 3) \sigma^2$

Q6. ~~AR~~ MA(q) process

$Y_t = \mu + a_t - \theta_1 a_{t-1} - \cdots - \theta_q a_{t-q}, \; a_t \sim NID(0, \sigma^2)$

(1). $E(Y_t) = \mu$

(2). $\gamma(0) = Cov(Y_t, Y_t) = Var(Y_t)$

$= E(Y_t^2) - E(Y_t)^2$

$= \mu^2 + \sigma^2(1+\theta_1^2 + \cdots + \theta_q^2) - \mu^2$

$= \sigma^2(1+\theta_1^2 + \cdots + \theta_q^2)$

(3). $\gamma(h)$

$= Cov(Y_t, Y_{t+h})$

$= \begin{cases} \cancel{\mu^2 - \mu^2} + (-\theta_h + \theta_1\theta_{1+h} + \cdots + \theta_{q-h}\theta_q), & h = 1, \cdots, q \\ 0 & h > q \end{cases}$

(4). Invertible.

$\pi(B) X_t = a_t$

$\frac{\Phi(B)}{\Theta(B)} X_t = a_t$

$\pi(B) = \frac{\Phi(B)}{\theta(B)} \Rightarrow \pi(B)\theta(B) = \phi(B)$

$(1 + \pi_1 B + \pi_2 B^2 + \cdots)(1 - \theta_1 B - \theta_2 B^2 - \cdots - \theta_q B^q) = 1$

$B^1 : \pi_1 - \theta_1 = 0 \quad \pi_1 = \theta_1$

$B^2 : \pi_1 \cdot (-\theta_1) + \pi_2 - \theta_2 = 0$

$\pi_2 - \theta_1^2 - \theta_2 = 0$

$\pi_2 = \theta_2 + \theta_1^2$

$B^3 :$

$B^4 :$

$B^5 :$

Question 6 Stationary AR(2) processes

$$Y_t - \phi_1 Y_{t-1} - \phi_2 Y_{t-2} = \mu + a_t, \quad a_t \sim NID(0, \sigma^2)$$

(1). $E(Y_t) = E(\phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \mu + a_t)$

$= \phi_1 E(Y_{t-1}) + \phi_2 E(Y_{t-2}) + \mu$

mean stationary so $E(Y_t) = \dfrac{\mu}{1-\phi_1-\phi_2}$

(2).

$Y_t = \phi_{21} Y_{t-1} + \phi_{22} Y_{t-2} + \cancel{\mu} + a_t$ WLOG，不看常数项

(3). $Y_t^2 = \phi_{21} Y_{t-1} Y_t + \phi_{22} Y_{t-2} Y_t + \cancel{\mu Y_t} + a_t Y_t$

$Y_{t-1} Y_t = \phi_{21} Y_{t-1}^2 + \phi_{22} Y_{t-1} Y_{t-2} + \cancel{\mu Y_{t-1}} + a_t Y_{t-1}$

$Y_{t-2} Y_t = \phi_{21} Y_{t-1} Y_{t-2} + \phi_{22} Y_{t-2}^2 + \cancel{\mu Y_{t-2}} + a_t Y_{t-2}$

$Y_{t-3} Y_t = \phi_{21} Y_{t-1} Y_{t-3} + \phi_{22} Y_{t-2} Y_{t-3} + \cancel{\mu Y_{t-3}} + a_t Y_{t-3}$

Take E( )

$\gamma(0) = \phi_{21} \gamma(1) \overset{+?^2)}{} + \phi_{22} \gamma(2) \overset{+(?^2)}{} + \cancel{\dfrac{\mu^2}{1-\phi_1-\phi_2}} + \sigma^2$

$\gamma(1) = \phi_{21} \gamma(0) + \phi_{22} \gamma(1) + \cancel{\dfrac{\mu^2}{1-\phi_1-\phi_2}} + 0$

$\gamma(2) = \phi_{21} \gamma(1) + \phi_{22} \gamma(0) + (\quad) + 0$

$\gamma(3) = \phi_{21} \gamma(2) + \phi_{22} \gamma(1) + (\quad) + 0$

$\phi_{11} = \dfrac{\gamma(1)}{\gamma(0)}$

$\phi_{22} = \dfrac{\det\begin{pmatrix} 1 & \gamma(1) \\ \gamma(1) & \gamma(2) \end{pmatrix}}{\det\begin{pmatrix} 1 & \gamma(1) \\ \gamma(1) & 1 \end{pmatrix}}$

$\phi_{33} = \dfrac{\det\begin{pmatrix} 1 & \gamma(1) & \gamma(1) \\ \gamma(1) & 1 & \gamma(2) \\ \gamma(2) & \gamma(1) & \gamma(3) \end{pmatrix}}{\det\begin{pmatrix} 1 & \gamma(1) & \gamma(2) \\ \gamma(1) & 1 & \gamma(1) \\ \gamma(2) & \gamma(1) & 1 \end{pmatrix}} = 0.$

(4). Causal.

$X_t = \psi(B)a_t$ $\quad\quad$ $\Phi(B)X_t = \Theta(B)a_t$

~~$\phi(B)\psi(B)$~~ $\quad\quad$ $\psi(B)\Phi(B) = \Theta(B)$

$(1+\psi_1 B+\psi_2 B^2+\cdots)(1-\phi_1 B-\phi_2 B^2)=1$

solve for $\psi_1, \cdots, \psi_5$. skip.

Q7 Method of moment estimation.

(1). AR(2).

B/c ACF tails off as a damped sine wave.

PACF cuts off at lag 1 & 2. so AR(2) is a good fit

(AR(1) is possible?)

(2).

$\phi_1 = \rho_1 = \frac{\phi_1}{1-\phi_2}$ =>

$\phi_{22} = \frac{\rho_2-\rho_1^2}{1-\rho_1^2} = \phi_2 = \frac{0.64-0.78^2}{1-0.78^2} = \frac{0.0316}{0.3916} = 0.0807 \doteq 0.1$

$-0.78 = \frac{\phi_1}{1-0.0807}$

$\phi_1 = -0.7171 \doteq -0.7$

(3). stationary?

