Ch2. Stationary Processes

1. Linear Processes
2. ARMA processes
3. Properties of sample mean and SACF
4. Forecasting stationary time series

General Data Analysis Step

Consider the model given by

$$\boxed{X_t = m_t + s_t + Y_t}$$

(you can set $m_t = 0$, $s_t = 0$ if there is no trend/seasonality.)

1. Plot data: Time plot and Correlogram
2. Look for trend/seasonality etc.
3. If so, apply (polynomial) regression, smoothing or differencing to remove trend/seasonality. Then obtain "stationary errors"
$$\widehat{Y}_t = X_t - \hat{m}_t - \hat{s}_t$$
4. Apply Ljung-Box and variants, SACF plots of residuals, turning point test etc to test randomness of residuals.
5. If statioanry errors are WN/IID$(0, \sigma^2)$, you only need to estimate $\sigma^2 = \gamma(0)$. Otherwise, we need to find the variance covariance matrix (Γ) of Y_t from the residuals $\{\widehat{Y}_t\}$ to explain dependent error structure.
6. Do forecasting/prediction for future values.

More on Γ

For stationary sequence Y_t, the variance-covariance matrix is

$$\Gamma = \begin{pmatrix} \mathrm{Cov}(Y_1, Y_1) & \mathrm{Cov}(Y_1, Y_2) & \dots & \mathrm{Cov}(Y_1, Y_n) \\ \mathrm{Cov}(Y_2, Y_1) & \mathrm{Cov}(Y_2, Y_2) & \dots & \mathrm{Cov}(Y_2, Y_n) \\ \vdots & \vdots & \vdots & \vdots \\ \mathrm{Cov}(Y_n, Y_1) & \mathrm{Cov}(Y_n, Y_2) & \dots & \mathrm{Cov}(Y_n, Y_n) \end{pmatrix}$$

$$= \begin{pmatrix} \gamma(0) & \gamma(1) & \dots & \gamma(n-1) \\ \gamma(1) & \gamma(0) & \dots & \gamma(n-2) \\ \vdots & \vdots & \vdots & \vdots \\ \gamma(n-1) & \gamma(n-2) & \dots & \gamma(0) \end{pmatrix}$$

More on Γ

- However, in general, if $\{Y_t\}$ are correlated, then we need to estimate entire Γ.
- In Chapter 1, we learned that ACVF can be estimated by SACVF

$$\widehat{\gamma}(h) = \frac{1}{n}\sum_{j=1}^{n-h}(\widehat{Y}_j - \overline{\widehat{Y}})(\widehat{Y}_{j+h} - \overline{\widehat{Y}})$$

- But, consider the sample size you can use to estimate $\widehat{\gamma}(n-1)$. You have only one observation. It means that however large sample you may have, estimating $\gamma(n-1)$ is extremely unstable!
- We can do better by incorporating parametric modeling of Γ.

Ch2.2 Linear Processes

Linear Processes

Definition

$\{X_t\}$ is called a linear process if

$$X_t = \sum_{j=-\infty}^{\infty} \psi_j Z_{t-j}$$

for all t where $\{Z_t\} \sim WN(0, \sigma^2)$ and $\sum_j |\psi_j| < \infty$ (absolutely summable).

Examples:

- $X_t = \psi_0 Z_t + \psi_1 Z_{t-1} + \psi_2 Z_{t-2} + \ldots + \psi_q Z_{t-q}$
- $X_t = Z_t + .2 Z_{t-1}$

Linear process with backshift operator

- Backshift opertator B

$$BZ_t = Z_{t-1}, \quad B^2 Z_t = Z_{t-2}, B^{-1} Z_t = Z_{t+1}$$

- Now, rewrite the linear process

$$\sum_j \psi_j Z_{t-j} = \Big(\sum_j \psi_j B^j Z_t \Big) = \Big(\sum_j \psi_j B^j \Big) Z_t := \psi(B) Z_t,$$

where

$$\psi(B) = \sum_j \psi_j B^j$$

- Example:

$$X_t = Z_t + .2Z_{t-1} = (1 + .2B)Z_t,$$

$$\psi(B) = 1 + .2B.$$

Why linear process?

