ARMA Models (Chapter 3)

Motivation: For 'most' time series $\{X_t\}$,

Wold Decomposition $\Longrightarrow$ $\{X_t\}$ is a linear TS

$$X_t = \sum_{j=0}^{\infty} \psi_j Z_{t-j}, \quad \{Z_t\} \sim \mathrm{WN}(0,\sigma^2),$$
$$= \psi(B) Z_t ,$$

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

We approximate $\psi(B)$ using a ratio of polynomials which leads us to the class of ARMA models.

1

3.1 ARMA(p,q) Processes

DEFINITION: $\{X_t\}$ is an **ARMA(p,q) process** if $\{X_t\}$ is stationary and if for every t,

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

where $\{Z_t\} \sim \mathrm{WN}(0,\sigma^2)$.

Write ARMA equations as:
$$\phi(B) X_t = \theta(B) Z_t$$

where $\phi(z)$ and $\theta(z)$ are the polynomials

$$\phi(z) = 1 - \phi_1 z - \cdots - \phi_p z^p, \quad \theta(z) = 1 + \theta_1 z + \cdots + \theta_q z^q.$$

2

Existence and Uniqueness: The ARMA eqns $\phi(B)\, X_t = \theta(B)\, Z_t$ have a stationary solution (which is unique) if and only if

$$\phi(z) \neq 0 \quad \text{for } |z| = 1$$

Causality: Solution $\{X_t\}$ is **causal** if we can write

$$X_t = \sum_{j=0}^{\infty} \psi_j Z_{t-j}, \quad \text{with} \quad \sum_{j=0}^{\infty} |\psi_j| < \infty.$$

$\{X_t\}$ causal $\Longleftrightarrow$ $\phi(z) \neq 0$ for $|z| \leq 1$

$$\psi(B) = \frac{\theta(B)}{\phi(B)}$$

(see p.85)

Invertibility: Solution $\{X_t\}$ is **invertible** if we can write

$$Z_t = \sum_{j=0}^{\infty} \pi_j X_{t-j}, \quad \text{with} \quad \sum_{j=0}^{\infty} |\pi_j| < \infty.$$

$\{X_t\}$ invertible $\Longleftrightarrow$ $\theta(z) \neq 0$ for $|z| \leq 1$

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

(see p.86)

Ex 3.1.1 (An ARMA(1,1) process).

$$X_t - .5\,X_{t-1} = Z_t + .4\,Z_{t-1},\ \ \{Z_t\} \sim \mathrm{WN}(0,\sigma^2),$$

AR polynomial: $\phi(z)=1-.5z$ (zero at z=2, so process is causal).

MA polynomial: $\theta(z)=1+.4z$ (zero at z=-2.5, so process is invertible).

MA(∞) Coefficients ψ_j:

$$\psi_0 = 1$$
$$\psi_1 = .4 + .5\,\psi_0 = .4+.5$$
$$\psi_2 = 0 + .5\,\psi_1 = .5(.4+.5),$$
$$\psi_j = .5^{\,j-1}(.4+.5),\quad j > 0.$$

5

Ex 3.1.2 (An AR(2) process).

$$X_t - .7\,X_{t-1} + .1\,X_{t-2} = Z_t\ ,\ \ \{Z_t\} \sim \mathrm{WN}(0,\sigma^2),$$

AR polynomial: $\phi(z)=1-.7z+.1z^2=(1-.5z)(1-.2z)$

(Zeros at z=2 and z=5, so process is causal).

MA(∞) Coefficients ψ_j:

$$\psi_0 = 1$$
$$\psi_1 = .7\,\psi_0 = .7$$
$$\psi_2 = .7\psi_1 - .1\,\psi_0 = .39$$
$$\psi_j = .7\psi_{j-1} - .1\,\psi_{j-2},\quad j > 1.$$

(Coefficients are computed in ITSM.)