$1-\phi_1 B-\phi_2 B^2=0$

check $B = \frac{-\phi_1 \pm\sqrt{\phi_1^2+4\phi_2}}{2\phi_2} = \frac{0.7\pm\sqrt{0.49+0.4}}{0.2}$

=> soln outside ⊙1

(4). $Q_{LB}(10) = n(n+2)\sum_1^{10} \frac{\hat{\rho}_k^2}{n-k}$ = we don't know (n) observations.

n=100

Q8. Definitions

(1). Strictly stationary:

A stochastic process, the probabilistic behaviour of every collection of values $(X_{t+1}, X_{t+2}, \cdots, X_{t+n})$ is identical to that of time shift set $(X_{t+1+h}, \ldots, X_{t+n+h})$

Weakly stationary:

$E(X_t) = \mu$

$Var(X_t) = \gamma(0) = \sigma^2 < \infty$

does not depend on time t.

Covariance stationary

Strictly $\Rightarrow$ weakly

(2) General approach to t.s. modeling

1. Identify a preliminary t.s model
2. Estimation of model parameters
3. Check adequacy

(3) ARMA(p, q).

$\{X_t\}$ process. stationary

$$X_t - \phi_1 X_{t-1} - \cdots - \phi_p X_{t-p} = a_t + \theta_1 a_{t-1} + \cdots + \theta_q a_{t-q}, \quad a_t \sim WN(0, \sigma^2)$$

(4) Dual relationship AR $\Leftrightarrow$ MA

A finite order stationary AR(p) corresponds to a MA(∞)

& MA(q) $\Leftrightarrow$ AR(∞)

Causal/stationary AR(p) $\rightarrow$ MA(∞)

invertible MA(q) $\rightarrow$ AR(∞)

(5). Wold decomposition:

Any zero-mean process $\{X_t\}$ which is not deterministic can be expressed as a sum $X_t = U_t + V_t$

U_t is $MA(\infty)$ and V_t is deterministic, not related with U_t and $AR(p)$.

(6). Yule-Walker eq'tn for $AR(p)$

derive? skip

(7). PACF?

The correlation between X_t & X_{t+k} after mutual linear dependency on the intervening variables $X_{t+1}, \cdots, X_{t+k-1}$ removed.

$\phi_{kk} = corr(X_t, X_{t+k} \mid X_{t+1}, \cdots, X_{t+k-1})$

And PACF can form a matrix with ACF

(8). 2 methods model selection

AIC : $-2\log ML + 2k$ } smaller, better.
BIC : $-2\log ML + k\log(n)$ }

Q10. ~~causal/stationary~~ ~~invertible~~

C I.

~~(1) $X_t + 0.2X_{t-1} - 0.48X_{t-2} = a_t$~~

~~$\Phi(B)X_t = a_t$ is causal iff $|\phi| < 1$~~

~~$X_t = \Theta(B)a_t$ is invertible iff $|\theta| < 1$~~

Q10. Causal/stationary $\quad$ Invertible

C $\quad$ I.

(1). $X_t + 0.2X_{t-1} - 0.48X_{t-2} = a_t$

$X_t + 0.2BX_t - 0.48B^2X_t = a_t$

$(1 + 0.2B - 0.48B^2)X_t = a_t$ $\quad$ C & I

(2). $X_t + 1.9X_{t-1} + 0.88X_{t-2} = a_t + 0.2a_{t-1} + 0.7a_{t-2}$

$X_t(1 + 1.9B + 0.88B^2) = a_t(1 + 0.2B + 0.7B^2)$ $\quad$ C & I

(3). $X_t + 0.6X_{t-2} = a_t + 1.2a_{t-1}$

$X_t(1 + 0.6B^2) = a_t(1 + 1.2B)$ $\quad$ C but not I

✓ $\quad$ X

(4). $X_t + 1.8X_{t-1} + 0.81X_{t-2} = a_t$ $\quad$ C & I.

$X_t(1 + 1.8B + 0.81B^2) = a_t$ $\quad$ ~~NOT C but I~~

solution to $0.81z^2 + 1.8z + 1$

$= (0.9z + 1)^2 = 0$

$0.9z = 1$

$z = \frac{1}{0.9} > 1$

(5). $X_t + 1.6X_{t-1} = a_t - 0.4a_{t-1} + 0.04a_{t-2}$

$X_t(1 + 1.6B) = a_t(1 - 0.4B + 0.04B^2)$ $\quad$ NOT C but I.

X $\quad$ ✓ $\quad$ ✓

Selected practice questions from past ~~tests~~ exams

1). Method to remove seasonality of t.s.

Say the ~~season~~ seasonal period is d.

Then differencing at lag d.

i.e. $\nabla_d = X_t - X_{t-d}$

$\boxed{\text{2). DF unit root test}}$

3). Strictly/Weakly stationary

- ~~Df~~ joint prob distribution of $(X_t, \ldots, X_{t+n})$ is the same as $(X_{t+h}, \ldots, X_{t+n+h})$. $\forall t, \cancel{t}\, n, h \in \mathbb{Z}$
- ① $E(X_t) = m < \infty$

② $Var(X_t) = \sigma^2 < \infty$

③ $\gamma(t, t+h)$ is independent of time & a function of the distance between 2 time points. $\forall t, h \in \mathbb{Z}$.

4). Wold decomp/Support for ARMA

a). A stochastic process $\{V_t\}$ is called deterministic if the values of V_{t+j}, $j \geq 1$ are ~~per~~ predictable in terms of the span of its past observations

A deterministic process => $AR(\infty)$ process.

b). Wold decomp: zero-mean nondeterministic = deterministic + $MA(\infty)$

c). ~~Wold : ~~ same as ARMA

5). Inverstible ARMA(p, q)

(a). $\Phi(B)X_t = \theta(B)a_t$, $a_t \sim WN(0, \sigma^2)$

where $\Phi(B) = 1 - \phi_1 B - \cdots \phi_p B^p$

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

invertible if $\theta(B) = 0$, the solutions are outside of unit circle.

(b). Studying only the invertible process, we ensure that we are ~~un~~able to match any set of ~~AC~~ autocorrelation functions to a unique ARMA model.

6). PACF. / How to use PACFs for model identification.

a). ① The correlation between X_t & X_{t+h} after mutual linear dependency on the intervening variables $X_{t+1}, \cdots X_{t+h-1}$ are removed.