Reason 1 The way of generating dependence is very simple. For example

$$\mathrm{Cov}(a_1Z_1 + a_2Z_2 + a_3Z_3, b_2Z_2 + b_3Z_3) =$$

$$\mathrm{Cov}(a_1Z_1, b_2Z_2 + b_3Z_3) =$$

Reason 2 Well studied in regression/linear algebra. It is easy to interpret and estimation (least squares method) is well developed.

Reason 3 Provides general framework for stationary TS. Linear combination of a given stationary process is again stationary process.

Reason 4 Wold decomposition (Ch2.6): Every weakly stationary TS can be represented as the sum of linear process and deterministic part.

Why linear process?

Proposition (2.2.1)

Let $\{Y_t\}$ *be any (weakly) stationary process with mean zero and ACVF* γ_Y*. Now define new process*

$$X_t = \sum_j \psi_j Y_{t-j}, \quad \sum_j |\psi_j| < \infty$$

is again statioanry process with mean zero and

$$\gamma_X(h) = \sum_{j=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} \psi_j \psi_k \gamma_Y(h + k - j)$$

In particualr, if $\{Y_t\} \sim WN(0, \sigma^2)$*, then*

$$\boxed{\gamma_X(h) = \sum_j \psi_j \psi_{j+h} \sigma^2}$$

Examples of linear process: MA(q)

MA(q): moving average of order q

$$\boxed{\begin{aligned} X_t &= Z_t + \theta_1 Z_{t-1} + \ldots + \theta_q Z_{t-q} \\ &= (1 + \theta_1 B + \ldots + \theta_q B^q) Z_t := \theta(B) Z_t \end{aligned}}$$

- Is it linear process?

- Is it stationary?

Examples of linear process: AR(1)

Consider linear process determined only by one parameter ϕ by

$$X_t = \sum_{j=0}^{\infty} \phi^j Z_{t-j}$$

- Is it linear process?

- Is it stationary?

Examples of linear process: AR(1)

Claim: If $|\phi| < 1$, it is equivalent to stationary process given by

$$\boxed{X_t = \phi X_{t-1} + Z_t} \tag{1}$$

Deduce that

$$X_t = Z_t + \phi X_{t-1} = Z_t + \phi(Z_{t-1} + \phi X_{t-2})$$

$$= ... = Z_t + \phi Z_{t-1} + \phi^2 Z_{t-2} + \ldots + \phi^{M+1} X_{t-M-1}. \tag{2}$$

Hence, as $M \to \infty$, (2) converges to $\sum_{j=0}^{\infty} \phi^j Z_{t-j}$.

AR(1) process with $|\phi| \geq 1$

- If $|\phi| = 1$, then AR(1) process is non-stationary

$$X_t = X_{t-1} + Z_t \quad \text{(Random Walk)}$$
$$X_t = -X_{t-1} + Z_t$$

- If $|\phi| > 1$, then $X_t = \phi X_{t-1} + Z_t$ does not converge.
  However,

$$X_{t+1} = \phi X_t + Z_{t+1}$$

$$X_t = \frac{1}{\phi} X_{t+1} - \frac{1}{\phi} Z_{t+1} = \frac{1}{\phi}\left(\frac{1}{\phi} X_{t+2} - \frac{1}{\phi} Z_{t+2}\right) - \frac{1}{\phi} Z_{t+1}$$

$$= \frac{1}{\phi^2} X_{t+2} - \frac{1}{\phi^2} Z_{t+2} - \frac{1}{\phi} Z_{t+1} = \frac{1}{\phi^k} X_{t+k} - \frac{1}{\phi^k} Z_{t+k} - \ldots - \frac{1}{\phi} Z_{t+1}.$$

$$\therefore X_t = -\sum_{k=1}^{\infty} \phi^{-k} Z_{t+k}.$$

  It implies that if $|\phi| > 1$, then X_t depends on the future values of Z_s.