6

3.2 ACF and PACF of ARMA Processes

Moving average processes:

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

This is a special case of a linear process so that

$$\gamma(h) = \begin{cases} \sigma^2 \sum_{j=0}^{q-|h|} \theta_j \theta_{j+|h|}, & \text{if } |h| \le q, \\ 0, & \text{if } |h| > q. \end{cases}$$

Remark: ACVF is 0 for all lags $|h| > q$. Converse is also true. If the ACVF of a stationary process $\{X_t\}$ is 0 beyond some lag q, then $\{X_t\}$ is a MA(q) process.

7

ARMA processes: Suppose $\{X_t\}$ is a causal ARMA

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

Take expectations of both sides with X_{t-k},

(1) $\gamma(k) - \phi_1\gamma(k-1) - \cdots - \phi_p\gamma(k-p) = \sigma^2 \sum_{j=0}^{m} \theta_{k+j}\psi_j, \quad 0 \le k < m$

(2) $\gamma(k) - \phi_1\gamma(k-1) - \cdots - \phi_p\gamma(k-p) = 0, \quad k \ge m,$

where m=max(p,q+1). Solve (1) for $\gamma(0), \ldots, \gamma(p)$,

and then use the recursion in (2) to solve for $\gamma(p+1), \ldots$

This is the method used in ITSM (Method 3 in text).

8

Method 2: Difference equation approach:

The general solution to the difference equation (2) is given by

$$\gamma(h) = \sum_{i=1}^{k} \sum_{j=0}^{r_i - 1} A_{i,j} h^j \xi_i^{-h}, \quad h \geq m - p,$$

where the p constants $A_{i,j}$ and $\gamma(h)$, $h < m\text{-}p$ are found by the boundary conditions (1). (The ξ_i are the distinct zeros of the AR polynomial, and the r_i is its multiplicity.)

Example : $X_t - X_{t-1} + 1/4\, X_{t-2} = Z_t + Z_{t-1}$

$\phi(z) = (1-z+.25z^2) = (1-.5z)^2$, $\theta(z) = (1+z)$.

Here

$$m = \max(p, q+1) = 2.$$

$$\gamma(k) - \gamma(k-1) + .25\gamma(k-2) = 0, \quad k \geq 2.$$

General solution

$$\gamma(k) = (A+Bk)\, 2^{-k}, \quad k \geq 0.$$

9

The boundary conditions (1) are

$$\gamma(0) - \gamma(1) + .25\, \gamma(2) = \sigma^2(\psi_0 + \theta_1\psi_1) = \sigma^2(1 + \theta_1 + \phi_1) = 3\sigma^2$$

$$\gamma(1) - \gamma(0) + .25\gamma(-1) = \sigma^2\psi_0 = \sigma^2$$

Replacing $\gamma(0)$, $\gamma(1)$, $\gamma(2)$ with the general solution $\gamma(k) = (A+Bk)\, 2^{-k}$, we obtain

$$3A - 2B = 16\sigma^2$$

$$-3A + 5B = 8\sigma^2$$

Solving, we get

$$A = 32\sigma^2/3, \; B = 8\sigma^2$$

and hence

$$\gamma(k) = (32/3 + 8k)\, 2^{-k}, \quad k \geq 0.$$

10

Complex zeros: For complex valued zeros of the AR polynomial, the solution to the difference equations includes a sinusoidal component. Consider an AR(2) process

$$X_t - \phi_1 X_{t-1} - \phi_2 X_{t-2} = Z_t$$

where $\phi(z) = 1 - \phi_1 z - \phi_2 z^2$ has complex valued zeros ξ_1, ξ_2. In this case, we can write

$$\xi_1 = re^{i\theta} \quad \text{and} \quad \xi_2 = re^{-i\theta} \text{ (complex conjugates).}$$

Solution is given by

$$\gamma(k) = A\, r^{-k} e^{i\theta k} + B\, r^{-k} e^{-i\theta k} \qquad k \geq 0.$$

where A and B are complex conjugates (see Example 3.2.4 for more details.)