~~②~~ ie. $\phi_{kk} = corr(X_t, X_{t+k} \mid X_{t+1}, \cdots, X_{t+k-1})$

② (the correlation between 2 prediction errors)

$\phi_{11} = Corr(X_{t+1}, X_t) = \rho(1)$

$\phi_{hh} = Corr(X_{t+h} - \hat{X}_{t+h}, X_t - \hat{X}_t)$ $h \geq 2$

where

$\hat{X}_{t+h} = \beta_1 X_{t+h-1} + \beta_2 X_{t+h-2} + \cdots \beta_{h-1} X_{t+1}$

$\hat{X}_t = \beta_1 X_{t+1} + \cdots + \beta_{h-1} X_{t+h-1}$

β's are obtained by Minimized MSE forecast.

b). Since the theoretical value of PACF at lag h of AR(p) is zero if h>p.
we guess that AR(p) ~~be~~ cuts off after lag h ~~when~~ in its PACF function graph.

BOX-JENKINS APPROACH

1. AR(2)

(1). $(1-\frac{1}{2}B)(1-\frac{1}{10}B)X_t = a_t, \quad a_t \sim NID(0,1)$

2, 10, both outside unit circle.
So stationary.

(2). $\rho(1)$ & $\rho(2)$

$X_t - 0.6X_{t-1} + 0.05X_{t-2} = a_t$

$\phi_1 = 0.6$
$\phi_2 = -0.05$

$\boxed{\rho(k) = \phi_1 \rho(k-1) + \phi_2 \rho(k-2)}, \ k \geq 1$

$\rho(1) = \phi_1 + \phi_2 \rho(-1)$

$(1-\phi_2)\rho(1) = \phi_1$

$\rho(1) = \frac{\phi_1}{1-\phi_2} = \frac{0.6}{1.05} = 0.5714286$

$\rho(2) = \phi_1 \rho(1) + \phi_2$

$= 0.6 \times 0.5714286 + (-0.05) = 0.2928571$

(3). $\phi_{11} = \rho(1) = \frac{\phi_1}{1-\phi_2} = 0.5714286$

$\phi_{22} = \frac{\det\begin{vmatrix} 1 & \rho(1) \\ \rho(1) & \rho(2) \end{vmatrix}}{\det\begin{vmatrix} 1 & \rho(1) \\ \rho(1) & 1 \end{vmatrix}} = \phi_2 = -0.05$

$\phi_{33} = 0$

2. Method of ~~moments~~ moment estimation

(1). $\phi_1\gamma(1)+\phi_2\gamma(2)+\sigma^2=\gamma(0)$

$\phi_1\gamma(0)+\phi_2\gamma(1) \quad =\gamma(1)$

$\phi_1\gamma(1)+\phi_2\gamma(0) \quad =\gamma(2)$

$1200\,\cancel{1800}\,\phi_1+600\phi_2+\sigma^2=1800$

$1800\phi_1+1200\phi_2 \quad =1200$

$1200\phi_1+1800\phi_2 \quad =600$

$2\phi_1+\phi_2+\frac{\sigma^2}{600}=3$

$3\phi_1+2\phi_2 \quad =2$

$2\phi_1+3\phi_2=1$

$9\phi_1+6\phi_2=6$

$4\phi_1+6\phi_2=2$

$5\phi_1=4$

$\phi_1=\frac{4}{5}$

$\phi_2=\frac{1-2\times\frac{4}{5}}{3}=-\frac{1}{5}$

$\frac{\sigma^2}{600}=3-2\times\frac{4}{5}+\frac{1}{5}=\frac{15-8+1}{5}=\frac{8}{5}$

$\sigma^2=\frac{8}{5}\times600=960$

So $\phi_1=\frac{4}{5}$, $\phi_2=-\frac{1}{5}$, $\sigma^2=960$

(2). 95% CI

$\hat{\Phi}-\Phi\sim N(0,\boxed{\frac{\sigma^2\Gamma_p^{-1}}{n}})$

$n=0$. $\Gamma_p=\begin{bmatrix}\gamma(0) & \gamma(1)\\ \gamma(1) & \gamma(0)\end{bmatrix}$

$0.8\pm1.96\sqrt{?}$

$-0.2\pm1.96\sqrt{?}$

FORECAST

1. ARIMA(1,1,0).

$(1-0.5B)(1-B)X_t = a_t, \quad a_t \sim NID(0,1)$

(1). Forecast function for origin t

(2). var of 1-step ahead forecast error.

$$(1-0.5B)(1-B)X_t = a_t$$

$$(1-1.5B+B^2)X_t = a_t$$

$$X_t - \cancel{\ } BX_t + B^2X_t = a_t$$

$$BX_t = X_{t-1}$$

$$B^2X_t = X_{t-2}$$

$$X_t - 1.5X_{t-1} + X_{t-2} = a_t$$

AR(2)

$$X_{t+1} = 1.5X_t - X_{t-1} + a_{t+1}$$

$$\hat{X}_t(1) = E(X_{t+1} \mid data) = 1.5X_t - X_{t-1}$$

$$X_t = 1.5X_{t-1} - X_{t-2} + a_t$$

2. AR (1).

$(1-0.6B)(X_t-9)=a_t$, $a_t \sim NID(0,1)$

Observe $(X_{97}, X_{98}, X_{99}, X_{100}) = (9.6, 9, 9, 8.9)$

(1). This is a AR(1) process with mean $\mu=9$.

stationary?

$(1-0.6B)=0$

$B=\frac{5}{3}>1$. so stationary

(2). Forecast $\{X_t\}$, $t=101, 102, 103, 104$.

and their associated 95% ~~ ~~ forecast limits

causal

$\hat{X}_{101} = 9+0.6(X_{100}-9)$

$= 9+0.6 \times \cancel{\ } -0.1$

$= 8.94$

$\hat{X}_{102} = 9+0.6(\hat{X}_{101}-9)$

$= 9+0.6\times \cancel{0.06} (-0.0\ 6) = \cdots$

- - -

Forecast limits

use integrated form

$\hat{X}_t(L) = \sum_0^\infty \psi_i | a_{t+L-i} |$

$\psi(B) = \theta(B)/\phi(B)$

$\theta(B)=1$, $\phi(B)=1-0.6B$

$\psi(B) = \cancel{\ } \sum_0^\infty 0.6^j B^j$

$\psi_j = 0.6^j$

~~sd~~ standard deviation of $\hat{X}_t(l)$ to be $\sqrt{1+\sum_{j=1}^{l-1} 0.6^{2j}}$, $\sigma=1$

sd of $\hat{X}_{101}$ is $=1$

sd of $\hat{X}_{102}$ is $\sqrt{1+0.6^2} = 1.16619$

So ~~forec~~ 95% forecast limits for $\hat{X}_{101}$ is $8.94 \pm 1.96 \cdot 1$

$\hat{X}_{102}$ is $\cdots \pm 1.96 \times 1.16619$

$\hat{X}_{103}$ - - -

$\hat{X}_{104}$ - - -

(3). Updating forecast formula:

$$\hat{X}_{t+1}(l) = \hat{X}_t(l+1) + \psi_l a_{t+1}$$
$$a_t = X_t - \hat{X}_{t-1}(1)$$

$a_{101} = 8.8 - 8.94 = -0.14$

So $\hat{X}_{101}(1) = 8.94 + 0.6 \times (-0.14) = 8.856$.

