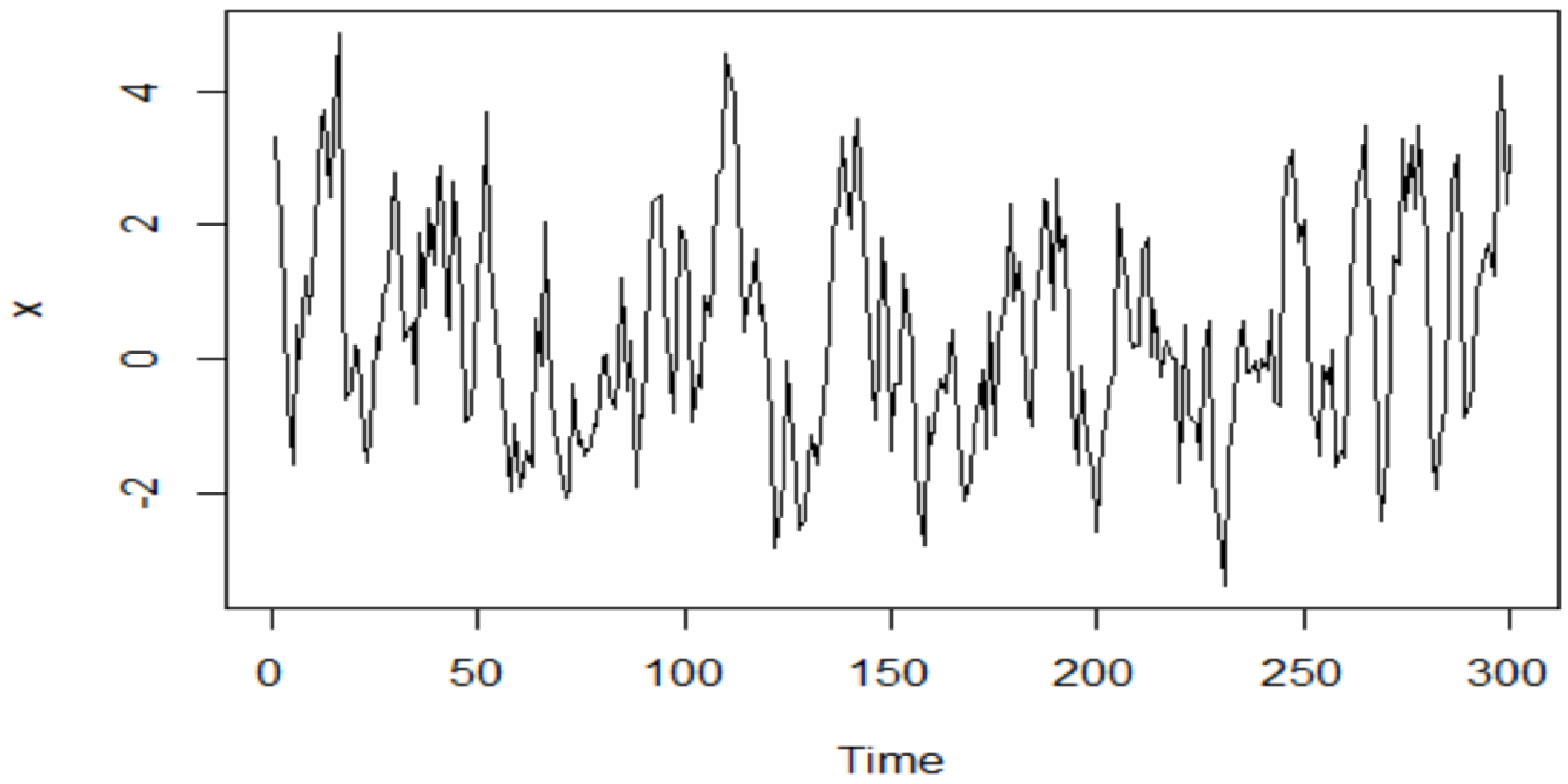
Autocorrelation and Partial Autocorrelation (ACF and PACF)[1],[2]