Fig 3.1. AR(2): $(1-.5B)(1-.2B)X_t = Z_t$

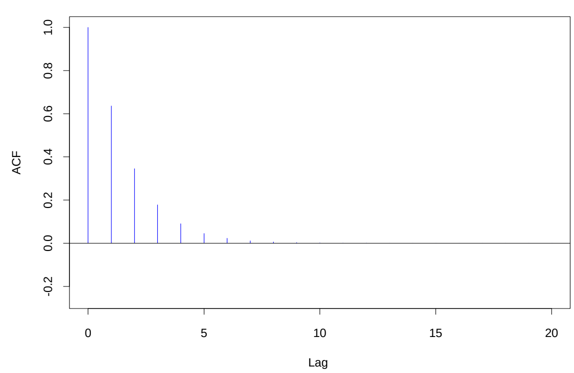


Fig 3.2. AR(2): $(1-.9B)(1-.5B)X_t = Z_t$

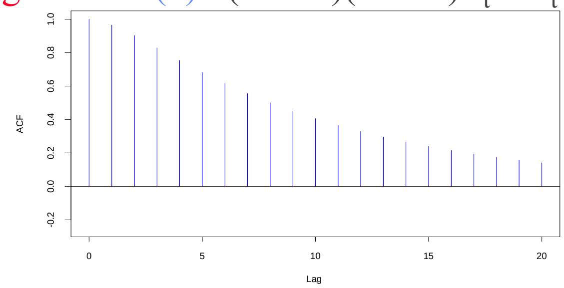


Fig 3.3. AR(2): $(1+.9B)(1-.5B)X_t = Z_t$

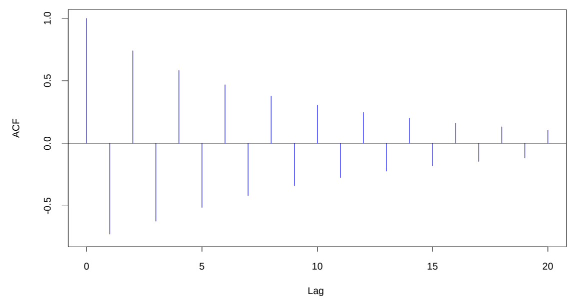
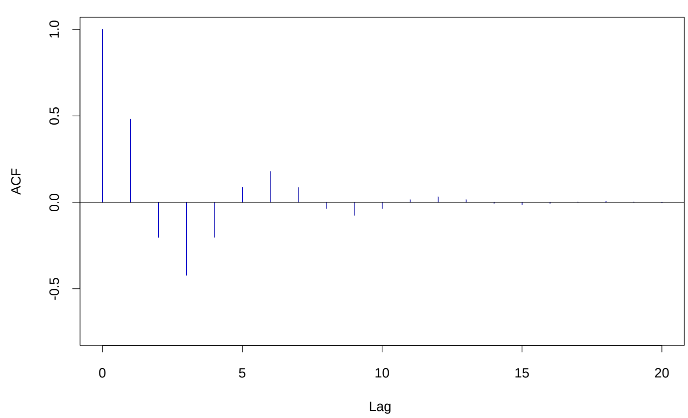


Fig 3.4. AR(2): $(1-.75\,B + .05625B^2)\,X_t = Z_t$

$$\phi(z) = (1-.375(1-i\sqrt{3})z)(1-.375(1+i\sqrt{3})z)$$

with zeros at $2.667e^{i1.047}$ and $2.667e^{-i1.047}$



3.2.3 The Partial Autocorrelation Function

DEFINITION: The **partial autocorrelation function (PACF)** of a stationary process $\{X_t\}$ is the function α given by

$$\alpha(0) = 1$$

$$\alpha(h) = \phi_{hh}\ ,\ \ h > 0$$

where ϕ_{hh} is the coefficient of X_1 for predicting X_{h+1} in terms of $X_1, \ldots, X_h$, i.e.

$$P_h X_{h+1} = \phi_{h1} X_h + \cdots + \phi_{hh} X_1$$

15

Interpretation of the PACF:

$\alpha(h)$ = correlation between $X_1 - P(X_1|X_2, \ldots, X_h)$ and $X_{h+1} - P(X_{h+1}|X_2, \ldots, X_h)$.