Causal process

In practice, we are only interested in the stationary process depending on past values of innovations $\{Z_s, s \leq t\}$.

Definition (Causal Process)

Linear process $\{X_t\}$ *is causal if*

$$\psi_j = 0, \quad \forall j < 0,$$

that is,

$$X_t = \sum_{j=0}^{\infty} \psi_j Z_{t-j}.$$

Invertibility of MA(1)

Consider MA(1)

$$X_t = Z_t + \theta Z_{t-1}$$

and represent Z_t in terms of X_t by

$$Z_t = X_t - \theta Z_{t-1} = X_t - \theta(X_{t-1} - \theta Z_{t-2})$$

$$= .. = X_t - \theta X_{t-1} + \theta^2 X_{t-2} + \ldots + (-\theta)^M X_{t-M}.$$

As argued similarly,

$$Z_t = \sum_{j=0}^{\infty} (-\theta)^j X_{t-j}$$

converges iff $|\theta| < 1$.

Invertibility of MA(1)

In a shorthand notation,

$$X_t = \theta(B)Z_t \Rightarrow \theta(B)^{-1}X_t = Z_t.$$

Thus, for MA(1)

$$(1+\theta B)^{-1}X_t = Z_t,$$

where

$$(1+\theta B)^{-1} = \frac{1}{1-(-\theta B)} = 1 - \theta B + (\theta B)^2 - (\theta B)^3 + \dots,$$

and such expansion holds iff $|\theta| < 1$.

Definition (Invertibility)

MA(1) process is called invertible if there is $\{\pi_j\}$ *such that* $\sum_{j=0}^{\infty} |\pi_j| < \infty$ *and*

$$Z_t = \sum_{j=0}^{\infty} \pi_j X_{t-j} \quad \textit{for all } t.$$

ARMA processes

Definition (ARMA(1,1))

The time series $\{X_t\}$ *is an ARMA(1,1) process if*

$$X_t - \phi X_{t-1} = Z_t + \theta Z_{t-1}, \quad Z_t \sim WN(0, \sigma^2)$$

$$(1 - \phi B)X_t = (1 + \theta B)Z_t.$$

- Causal/stationarity: To be a causal process, note that

$$X_t = \frac{1 + \theta B}{1 - \phi B} Z_t = (1 + \theta B)(1 + \phi B + \phi^2 B^2 + \ldots)Z_t$$

$$= (1 + (\theta + \phi)B + (\theta\phi + \phi^2)B^2 + \ldots)Z_t$$

if $|\phi| < 1$. (Can further show this is the unique stationary solution of ARMA(1,1))

ARMA processes

- Invertible: Similarly,

$$Z_t = \frac{1-\phi B}{1+\theta B} X_t = (1-\phi B)(1-\theta B+\theta^2 B^2-\theta^3 B^3+\ldots)X_t$$

$$= (1-(\theta+\phi)B+(\theta+\phi)\theta B^2+\ldots)X_t$$

if $|\theta| < 1$.

- Identifiability: However, this is not everything. If

$$(1-\phi B) = (1+\theta B)$$

that is if $\phi = -\theta$, then we simply have

$$X_t = Z_t,$$

so it is NOT a ARMA(1,1) process but WN. Thus we need $\phi + \theta \neq 0$.

ARMA processes

We have defined AR(1), MA(1) and ARMA(1,1) process in general, but we wish them to have some nice properties for the practical implications and forecasting/estimation purposes. Those are

- AR(1): $|\phi| < 1$ is needed to have causal and stationary solution.
- MA(1):$|\theta| < 1$ is needed to have invertibility.
- ARMA(1,1): above two and identifiability.