...

3. Forecast ARMA(1,1)

$X_t - 0.5X_{t-1} = a_t + 0.25a_{t-1}$, $a_t \sim NID(0, 1)$.

$(X_{97}, X_{98}, X_{99}, X_{100}) = (-0.7, -1, -0.8, -0.4)$

(a). Forecast function:

Lead time $= 1$:

$$\hat{X}_t(1) = 0.5X_t + 0.25(X_t - \hat{X}_{t-1}(1))$$

Lead time h, $h>1$ $\hat{X}_t(h) = 0.5\hat{X}_t(h-1)$

(b). Best linear forecast $X_{101} + X_{102} + X_{103}$

$\hat{X}_{100}(1) = 0.5 \times (-0.4) + 0.25(-0.4 - 0) = -0.3$

$\hat{X}_{102}(2) = 0.5 \times (-0.3) = -0.15$

$\hat{X}_{103}(3) = 0.5 \times (-0.15) = -0.075$

Best linear forecast is $-0.3 - 0.15 - 0.075 = -0.525$

(c). 95%

5. $(1-0.5B)(X_t-4)=(1+0.5B)a_t \quad a_t \sim NID(0,1)$. ARMA(1,1)

1-step forecast at origin $t=99$ is $\hat{X}_{99}(1)=2.09$

and $\{X_{99}, X_{100}, \cdots, X_{105}\}$

$= 2.11, 1.39, 2.57, 4.11, 6.28, 4.89, 5.94$

(a). $\hat{X}_{100}(l)$

for $l=1,2,3$

$\cancel{\hat{X}_{100}(1)}$

$X_t - 4 - 0.5X_{t-1} + 2 = a_t + \frac{1}{2}a_{t-1}$

$X_{t+l} = 0.5(X_{t+l-1}-4)+4+a_{t+l}+0.5a_{t+l-1}$

make conditional expectation filtration (4 rules) $\Rightarrow$ $\hat{X}_t(1) = 4+0.5(X_t-4)+0.5\hat{a}_t$

and $\hat{a}_t = X_t - \hat{X}_{t-1}(1)$

$\hat{a}_{100} = X_{100} - \hat{X}_{99}(1) = 1.39-2.09 = -0.7$

so $\hat{X}_{100}(1) = 4+0.5\times(1.39-4)-0.5\times 0.7 = 2.345$

$\hat{X}_{100}(2) = 4+0.5\times(2.345-4) = 3.1725$

$\hat{X}_{100}(3) = 4+0.5\times(3.1725-4) = 3.58625$

~~FORECAST~~

Forecast error Variance

$$(1-0.5B)(1+\psi_1 B+\dots) = 1-\theta B$$

$$(1-0.5B)(1+\cdot \dots) \quad 1+0.5B$$

$$-0.5B+\psi_1 B = 0.5B$$

$$\psi_1 = 1$$

$$\psi_2 = 0.5$$

$$\psi_3 = 0.25$$

$$\dots$$

So $Var(e_{100}(1)) = 1$

$Var(e_{100}(2)) = 1+(0.5^0)^2 = 2$

$Var(e_{100}(3)) = 1+(0.5^0)^2+(0.5^1)^2 = 2.25$

$$\boxed{Var = (1+\psi_1^2+\psi_2^2+\dots)\sigma^2}$$

$$\cancel{\hat{X}_{t+l} = \hat{X}_t(l)}$$

4.

$(1-B)(1+0.9B)X_t = a_t$

$\Leftrightarrow \left(1-B+0.9B-0.9B^2\right)X_t = a_t$

$\Leftrightarrow$

$X_t - 0.1X_{t-1} - 0.9X_{t-2} = a_t$

$X_t = 0.1X_{t-1} + 0.9X_{t-2} + a_t$

$\hat{X}_t(l) = E(X_t \mid \text{data given}) = E\left(0.1X_{t-1+l} + 0.9X_{t-2+l} + a_{t+l}\right)$

when $l=1$

$\hat{X}_{t+1} = E(0.1X_t + 0.9X_{t-1} + a_{t+1})$

$= 0.1X_t + 0.9X_{t-1} + 0$

$l=2$, $\hat{X}_{t+2} = E(0.1X_{t+1} + 0.9X_t + a_{t+2})$

$= 0.1\hat{X}_{t+1} + 0.9X_t + 0$

$l=3$, $\hat{X}_{t+3} = E(0.1X_{t+2} + 0.9X_{t+1} + a_{t+3})$

$= 0.1\hat{X}_{t+2} + 0.9\hat{X}_{t+1} + 0$

Lecture Four
Vector Autoregressive & Causality

VAR(1)
$r_t = \Phi_0 + \Phi_1 r_{t-1} + a_t$

Φ_0 is k-dim vector, Φ_1 is a $k \times k$ matrix, a_t is a multivariate white noise series, IID

e.g.
$r_t = \begin{bmatrix} r_{1,t} \\ r_{2,t} \end{bmatrix}$, $a_t = \begin{bmatrix} a_{1,t} \\ a_{2,t} \end{bmatrix}$

$r_{1,t} = \phi_{10} + \phi_{11} r_{1,t-1} + \phi_{12} r_{2,t-1} + a_{1,t}$
$r_{2,t} = \phi_{20} + \phi_{21} r_{1,t-1} + \phi_{22} r_{2,t-1} + a_{2,t}$

In matrix form:

$$\begin{bmatrix} r_{1,t} \\ r_{2,t} \end{bmatrix} = \begin{bmatrix} \phi_{10} \\ \phi_{20} \end{bmatrix} + \begin{bmatrix} \phi_{11} & \phi_{12} \\ \phi_{21} & \phi_{22} \end{bmatrix} \begin{bmatrix} r_{1,t-1} \\ r_{2,t-1} \end{bmatrix} + \begin{bmatrix} a_{1,t} \\ a_{2,t} \end{bmatrix}$$

Stationary VAR(1) Model
Assume $\Phi_0 = 0$, iteration:

$$r_t = a_t + \Phi_1 a_{t-1} + \Phi_1^2 a_{t-2} + \cdots + \Phi_1^j a_{t-j} + \cdots$$
$$= \sum_{j=0}^{\infty} \Phi_1^j a_{t-j}$$

- The impact of past shocks, say a_{t-j}, on r_t is given by Φ_1^j.