Ex 3.2.6 (An AR(p) process).

$$X_t = \phi_1 X_{t-1} + \cdots + \phi_p X_{t-p} + Z_t$$

Now for $h = p, p+1, p+2, \ldots$,

$$P_h X_{h+1} = \phi_1 X_h + \cdots + \phi_p X_{h+1-p}\ ,$$

so that $\alpha(p) = \phi_p$

and $\alpha(h) = 0,\ \ h > p.$

16

Summary of ACF and PACF:

AR(p) processes:

$$\begin{cases} \rho(h) \text{ decreases geometrically as } h \longrightarrow \infty \\ \alpha(h) \text{ is } 0 \text{ for } h > p. \end{cases}$$

MA(q) processes:

$$\begin{cases} \rho(h) \text{ is } 0 \text{ for } h > q \\ \alpha(h) \text{ decreases geometrically as } h \longrightarrow \infty \end{cases}$$

17

3.2.4 Examples

Ex 3.2.8 (Overshorts; OSHORTS.TSM)

y_t = measured amount of fuel in storage tank at the end of day t.

a_t = measured amount sold – amount delivered

$x_t = y_t - y_{t-1} + a_t$ (overshorts)

18

Figure 3.5 (TS plot of overshorts)

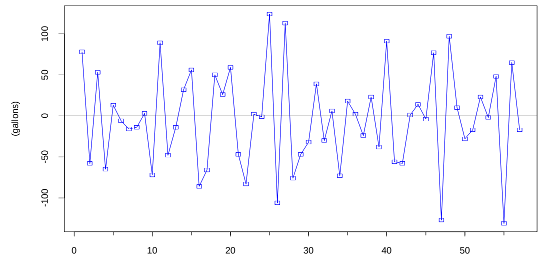
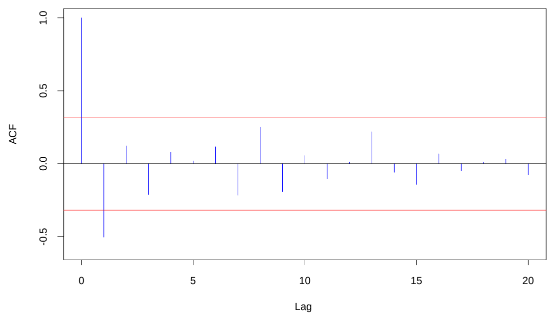


Figure 3.6 (Sample ACF of overshorts)



ACF plot suggests MA(1) model

$$X_t = \mu + Z_t + \theta Z_{t-1} \text{ , } \{Z_t\} \sim \mathrm{WN}(0,\sigma^2)$$

Estimation of parameters:

mean: $\bar{x}_{57} = -4.035$

(Use method of moments for estimating θ and σ^2.)

$$(1+\theta^2)\sigma^2 = \hat{\gamma}(0) = 3415.72$$

$$\theta \sigma^2 = \hat{\gamma}(1) = -1719.95$$

Using the approximate solution $\theta = -1$ and $\sigma^2 = 1708$,

$$X_t = -4.035 + Z_t - Z_{t-1} \text{ , } \{Z_t\} \sim \mathrm{WN}(0,1708)$$

A Structural Model Derivation

$$Y_t = y_t^* + U_t \qquad \{U_t\} \sim \mathrm{WN}(0,\sigma_u^2)$$

$$\begin{cases} y_t^* = \text{true amount of fuel in tank} \\ U_t = \text{measurement error} \end{cases}$$

$$A_t = a_t^* + V_t \qquad \{V_t\} \sim \mathrm{WN}(0,\sigma_v^2)$$

$$\begin{cases} a_t^* = \text{true amount of fuel sold - delivered} \\ V_t = \text{measurement error} \end{cases}$$