- In Chapter 3, we will study higher-order ARMA(p, q) processes and impose similar conditions on coefficients to achieve stationarity, causality, invertibility and identifiability. The key condition is to see when we can expand geometric sum!

Ch2.4 Properties of sample mean and SACF

Estimation of μ

Natural estimator: From LLN, use sample average

$$\overline{X} = \frac{X_1 + \ldots + X_n}{n}.$$

Properties

- Unbiased. $E(\overline{X}) = \mu$.
- Central Limit Theorem for **stationary** TS

$$\frac{\overline{X} - \mu}{\sqrt{\nu/n}} \xrightarrow{d} \mathcal{N}(0,1), \quad \nu = \sum_{h=-\infty}^{\infty} \gamma(h).$$

  ν is calle the **long-run variance**.
- From CLT, we can construct (asymptotic) 95% CI by

$$\overline{X} \pm 1.96\sqrt{\frac{\hat{\nu}}{n}}.$$

Example: MA(1)

Consider the MA(1) process,

$$X_t = Z_t + \theta Z_{t-1}$$

with $Z \sim WN[0, \sigma^2]$. In this case

$$\gamma_X(h) = \begin{cases} (1+\theta^2)\sigma^2 & \text{if } h = 0, \\ \theta\sigma^2 & \text{if } h = \pm 1, \\ 0 & \text{all other } h. \end{cases}$$

So $\nu = (1 + \theta^2 + 2\theta)\sigma^2 = (1+\theta)^2\sigma^2$. A 95% confidence interval for μ would be

$$\bar{x} \pm \frac{1.96(1+\theta)\sigma}{\sqrt{n}}.$$

In practice, we plug-in estimates of θ and σ. Will be covered in Ch5.

Example: AR(1)

Consider the AR(1) process,

$$X_t = \phi X_{t-1} + Z_t \text{ with } |\phi| < 1.$$

From Prop. 2.2.1, note that for $h > 0$,

$$\gamma(h) = \mathrm{Cov}(X_t, X_{t+h}) = \sum_{j=0}^{\infty} \psi_j \psi_{j+h} \sigma^2 = \sigma^2 \sum_{j=0}^{\infty} \phi^j \phi^{j+h}$$

$$= \sigma^2 \phi^h \sum_{j=0}^{\infty} \phi^{2j} = \sigma^2 \phi^h \frac{1}{1-\phi^2}$$

Thus,

$$\boxed{\gamma(h) = \frac{\sigma^2 \phi^{|h|}}{1-\phi^2}.}$$

Example: AR(1) -continued

Therefore

$$\sum_{h=-\infty}^{\infty} \gamma(h) = \frac{\sigma^2}{1-\phi^2} + 2\sum_{h=1}^{\infty} \sigma^2 \frac{\phi^h}{1-\phi^2}$$

$$= \frac{\sigma^2}{1-\phi^2} + \frac{2\sigma^2}{1-\phi^2}\frac{\phi}{1-\phi} = \frac{\sigma^2}{1-\phi^2} \cdot \frac{1+\phi}{1-\phi} = \frac{\sigma^2}{(1-\phi)^2}.$$

Note: Don't confuse $\frac{\sigma^2}{1-\phi^2}$ (formula for $\gamma(0)$) with $\frac{\sigma^2}{(1-\phi)^2}$ (formula for ν).

Example

Suppose that $\sigma^2 = 2$ and $\phi = .75$, and $\bar{x} = .95$ based on 54 observations. Find 95% CI for mean μ.

Q: What is the test result for $H_0 : \mu = 0 \quad H_1 : \mu \neq 0$?