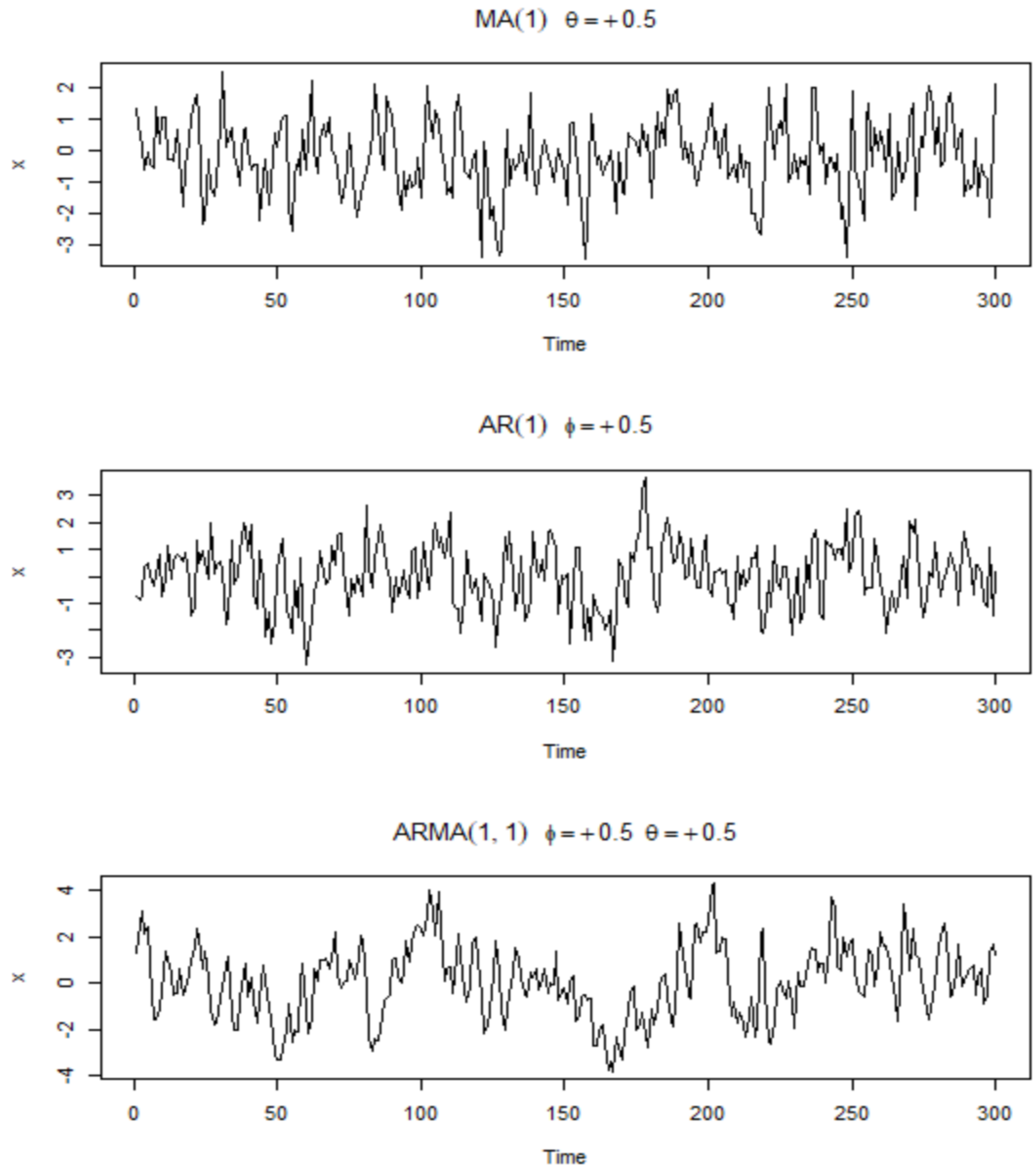
MA(q)? q=?