For r_t to be stationary:

All eigenvalues of Φ_1 are less than 1 in modulus

$\iff r_t$ is weakly stationary, provided the covariance matrix of a_t exists.

· Cross-covariance matrix of VAR(1):

$$\text{cov}(r_t, r_{t-l}) = \Gamma_l = \Phi_1 \Gamma_{l-1}, \quad \forall l > 0$$

$$\text{cov}(r_t, r_t) = \Gamma_0 = \sum_{j=0}^{\infty} \Phi_1^j \Sigma (\Phi_1^j)^T$$

where $\Phi_1^0 = I_k$

and Γ_l is the lag-j cross covariance matrix of r_t.

~~GENERALIZATION~~: VAR(p) model

Lecture 5

4 variable $\begin{bmatrix} P_t \\ e_t \\ rw_t \\ U_t \end{bmatrix} = \begin{bmatrix} \\ \\ \\ \\ \end{bmatrix} + \begin{bmatrix} \\ \\ \\ \\ \end{bmatrix} + \begin{bmatrix} \phi_{11} & \phi_{12} & \phi_{13} & \phi_{14} \\ \phi_{21} & \phi_{22} & \phi_{23} & \phi_{24} \\ \phi_{31} & \phi_{32} & \phi_{33} & \phi_{34} \\ \phi_{41} & \phi_{42} & \phi_{43} & \phi_{44} \end{bmatrix} \begin{bmatrix} P_{t-1} \\ e_t \\ rw_{t-1} \\ U_{t-1} \end{bmatrix} + \begin{bmatrix} a_{1t} \\ a_{2t} \\ a_{3t} \\ a_{4t} \end{bmatrix}$

July 30th.

Forecast (Accuracy) Evaluation

1. Minimum MSPE (Mean Square Prediction ERROR)
   - Compare different models.
   - smaller MSPE, better.

2. F-statistic

$\boxed{\text{3 assumptions}}$

a). Forecast errors have zero mean, normally distributed.

b). Forecast errors serially uncorrelated.

c). ——— ~~contemporary corr~~ contemporaneously uncorrelated with each other.

- The third is the worst. (problematic)

3. So we have another statistics: called Granger-Newbold Test about how to over-come c) in 2.

Consider transformation:

$x_i = e_{1i} + e_{2i}$

$y_i = e_{1i} - e_{2i}$

H_0: $var(e_{1i}) = var(e_{2i})$ ⟵

$Cov(x_i, y_i) = E(x_i \cdot y_i) = E(e_{1i}^2) - E(e_{2i}^2) = 0$

$\Updownarrow$

$x_i \cdot y_i = (e_{1i} + e_{2i})(e_{1i} - e_{2i}) = e_{1i}^2 - e_{2i}^2$

r_{xz} : sample correlation coefficient between x_i & y_i

$$r_{xz} = \frac{\sum(x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\sum(x_i-\bar{x})^2}\sqrt{\sum(y_i-\bar{y})^2}}$$

$$t = \frac{Cov(x_i, y_i)}{\underbrace{\sqrt{Var(x_i)}\sqrt{Var(y_i)}}_{\Delta}}$$

So depends on the difference of 2 model's error square.

rewrite:

If: $t = \frac{Cov(x_i, y_i)}{\Delta} = \frac{E(e_{1i}^2) - E(e_{2i}^2)}{\Delta} > 0$

Δ → positive

$\Rightarrow E(e_{1i}^2) - E(e_{2i}^2) > 0$

$\Rightarrow E(e_{1i}^2) > E(e_{2i}^2)$

$\Rightarrow$ model 2 is better.

If: $t = \text{- - - -} < 0$

$\Rightarrow E(e_{1i}^2) < E(e_{2i}^2)$

$\Rightarrow$ model 1 is better

4. (Another Test)
The Diebold-Mariano Test (skipped?)

- 3 & 4 not work for nested model like $AR(1) \subseteq AR(2) \subseteq \ldots$

5. Clark & West Test
(For Nested model)
(skipped)

Introduction to Transfer Function Noise Model

~~VAR(1):~~

$$\cancel{\begin{bmatrix} Y_{1t} \\ Y_{2t} \end{bmatrix} = \begin{bmatrix} \phi_{11} & \phi_{12} \\ \phi_{21} & \phi_{22} \end{bmatrix} \begin{bmatrix} Y_{1t-1} \\ Y_{2t-1} \end{bmatrix} + \begin{bmatrix} a_{1t} \\ a_{2t} \end{bmatrix}}$$

$Y_t = \cancel{X_t + X_{t-1} +}$

$v_0 X_t + v_1 X_{t-1} + \cdots = v(B) X_t$ (*)

$v_0, v_1, \cdots$ are called impulse response function

$v(B) = v_0 + v_1 B + v_2 B^2 + \cdots$

For (*) meaningful impulse responses are absolutely summable.

$\sum_{j=0}^{\infty} |v_j| < \infty$

$Y_t = \sum_{j=0}^{\infty} v_j X_{t-j} = \sum_{j=0}^{\infty} v_j$?