Remark: AR(1) with nonzero μ

- Consider

$$X_t - \mu = \phi(X_{t-1} - \mu) + Z_t, \quad |\phi| < 1, \quad Z_t \sim \mathrm{WN}(0, \sigma^2). \quad (3)$$

Then, arguing similarly we obtain

$$\overline{X} \approx \mathcal{N}\left(\mu, \frac{1}{n} \frac{\sigma^2}{(1-\phi)^2}\right)$$

- Note that (3) can be written as

$$X_t = (1-\phi)\mu + \phi X_{t-1} + Z_t$$

so be careful in notation.

Estimation of ACVF/ACF

Natural estimator: Based on data $X_1, \ldots, X_n$, use sample ACVF and SACF

$$\hat{\gamma}(h) = \frac{1}{n} \sum_{t=1}^{n-|h|} (X_{t+|h|} - \overline{X})(X_t - \overline{X}), \quad \hat{\rho}(h) = \frac{\hat{\gamma}(h)}{\hat{\gamma}(0)}$$

Properties:

- $\hat{\rho}(h)$ is biased estimator for $\rho(h)$, but asymptotically

$$E(\rho(\hat{h})) \to \rho(h) \text{ as } n \to \infty$$

- Note that

$$\hat{\Gamma}_k := \begin{pmatrix} \hat{\gamma}(0) & \hat{\gamma}(1) & \ldots & \hat{\gamma}(k-1) \\ \hat{\gamma}(1) & \hat{\gamma}(0) & \ldots & \hat{\gamma}(k-2) \\ \vdots & \vdots & \ddots & \vdots \\ \hat{\gamma}(k-1) & \ldots & \hat{\gamma}(1) & \hat{\gamma}(0) \end{pmatrix},$$

  is non-negative definite!

Estimation of ACVF/ACF

Indeed:

$$\Gamma_k = \frac{1}{n} TT',$$

where

$$T = \begin{pmatrix} 0 & \dots & 0 & (X_1 - \overline{X}) & (X_2 - \overline{X}) & \dots & (X_k - \overline{X}) \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots \\ 0 & (X_1 - \overline{X}) & \dots & (X_k - \overline{X}) & 0 & 0 & 0 \end{pmatrix}.$$

Then, we can factor out n^{-1} to get

$$a'\Gamma_k a = n^{-1}(a'T)(T'a) \geq 0.$$

If it is replaced by $(n-h)$, then $(n-h)$ appears on each element in Γ_k hence n.n.d is not achieved.

Estimation of ACVF/ACF

- Asymptotic Normaltiy

$$\begin{pmatrix} \hat{\rho}(1) \\ \hat{\rho}(2) \\ \vdots \\ \hat{\rho}(k) \end{pmatrix} \overset{d}{=} \mathcal{N}\left(\begin{pmatrix} \rho(1) \\ \rho(2) \\ \vdots \\ \rho(k) \end{pmatrix}, \frac{1}{n} \begin{pmatrix} w_{11} & w_{12} & \dots & w_{1k} \\ w_{21} & w_{22} & \dots & 2_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ w_{k1} & \dots & \dots & w_{kk} \end{pmatrix} \right)$$

$$w_{ij} = \sum_{k=1}^{\infty} \{\rho(k+i) + \rho(k-i) - 2\rho(i)\rho(k)\}$$
$$\times \{\rho(k+j) + \rho(k-j) - 2\rho(j)\rho(k)\}$$

is known as *Bartlett's* formula.

Example: IID$(0, \sigma^2)$

- Example: If $\{X_t\}$ is IID$(0, \sigma^2)$, then

$$w_{ii} = \sum_{k=1}^{\infty} \{\rho(k+i) + \rho(k-i) - 2\rho(i)\rho(k)\}^2 = 1$$

b/c $\rho(k) = 0$ for all $k \geq 1$ and $\rho(k-i) = 1$ if $k = i$. Similarly,

$$w_{ij} = 0, \qquad \text{if } i \neq j.$$

Therefore,

$$\hat{\boldsymbol{\rho}} \approx \mathcal{N}\left(\mathbf{0}, \frac{1}{n} I_k\right).$$

Recall $\pm 2/\sqrt{n}$ rule to test $H_0 : \rho = 0$ vs $H_1 : \rho \neq 0$!