AR(p) p=?

ARMA(p,q)?

[1] Robert Shumway, David Stoffer. Time Series Analysis and its Applications with R Examples, 4th edition, Springer, 2017 (chapter 3)
[2] Peter Brockwell, Richard Davis. Time Series: Theory and Methods, Springer-Verlag, 1987 (chapter 3)

ACF and PACF



ACF

1) The MA(q) model

$$X_t = \theta(B)\, w_t \qquad \theta(B) = 1 + \theta_1 B + \dots + \theta_q B^q, \quad \theta_q \neq 0$$

$$E(X_t) = \sum_{j=0}^{q} \theta_j\, E(w_{t-j}) = 0 \qquad \theta_0 = 1$$

$$\gamma(h) = \operatorname{cov}(x_{t+h}, x_t) = \operatorname{cov}\left(\sum_{j=0}^{q} \theta_j w_{t+h-j}, \sum_{k=0}^{q} \theta_k w_{t-k}\right)$$

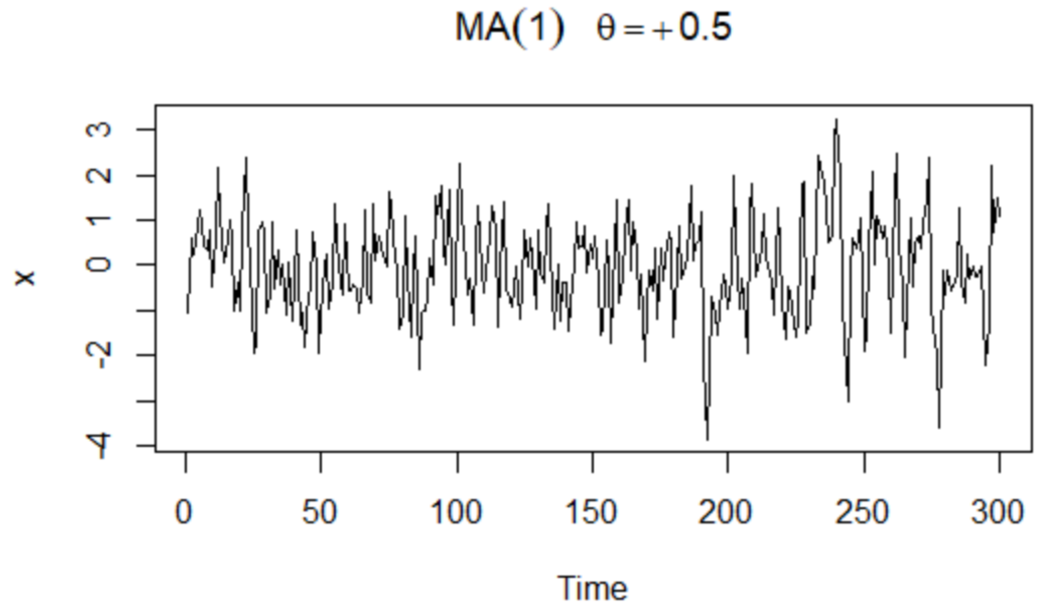
$$= \begin{cases} \sigma_w^2 \sum_{j=0}^{q-h} \theta_j \theta_{j+h}, & 0 \le h \le q \\ 0, & h > q \end{cases}$$