$Y_{t+1} = \sum_{j=0}^{\infty} v_j$

we say the system is stable

$g = \sum_{j=0}^{\infty} v_j$ is called the steady-state gain as it represents the impact on Y when X_{t-j} are held constant over time.

in infinite order polynomial

Now we can approximate $v(B)$ with 2 finite order polynomials:

$$\omega(B) = \omega_0 + \omega_1 B + \cdots + \omega_s B^s$$

$$\delta(B) = 1 - \delta_1 B - \cdots - \delta_r B^r$$

$\omega_0 \neq 0$.

$$v(B) = \frac{\omega(B) B^b}{\delta(B)}, \quad b \geq 0 \Rightarrow \text{time delay of the system.}$$

N_t noise term

$$Y_t = \sum_{j=0}^{\infty} v_j X_{t-j} + N_t = \frac{\omega(B) B^b}{\delta(B)} \cdot X_t + N_t$$

$$\phi(B) N_t = \theta(B) a_t$$

Model Building Process

(1). ~~The~~ Preliminary identification of the impulse response coefficients v_i's.

$$\boxed{\begin{array}{l} (1)\ \{v_j\} \overset{(2)}{\Rightarrow} Y_t - \sum \hat{v}_j X_{t-j} = \hat{N}_t \\ \searrow (2).\ \text{can guess } \frac{\omega(B) B^b}{\delta(B)} \end{array}}$$

(2). Specification of N_t.

(3). Specification of transfer function

(4). Estimation of TFN model specified in Step 2 & 3

(5). Model diagnostic checks

In specific:

(1) $Y_t = \cancel{\sum_{j=0}^{\infty} v_j X_{t-j}} + N_t = \frac{\omega(B) B^b}{\delta(B)} X_t + N_t$

$v(B) X_t$

$\phi_x(B) X_t = \theta_x(B) \alpha_t$

$\Downarrow$

Stationary / Causal

$\Rightarrow \frac{\phi_x(B)}{\theta_x(B)} = \alpha_t$

Apply this on both sides

$$\underbrace{\frac{\phi_x(B)}{\theta_x(B)} Y_t}_{\beta_t} = v(B) \underbrace{\frac{\phi_x(B)}{\theta_x(B)} X_t}_{\alpha_t} + \underbrace{\frac{\phi_x(B)}{\theta_x(B)} N_t}_{\text{still a noise term } \cancel{n_t}}$$

Then multiply both sides by α_{t-j}, $j \geq 0$.

$\beta_t \alpha_{t-j} = v(B) \alpha_t \alpha_{t-j} + n_t \alpha_{t-j}$

$E(\beta_t \alpha_{t-j}) = E(v(B) \alpha_t \alpha_{t-j}) + E(n_t \alpha_{t-j})$

* noise term actually

$\text{cov}(\beta_t, \alpha_{t-j}) = \underbrace{v_j \cdot \text{var}(\alpha_{t-j})}_{} + 0$

Why?

$$\boxed{\begin{aligned} &(v_0 + v_1 B + \cdots) \alpha_t \alpha_{t-j} \\ &= v_0 \underbrace{\alpha_t \alpha_{t-j}}_{\substack{\gamma(j) \\ \| \\ 0}} + v_1 \underbrace{\alpha_{t-1} \alpha_{t-j}}_{\substack{\gamma(j-1) \\ \| \\ 0}} + \cdots + v_j \underbrace{\alpha_{t-j} \alpha_{t-j}}_{\substack{\gamma(0) \\ \| \\ \text{var}(\alpha_{t-j})}} \end{aligned}}$$

~~So var(α_{t-j})~~ $v_j = \cancel{\frac{\text{cov}(\beta_t, \alpha_{t-j})}{\text{var}(\alpha_{t-j})}} =$

$\underline{cov(\beta, \alpha)} = v \, Var(\alpha)$

$r_{\beta\alpha} \cdot se_{\beta} \cdot se_{\alpha} = v \, (se_{\alpha})^2$

$\Rightarrow r_{\beta\alpha} = v \cdot \frac{se_{\alpha}}{se_{\beta}}$

$\Rightarrow v = r_{\beta\alpha} \cdot \frac{se_{\beta}}{se_{\alpha}}$

(2). Specification of N_t

(3). Specification of transfer function

- ~~Goal~~: find $\omega(B)$ & $\delta(B)$, & b. to best approximate $v(B)$.

$$v(B) = \frac{\omega(B) B^b}{\delta(B)}$$

$$\Rightarrow v_0 + v_1 B + \cdots = \frac{\omega_0 B^b + \omega_1 B^{b+1} + \cdots + \omega_s B^{b+s}}{1 - \delta_1 B - \cdots - \delta_r B^r}$$

Then ...

(4). Estimation

(5). Diagnostic checking.

idea:

Tutorial 4. on July 31st

Conditional expectation ~~w~~ given filtration evaluation

(in forecast pdf practice a)

Sps. A, B are events

prob space $(\Omega, \mathcal{F}, P_r)$

① if $A \in \mathcal{F}$, $A \subset \Omega$
then $\bar{A} \in \mathcal{F}$

② $\Omega \in \mathcal{F} \Rightarrow \emptyset \in \mathcal{F}$

③ $A_1, A_2, \dots \in \mathcal{F}$
then $\bigcup_{i=1} A_i \in \mathcal{F}$

①③ $\Rightarrow$ if $A_1, \dots, \in \mathcal{F}$
then $\bigcap_{i=1}^{\infty} A_i \in \mathcal{F}$

$P_r(\emptyset) = 0$

$P_r(\Omega) = 1$

if $A_1, \dots$ disjoint, then $P_r(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P_r(A_i)$

$P_r(A) \in [0,1]$, $A \in \mathcal{F}$

$P_r(A|B) = \frac{P(A \cap B)}{P(B)}$ given $P(B) \neq 0$

X is r.v.

$X : \Omega \to \mathbb{R}$

$F(x) = P_r(X \le x)$
$= P_r(\omega \in \Omega : X(\omega) \le x)$ "hidden story here"

$E(X) = \int x f(x) dx = \int x \, dF(x) \Leftrightarrow \int_{\Omega} X(\omega) \, d P_r(\omega)$

$E(Y|X) \overset{\text{classic}}{=} \int y f(y|x) dy = g(x)$ r.v.