Example: MA(1)

Consider: $X_t = Z_t + \theta Z_{t-1}$, $Z_t \sim WN(0, \sigma^2)$ and compute w_{ii}.
Recall that

$$\gamma_X(h)=\begin{cases}\sigma^2(1+\theta^2) & \text{if } h=0,\\ \sigma^2\theta & \text{if } h=\pm 1,\\ 0 & \text{if } |h|>1.\end{cases} \quad \rho_X(h)=\begin{cases}1 & \text{if } h=0,\\ \frac{\theta}{1+\theta^2} & \text{if } h=\pm 1,\\ 0 & \text{if } |h|>1.\end{cases}$$

$$w_{11}=\sum_{k=1}^{\infty}\{\rho(k+1)+\rho(k-1)-2\rho(1)\rho(k)\}^2$$

$$=(\rho(0)-2\rho(1)^2)^2+\rho(1)^2=1-3\rho(1)^2+4\rho(1)^4.$$

$$w_{22}=\sum_{k=1}^{\infty}\{\rho(k+2)+\rho(k-2)-2\rho(2)\rho(k)\}^2$$

$$=\rho(-1)^2+\rho(0)^2+\rho(1)^2=1+2\rho(1)^2.$$

You can show that $w_{ii} = 1 + 2\rho(1)^2$ if $i \geq 2$. 95% Confidence bounds for $\rho(1)$ for MA(1) model is given by

$$\hat{\rho}(1) \pm \frac{1.96}{\sqrt{n}}(1 - 3\hat{\rho}(1)^2 + 4\hat{\rho}(1)^4)^{1/2}$$

Ch2.5 Forecasting Stationary Time Series

Forecasting basic setup

- Consider TS with observations $\{X_1, \ldots, X_n\}$. We want to predict future values based on **linear** combination of $\{1, X_1, \ldots, X_n\}$.
- Notation:

$$P_n X_{n+h} = a_0 \cdot 1 + a_1 X_n + \ldots + a_n X_1, \quad a_i \in \mathbb{R}$$

- Two cases will be considered
  - Prediction based on the infinite past (Ch.2.5.3)
    It actually provides simpler and elegant solution. Easy to understand in theoretical perspective.
  - Prediction based on the finite past (Ch.2.5.4)
    This is what actually used in practice. However, due to intensive computing, formulas are a bit complicate.
- Underlying theory is **projection"**, i.e, minimize mean squared prediction errors (MSPE).

Prediction based on infinite past

- Basically, we assume that $\{X_t\}$ is a causal and invertible linear process.

$$X_t = \sum_{j=0}^{\infty} \psi_j Z_{t-j}, \quad \text{(Causal)}$$

$$Z_t = \sum_{j=0}^{\infty} \pi_j X_{t-j}, \quad \text{(Invertible)}$$

  $\{Z_t\} \sim WN(0, \sigma^2)$, $\psi_0 = \pi_0 = 1$ and $\sum_j |\psi_j| < \infty$, $\sum_j |\pi_j| < \infty$.

- h-step ahead prediction (forecasting):

$$\boxed{\widetilde{P}_t X_{t+h} = a_0 + a_1 X_t + a_2 X_{t-1} + \ldots}$$

Prediction based on infinite past

- Consider $h = 1$ first. From the invertibility observe that

$$Z_{t+1} = \sum_{j=0}^{\infty} \pi_j X_{t+1-j} = X_{t+1} + \pi_1 X_t + \pi_2 X_{t-1} + \dots$$

$$X_{t+1} = Z_{t+1} - \pi_1 X_t - \pi_2 X_{t-1} - \dots$$

$$\boxed{X_{t+1} = Z_{t+1} - \sum_{j=1}^{\infty} \pi_j X_{t+1-j}}$$

- Candidate for one-step ahead predictor:

$$\widetilde{P}_t X_{t+1} = \widehat{Z}_{t+1} - \sum_{j=1}^{\infty} \pi_j X_{t+1-j} = -\sum_{j=1}^{\infty} \pi_j X_{t+1-j}$$

since $E(Z_{t+1}) = 0$ so $\widehat{Z}_{t+1} = 0$, and we have already observed $X_t, X_{t-1}, \dots$.