$$y_t^* = \mu + y_{t-1}^* - a_t^*, \quad \mu = \text{leakage term } (<0)$$

$$X_t = Y_t - Y_{t-1} + A_t = \mu + U_t - U_{t-1} + V_t$$

21

The series of overshorts is then given by

$$X_t = Y_t - Y_{t-1} + A_t = \mu + U_t - U_{t-1} + V_t$$

with

$$\mathrm{E}(X_t) = \mu,$$

$$\gamma(h) = \begin{cases} 2\sigma_u^2 + \sigma_v^2, & \text{if } h = 0, \\ -\sigma_u^2, & \text{if } |h| = 1, \\ 0, & \text{if } |h| > 1. \end{cases}$$

It follows that $\{X_t\}$ is an MA(1) process with

$$\rho(1) = \theta/(1+\theta^2) = -\sigma_u^2/(2\sigma_u^2 + \sigma_v^2)$$

and $\theta = -1$ if and only if $\sigma_v^2 = 0$. (If so then the model is noninvertible.)

22

The series of overshorts is then given by

$$X_t = Y_t - Y_{t-1} + A_t = \mu + U_t - U_{t-1} + V_t$$

with

$$E(X_t) = \mu,$$

$$\gamma(h) = \begin{cases} 2\sigma_u^2 + \sigma_v^2, & \text{if } h = 0, \\ -\sigma_u^2, & \text{if } |h| = 1, \\ 0, & \text{if } |h| > 1. \end{cases}$$

It follows that $\{X_t\}$ is an MA(1) process with

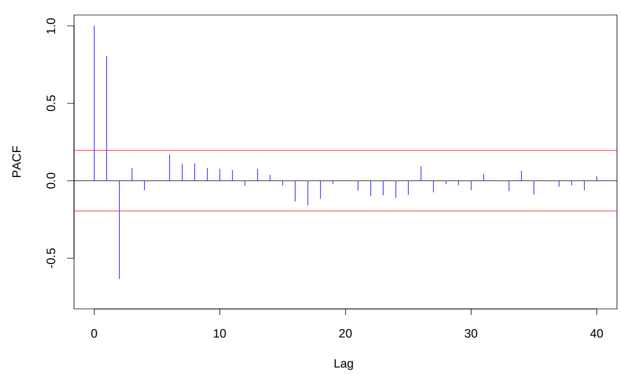
$$\rho(1) = \theta/(1+\theta^2) = -\sigma_u^2/(2\sigma_u^2 + \sigma_v^2)$$

and $\theta = -1$ if and only if $\sigma_v^2 = 0$.

23

Ex 3.2.9. (Sunspot Numbers; SUNSPOTS.TSM)

Figure 3.7. PACF of sunpsots



Plot suggests AR(2) model. Using method of moments,

$$X_t - 1.318\, X_{t-1} + .634\, X_{t-2} = Z_t, \quad \{Z_t\} \sim \text{WN}(0,232.9)$$

24

3.3 Forecasting ARMA Processes

Suppose $\{X_t\}$ is a stationary TS with mean zero.

Durbin-Levinson:

$$P_n X_{n+1} = \phi_{n1} X_n + \cdots + \phi_{n,n} X_1 , \quad n \geq 1$$

Innovations algorithm:

$$P_n X_{n+1} = \theta_{n1}(X_n - \hat{X}_n) + \cdots + \theta_{nn}(X_1 - \hat{X}_1),$$

25

AR(p) Processes:

$$P_n X_{n+1} = \phi_1 X_n + \cdots + \phi_p X_{n+1-p} , \quad n \geq p$$

MA(q) Processes:

$$P_n X_{n+1} = \theta_{n1}(X_n - \hat{X}_n) + \cdots + \theta_{nq}(X_{n+1-q} - \hat{X}_{n+1-q}),$$

where $\theta_{n1}, \ldots, \theta_{nq}$ are computed recursively from the innovations algorithm.