$E(Y|X) \overset{\text{def}}{=} g(X)$ s.t. $\int_A g(x) dP_r = \int_A Y dP_r$

$A \in G \subseteq F \leftarrow (\Omega, F, P_r)$

$E(Y|G) = g(Y)$ s.t. $\int_A g(Y)\,dP = \int_A Y\,dP$

$E(Y|X)$ is a special case of $E(Y|G)$

Let $G = \sigma(X)$

$\sigma(X)$ ~~is a~~ is defined as $\{A \subseteq F : A = \cdots\}$

~~idea~~

$\{4\}, \{5\}$

e.g. $\Omega = \{1, 2, 3, 4, 5, 6\}$

$F = \{\{1\}, \{2\}, \{3\}, \{1,2\}, \{2,3\}, \{1,3\}, \{1,2,3\}, \emptyset\}$

$X(\omega) = \begin{cases} 1 & \text{if } \omega \text{ is odd} \\ 0 & \text{if } \omega \text{ is even} \end{cases}$

$\sigma(X) = \{\{1, 3, 5\}, \{2, 4, 6\}, \emptyset, \Omega\}$

like $\{3, 4\} \rightarrow$ not observable through $X(\omega)$

if $X(\omega) = 1$

$\sigma(X) = \{\emptyset, \Omega\} \leftarrow$ non informative

idea

Filtration:

$G_0 \subset G_1 \subset \cdots$

$G_0 = \{\emptyset, \Omega\}$

information grows over time

$G_t = \sigma(X_1, X_2, \ldots, X_t)$

information up to t

$G_0 = \{\emptyset, \Omega\}$

$G_1 = \sigma(X_1)$ $\quad G_2 = \sigma(X_1, X_2)$

more info than G_1 unless X_1 and X_2 are the same. But we X_1 & X_2 are correlated we get

less info than when they are independent

if $X_1 \perp X_2$, then G_2 is optimal.
if $X_1 \equiv X_2$, no new info. in G_2 comparing with G_1.

$$E(\hat{X}_1(1) \mid X_1) \equiv E(\hat{X}_1(1) \mid \sigma(X_1))$$

① $\sigma(X_1) = \{\emptyset, \Omega\}$

$E(\hat{X}_1(1) \mid \sigma(X_1)) = E(\hat{X}_1(1))$

② $\sigma(X_1) = \mathcal{F}$

$E(\hat{X}_1(1) \mid \sigma(X_1)) = \hat{X}_1(1)$ a.s.

Thm: $\hat{X}_1(1) = g(X_1)$

① $\cancel{\text{Var}}$ Res $(\hat{X}_1(1)) = E(\hat{X}_1(1) - E(\hat{X}_1(1)))^2 \neq 0 > 0$

② $\cancel{\text{Var}}$ sq. residual

$$= E(\hat{X}_1(1) - E(\hat{X}_1(1) \mid X_1))^2 \neq 0$$

THIS LECTURE HAS NO SLIDES

STA457 Lecture August 6th

Final exam : 60 pts
- Before exam: 12 pts
- After exam : 48 pts
- Definition: 24 pts
- Calculation 36 pts

different breakdown.

calculator.

Topics:

1. Midterm ..
2. Vector autoregressive model
3. Granger Causality
4. Cointegration
5. Forecast evaluation (MPSE)...
6. Transfer function
7. Bayesian dynamic linear model (*)
8. Spectral analysis (Frequency domain analysis) (*)

not covered yet.

- very useful to do the practice problems.

$$y_t = \sum_{i=0}^{\infty} v_j x_{t-j} + e_t$$

notation doesn't matter

$$\gamma_{XY}(k) = \gamma_k(X, Y) = cov(X_{t+k}, Y_t) = cov(X_t, Y_{t-k})$$

$$\rho_k(X, Y) = \frac{\gamma_k(X, Y)}{\sqrt{\gamma_X(0)\gamma_Y(0)}}$$

$$\gamma_X(0) = Cov(X_t, X_t)$$

Sps the real data:

$$Y_t = \beta_0 + \beta_1 X_{t-d} + e_t$$

$$\rho_k(X,Y) = \frac{Cov(\cancel{Y_t, X_t}) X_{t+k}, Y_t)}{\sqrt{Var(Y_t)\, Var(X_t)}}$$

$$= \frac{Cov(X_{t+k}, \beta_1 X_{t-d} + e_t)}{\sigma_x \sqrt{\beta_1 \sigma_x^2 + \sigma_e^2}} \quad \text{(constants ignored)}$$

$$= \boxed{\frac{\beta_1\, Cov(X_{t+k}, X_{t-d})}{\sigma_x \sqrt{\beta_1 \sigma_x^2 + \sigma_e^2}}} + \cancel{0}$$

· unless $k = -d \Rightarrow \rho_k(X,Y) = 0$

$$X_t = \phi_x X_{t-1} + a_t$$
$$Y_t = \phi_Y Y_{t-1} + \zeta_t$$

if both AR(1) & independent

$$\phi_x(B) X_t = \theta_x(B) \alpha_t$$
$$y_t = v(B) X_t + n_t$$
$$\boxed{\frac{\phi_x(B)}{\theta_x(B)} x_t = \alpha_t}$$

β_t

$$\frac{\phi_x(B)}{\theta_x(B)} y_t$$

α_t

$$= v(B) \frac{\phi_x(B)}{\theta_x(B)} x_t + \frac{\phi_x(B)}{\theta_x(B)} n_t$$

~~$\sqrt{n}\,\gamma_k(X_t, Y_t)$~~

$$\sqrt{n}\,\gamma_k(X,Y) \to N\left(0, \frac{1+\phi_x\phi_Y}{1-\phi_x\phi_Y}\right)$$

if $X_t, Y_t \sim WN$, $\phi_x = \phi_Y = 0$

then $\sqrt{n}\,\gamma_k(X,Y) \to N(0, 1)$

But if $\phi_x \phi_Y = 0.99$

$\Rightarrow \sqrt{n}\,\gamma_k(X,Y) \to N(0, 100)$

$$\sqrt{n}\,\hat{\gamma}_k(X,Y) \to N(0,\ 1+2\underbrace{\textstyle\sum_{j=1}^{\infty} \rho_j(X)\rho_j(Y)}_{0})$$

$$\underbrace{\frac{\phi_x(B)}{\theta_x(B)} y_t}_{\beta_t} = v(B) \underbrace{\frac{\phi_x(B)}{\theta_x(B)} X_t}_{\alpha_t} + \frac{\phi_x(B)}{\theta_x(B)} n_t$$

$$\Rightarrow \phi_x(B) X_t = \alpha_t \theta_x(B)$$

$$\gamma_{\alpha_t}(k) = 0 \ \forall\ k \neq 0$$

$\boxed{\gamma_{\beta_t}(k) = ?\ \text{whatever we don't care}}$

Example:

Practice problem.

write ARMAX model in the form of a TFN model.

$$C_t = 0.68 C_{t-1} - 0.71 C_{t-2} + 0.69 C_{t-3} + 0.26 C_{t-4} - 0.07 X_{t-1} - 0.05 X_{t-2} + a_t - 0.08 a_{t-1} + a_{t-2}$$

$$\phi(B) C_t = \tilde{v}(B) X_t + \theta(B) a_t$$

$$C_t = \underbrace{\frac{\tilde{v}(B)}{\phi(B)}}_{v(B)} X_t + \overbrace{\frac{\theta(B)}{\phi(B)} a_t}^{n_t}$$

Next topic:

Dynamic Linear Model

$$y_t = F_t' \theta_t + v_t \quad \text{[observation equation]}$$

$$\theta_t = G_t \theta_{t-1} + w_t \quad \text{[system equation]}$$

$$\theta_0 \sim p(\theta_0 | D_0) \quad \text{[initial information]}$$

θ_t : $p \times 1$ state vector
F_t : $p \times 1$ vector of constants (known)
G_t : $p \times p$ evolution matrix
v_t : observation noise (scalar)
w_t : $p \times 1$ evolution noise (vector)

$v_t \sim N(0, V_t)$
$w_t \sim N(0, W_t)$
with v_s & w_t indpt $\forall t, s$ & indpt of $(\theta_0 | D_0)$

Shorthand Notation.

$\{F_t, G_t, V_t, W_t\}$

$$D_t = \{y_{1:t}, D_0\} = \{y_t, D_{t-1}\}$$

$$f_t(h) = E(y_{t+h} | D_t) = F_{t+h}' G_{t+h} \cdots G_{t+1} \underline{E(\theta_t | D_t)}$$

$*$ if $G_t = G \Rightarrow f_t(h) = F_{t+h}' G^h E(\theta_t | D_t)$

E.g. AR(p) model

$$y_t = \sum_{j=1}^{p} \phi_j y_{t-j} + v_t$$

$\{F_t, I_p, v_t, 0\}$ $\quad I_p = (1, \cdots, 1)$

$F_t' = (y_{t-1}, \cdots, y_{t-p})$

$\theta_t = (\phi_1, \cdots, \phi_p)$

Tutorial August 7th

$X_t \to Y_t \to Z_t$

$Z|Y, X$ the same as $Z|Y$

$Z|Y \perp X|Y$

X_t does not change $P_r(Z_t/Y_t)$

to predict Z_{t+1}

$Z_1, Z_2, \cdots, Z_t$

$X_1, X_2, \cdots X_t$

$E(Z_{t+1} | Z_1, \cdots, Z_t, X_1, \cdots, X_t)$

is better than

$E(Z_{t+1} | Z_1, \cdots, Z_t)$

General Fact:

if $X_1, X_2, \cdots, X_n \overset{iid}{\sim} N(0,1)$

$S = \sum_{i=1}^{n} X_i^2 \sim \chi_n^2$

H_0 : u and v has no causality

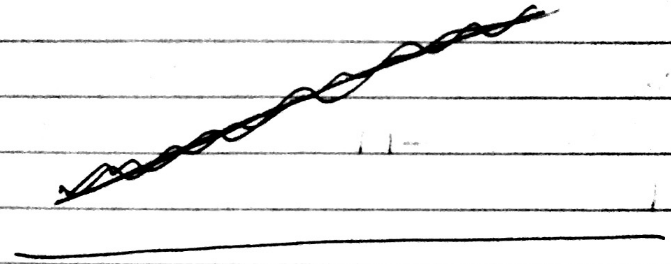
~~H_1~~: Cross-covariance

$r_{uv}(k) \approx N(0, n-k)$ approximately

$Q_L = n^2 \sum \frac{r_{uv}^2(k)}{n-k} \sim \chi^2_{L+1}$

$\left(\frac{r_{uv}(k)}{\sqrt{n-k}}\right) \approx N(0,1)$

eg



diff ∇ analogous to differentiation

$$y = at + b + \varepsilon$$

$$\frac{dy}{dt} = a$$

$d=1$

$$(1-B)^{1}\,\varphi(B)X_t = \theta(B)a_t$$

$$\pi(B) = (1-B)\phi(B)$$

$$\frac{\pi(B)}{1-B} =$$

- n). For eigenvalue of matrix

USE DET

$$\Phi = \begin{bmatrix} 1.1 & -0.6 \\ 0.3 & 0.2 \end{bmatrix}$$

$$\det(\lambda I - \Phi) = 0$$

$$\Phi v = \lambda v \quad v = (v_1, v_2)$$

$$\det\begin{bmatrix} 1.1-\lambda & -0.6 \\ 0.3 & 0.2-\lambda \end{bmatrix} = 0$$

$$\begin{cases} 1.1v_1 - 0.6v_2 = \lambda v_1 \\ 0.3v_1 + 0.2v_2 = \lambda v_2 \\ v_1^2 + v_2^2 = 1 \end{cases}$$

$$(1.1-\lambda)(0.2-\lambda) + 0.18 = 0$$

$$0.22 - 1.3\lambda + \lambda^2 + 0.18 = 0$$

✓

Lecture 7. Spectral Analysis

Recall: DLM

$$y_t = F_t'\theta_t + v_t$$

$$\theta_t = \boxed{G_t}\,\theta_{t-1} + w_t$$

$$F_t' = (x_{1t}, \dots, x_{pt}) \qquad \theta_t = \begin{pmatrix} \beta_{1t} \\ \beta_{2t} \\ \vdots \\ \beta_{pt} \end{pmatrix}$$

I G_t不断变

$$y_t = \sum_{i=1}^{p} \beta_{it} x_{it} + v_i$$

$$\psi(L) = 1 + \psi_1 L + \psi_2 L^2 + \cdots$$

Population spectrum / spectral density function

~~Let Y_t be a causal/stationary process and γ_j~~

$$Y_t = \sum_{i=0}^{\infty} \psi_i a_{t-i} = \psi(B) a_t, \quad a_t \sim WN(0, \sigma^2)$$

$$\psi(B) = 1 + \psi_1 B + \psi_2 B^2 + \cdots$$

$$\gamma_k = \sigma^2 \sum_{i=0}^{\infty} \psi_i \psi_{i+k}$$