Prediction based on infinite past

- For h-step ahead prediction, we can recursively predict

$$\ldots, X_0, X_1, X_2, \ldots, X_t \Rightarrow \widetilde{P}_t X_{t+1}$$

$$\ldots, X_0, X_1, X_2, \ldots, X_t, \widetilde{P}_t X_{t+1} \Rightarrow \widetilde{P}_t X_{t+2}$$

$$\underbrace{\ldots, X_0, X_1, X_2, \ldots, X_t,}_{\text{observed}} \underbrace{\widetilde{P}_t X_{t+1}, \widetilde{P}_t X_{t+2}, \widetilde{P}_t X_{t+h-1}}_{\text{estimated}} \Rightarrow \widetilde{P}_t X_{t+h}$$

- h-step prediction

$$\boxed{\begin{aligned} \widetilde{P}_t X_{t+1} &= -\sum_{j=1}^{\infty} \pi_j X_{t+1-j} \\ \widetilde{P}_t X_{t+h} &= -\sum_{j=1}^{h-1} \pi_j \widetilde{P}_t X_{t+h-j} - \sum_{j=h}^{\infty} \pi_j X_{t+h-j} \end{aligned}}$$

Prediction based on infinite past

We can, in fact, argue that above procedure is the best!

Theorem

$\widetilde{P}_t X_{t+h}$ is the BLP (Best Linear Predictor) that minimized the mean squared prediction error (MSPE) and resulting MSPE is given by

$$\boxed{\mathrm{MSPE} = E\left(X_{t+h} - \widetilde{P}_t X_{t+h}\right)^2 = \sigma^2 \sum_{j=0}^{h-1} \psi_j^2}$$

since

$$X_{t+h} - \widetilde{P}_t X_{t+h} = \sum_{j=0}^{h-1} \psi_j Z_{t+h-j}$$

Prediction based on infinite past

Keep this BLP/MSPE formula in mind. Also

- Prediction $\Rightarrow$ from invertibility
- MSPE $\Rightarrow$ from causality
- Furthermore, $100(1-\alpha)\%$ Prediction Interval will be

$$\boxed{\widetilde{P}_n X_{n+h} \pm z_{\alpha/2}\sqrt{\text{MSPE}}}$$

Prediction based on infinite past: AR(1)

- AR(1) process: $X_t = \phi X_{t-1} + Z_t$, $|\phi| < 1$
- From the invertibility, we have

$$Z_t = X_t - \phi X_{t-1} \Rightarrow \pi_0 = 1, \pi_1 = -\phi, \pi_j = 0, j \geq 2.$$

$$\begin{aligned} \widetilde{P}_t X_{t+1} &= -\pi_1 X_t = \phi X_t \\ \widetilde{P}_t X_{t+2} &= -\pi_1 \widehat{X}_{t+1} = \phi(\phi X_t) = \phi^2 X_t \end{aligned}$$

- MSPE becomes $\sigma^2 \sum_{j=0}^{h-1} \psi_j^2$. To calculate ψ_j note that

$$X_t = (1 - \phi B)^{-1} Z_t = Z_t + \phi Z_{t-1} + \phi^2 Z_{t-2} + \ldots$$

$$\begin{aligned} h = 1 &\Rightarrow \sigma^2 \psi_0^2 = \sigma^2 \\ h = 2 &\Rightarrow \sigma^2(\psi_0^2 + \psi_1^2) = \sigma^2(1 + \phi^2) \end{aligned}$$