ARMA(p,q) Processes:

$$P_n X_{n+1} = \phi_1 X_n + \cdots + \phi_p X_{n+1-p} + \theta_{n1}(X_n - \hat{X}_n) + \cdots + \theta_{nq}(X_{n+1-q} - \hat{X}_{n+1-q})$$

26

We could apply the innovations algorithm directly to predict causal ARMA processes. However, there is a drastic simplification if we apply the IA to a transformed process. Suppose $\{X_t\}$ satisfies the causal ARMA equations,

$$\phi(\mathrm{B})X_t = \theta(\mathrm{B})Z_t \qquad \{Z_t\} \sim \mathrm{WN}(0, \sigma^2).$$

Now consider the transformed process $\{W_t\}$ defined by

$$\begin{cases} W_t = \sigma^{-1}X_{t,} & t = 1,\ldots,m = \max(p,q) \\ W_t = \sigma^{-1}\phi(B)X_{t,} & t > m \end{cases}$$

Note for $t > m$, W_t is the MA(q) process,

$$W_t = \sigma^{-1}\phi(B)X_t = \sigma^{-1}\theta(B)Z_t$$

27

The ACVF for $\{W_t\}$ can be computed by the simple (but tedious) formula

$$\kappa(i,j) = \begin{cases} \sigma^{-2}\gamma(i-j), & 1 \le i, j \le m, \\ \sigma^{-2}(\gamma(i-j) - \sum_{r=1}^{p}\phi_r\gamma(r-|i-j|)), & \min(i,j) \le m < \max(i,j) \le 2m, \\ \sum_{r=0}^{q}\theta_r\theta_{r+|i-j|}, & \min(i,j) > m, \\ 0, & \text{otherwise} \end{cases}$$

From the definition of W_t, we have

$$H_n = sp\{X_1, X_2, \ldots, X_n\} = sp\{W_1, W_2, \ldots, W_n\}$$

and

$$\begin{cases} \hat{W}_{n+1} = \sum_{j=1}^{n}\theta_{nj}(W_{n+1-j} - \hat{W}_{n+1-j}), & 1 \le n < m, \\ \hat{W}_{n+1} = \sum_{j=1}^{q}\theta_{nj}(W_{n+1-j} - \hat{W}_{n+1-j}), & n \ge m, \end{cases}$$

The coefficients θ_{nj} and r_n are computed from the IA applied to W_t.

28

Now the predictor of the X_t process can be computed from W_t process via

$$\begin{cases} \hat{W}_t = \sigma^{-1}\hat{X}_t, & t = 1,\ldots,m, \\ \hat{W}_t = \sigma^{-1}(\hat{X}_t - \phi_1 X_{t-1} - \cdots - \phi_p X_{t-p}), & t > m, \end{cases}$$

and

$$X_t - \hat{X}_t = \sigma(W_t - \hat{W}_t) \quad \text{for all } t \geq 1$$

It follows that

$$\begin{cases} \hat{X}_{n+1} = \sum_{j=1}^{n} \theta_{nj}(X_{n+1-j} - \hat{X}_{n+1-j}), & 1 \leq n < m, \\ \hat{X}_{n+1} = \phi_1 X_n + \cdots + \phi_p X_{n+1-p} + \sum_{j=1}^{q} \theta_{nj}(X_{n+1-j} - \hat{X}_{n+1-j}), & n \geq m, \end{cases}$$

and

$$E(X_{n+1} - \hat{X}_{n+1})^2 = \sigma^2 E(W_{n+1} - \hat{W}_{n+1})^2 = \sigma^2 r_n \ .$$

29

Remarks:

1. The representation for the predictor

$$\hat{X}_{n+1} = \phi_1 X_n + \cdots + \phi_p X_{n+1-p} + \sum_{j=1}^{q} \theta_{nj}(X_{n+1-j} - \hat{X}_{n+1-j})$$

only requires storage of the p previous observations and q previous innovations.