Prediction based on infinite past: MA(1)

- Consider invertible MA(1) process $X_t = Z_t + \theta Z_{t-1}$, $|\theta| < 1$.
- From invertibility, we have

$$Z_t = (1 + \theta B)^{-1} X_t = X_t - \theta X_{t-1} + \theta^2 X_{t-2} - \theta^3 X_{t-3} + \dots$$

$$\pi_0 = 1, \quad \pi_1 = -\theta, \quad \pi_2 = \theta^2$$

$$\begin{aligned}
\widetilde{P}_t X_{t+1} &= -\pi_1 X_t - \pi_2 X_{t-1} - \pi_3 X_{t-2} - \dots \\
&= \theta X_t - \theta^2 X_{t-1} + \theta^3 X_{t-2} + \dots \\
\widetilde{P}_t X_{t+2} &= -\pi_1 \widehat{X}_{t+1} - \pi_2 X_t - \pi_3 X_{t-1} \dots \\
&= \theta(\theta X_t - \theta^2 X_{t-1} + \theta^3 X_{t-2}) - \theta^2 X_t + \theta^3 X_{t-1} = 0
\end{aligned}$$

- Something strange? No, because MA(1) process only has lag-1 dependency so for $h > 1$, X_{t+h} is uncorrelated with X_t, $X_{t-1}, \dots,$.

Prediction based on the finite past

- Observations: $X_1, \ldots, X_n$
- h-step predictor:
  $P_n X_{n+h} = a_0 + a_1 X_n + a_2 X_{n-1} + \ldots + a_n X_1$.
- Want to find (optimal) coefficients $\{a_1, \ldots, a_n\}$ by minimizing MSPE (Mean Squared Prediction Error)

$$\begin{aligned} \mathrm{MSPE} &= E\left(X_{n+h} - P_n X_{n+h}\right)^2 \\ &= E\left(X_{n+h} - (a_0 + a_1 X_n + a_2 X_{n-1} + \ldots + a_n X_1)\right)^2 \end{aligned}$$

- Need to solve "derivative is equal to 0"

$$\begin{aligned} \frac{\partial MSPE}{\partial a_0} &= 0 \\ \frac{\partial MSPE}{\partial a_j} &= 0, \quad j = 1, \ldots, n. \end{aligned}$$

Estimating coefficients I

$$E(X_{n+h} - (a_0 + a_1 X_n + a_2 X_{n-1} + \ldots + a_n X_1))(-2) = 0. \quad (4)$$

$$E\left(-2X_{n+1-j}(X_{n+h} - (a_0 + a_1 X_n + a_2 X_{n-1} + \ldots + a_n X_1))\right) = 0, \quad (5)$$

$j = 1, \ldots, n.$

Since $EX_t = \mu$ for all t, (4) becomes

$$a_0 = \mu \left(1 - \sum_{i=1}^{n} a_i\right) \quad (6)$$

Plug-in (6) to (5) gives that

$$EX_{n+1-j}\left((X_{n+h} - \mu) - \sum_{i=1}^{n} a_i (X_{n+1-i} - \mu)\right) = 0$$

$$E(X_{n+1-j} - \mu)\left((X_{n+h} - \mu) - \sum_{i=1}^{n} a_i (X_{n+1-i} - \mu)\right) = 0$$

Estimating coefficients II

- Equation (5) can be undersood as

$$P_n X_{n+h} \perp (X_{n+h} - P_n X_{n+h})$$

$$\iff E(X_{n+h} - P_n X_{n+h}) X_j = 0, \quad j = 1, \ldots, n.$$

  That is, $P_n X_{n+h}$ is an orthogonal projection onto $\mathrm{span}\{1, X_1, \ldots, X_n\}$

- In practice, use computers to get the solution.