2. If $\{X_t\}$ is invertible, then $\theta_{nj} \to \theta_j$ and $r_n \to 1$ geometrically as $n \to \infty$.

3. Innovations representation of $\{X_t\}$:

$$X_{n+1} = \phi_1 X_n + \cdots + \phi_p X_{n+1-p} + X_{n+1} - \hat{X}_{n+1} + \sum_{j=1}^{q} \theta_{nj}(X_{n+1-j} - \hat{X}_{n+1-j})$$

30

Example 3.3.3 (ARMA(1,1):

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

Then for $n \geq 1$

$$\hat{X}_{n+1} = \phi_1 X_n + \theta_{n1}(X_n - \hat{X}_n)$$

To compute θ_{n1}, we have $\gamma(0) = \sigma^2(1+2\theta\phi+\theta^2)/(1-\phi^2)$ and hence

with $W_1 = \sigma^{-1} X_1$, $W_t = \sigma^{-1}(X_t - \phi X_{t-1}) = \sigma^{-1}(Z_t + \theta Z_{t-1})$, $t > 1$,

$$\kappa(i,j) = \begin{cases} (1+2\phi\theta+\theta^2)/(1-\phi^2), & i=j=1, \\ 1+\theta^2, & i=j\geq 2, \\ \theta, & |i-j|=1, i\geq 1, \\ 0, & \text{otherwise} \end{cases}$$

31

We obtain the recursions

$$\begin{cases} r_0 = (1+2\theta\phi+\theta^2)/(1-\phi^2) \\ \theta_{n1} = \theta / r_n \\ r_n = 1+\theta^2 - \theta^2 / r_{n-1} \end{cases}$$

These recursions are demonstrated in the table below with

$\phi = 0.2$ and $\theta = 0.4$. $\hat{X}_{n+1} = \phi_1 X_n + \theta_{n1}(X_n - \hat{X}_n)$

n	X_{n+1}	r_n	θ_{n1}	$\hat{X}_{n+1}$
0	−1.100	1.3750		0
1	0.514	1.0436	0.2909	−0.5340
2	0.116	1.0067	0.3833	0.5068
3	−0.845	1.0011	0.3973	−0.1321
4	0.872	1.0002	0.3996	−0.4539
5	−0.467	1.0000	0.3999	0.7046
6	−0.977	1.0000	0.4000	−0.5620
7	−1.699	1.0000	0.4000	−0.3614
8	−1.228	1.0000	0.4000	−0.8748
9	−1.093	1.0000	0.4000	−0.3869
10		1.0000	0.4000	−0.5010

32

h-step prediction

From the representation (see Remark 3)

$$X_{n+h} = \phi_1 X_{n+h-1} + \cdots + \phi_p X_{n+h-p} + X_{n+h} - \hat{X}_{n+h} + \sum_{j=1}^{q} \theta_{n+h-1,j}(X_{n+h-j} - \hat{X}_{n+h-j})$$

we have, by applying the operator P_n = projection onto $\text{sp}\{X_1, \ldots, X_n\}$,

$$P_n X_{n+h} = \phi_1 P_n X_{n+h-1} + \cdots + \phi_p P_n X_{n+h-p} + \sum_{j=h}^{q} \theta_{n+h-1,j}(X_{n+h-j} - \hat{X}_{n+h-j}),$$

for $h > \max(p,q) - n$, where $P_n X_t = X_t$ for $t \leq n$.

Note: The h-step prediction errors are a bit more complicated to compute.

33

Prediction Bounds for Stationary Gaussian Processes

Recall that for a Gaussian time series,

$$P_n X_{n+h} = E(X_{n+h} \mid X_1, \ldots, X_n)$$

and

$$X_{n+h} - P_n X_{n+h} \sim N(0, \sigma_n^2(h)),$$

where

$$\sigma_n^2(h) = E(X_{n+h} - P_n X_{n+h})^2.$$

So,

$$P_n X_{n+h} \pm 1.96\sigma_n(h)$$

is a 95% ***prediction interval*** for X_{n+h}.

34