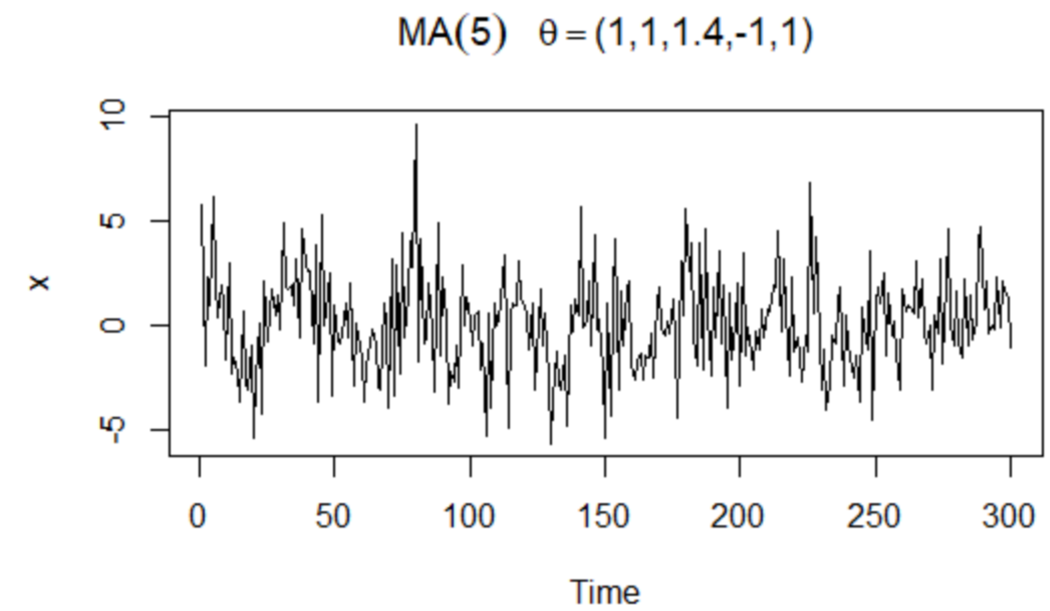
$$\gamma(h) = \gamma(-h)$$

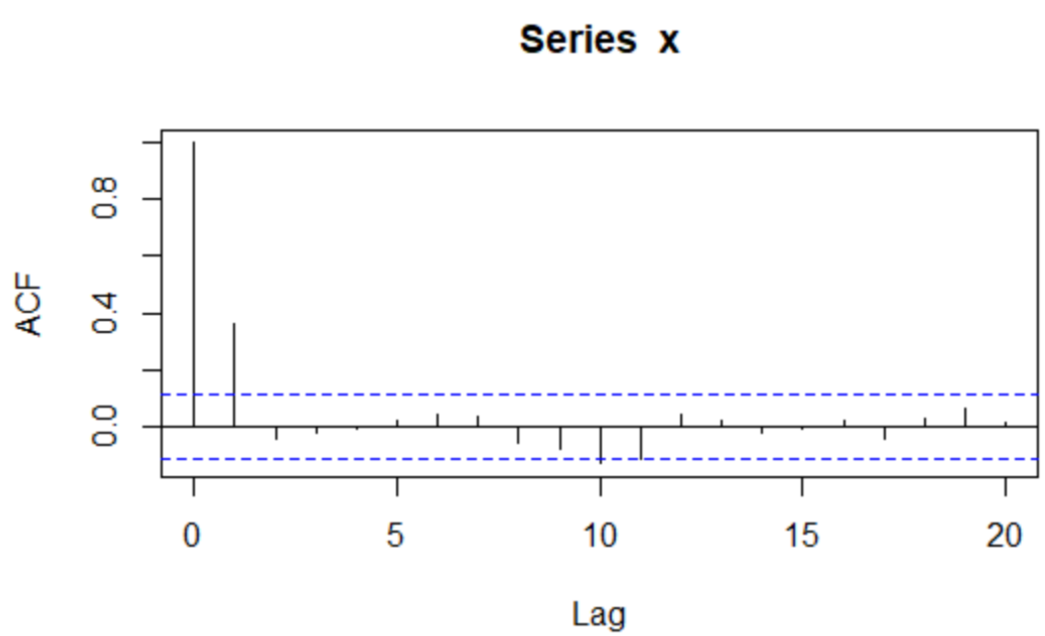
$$\rho(h) = \frac{\gamma(h)}{\gamma(0)}$$

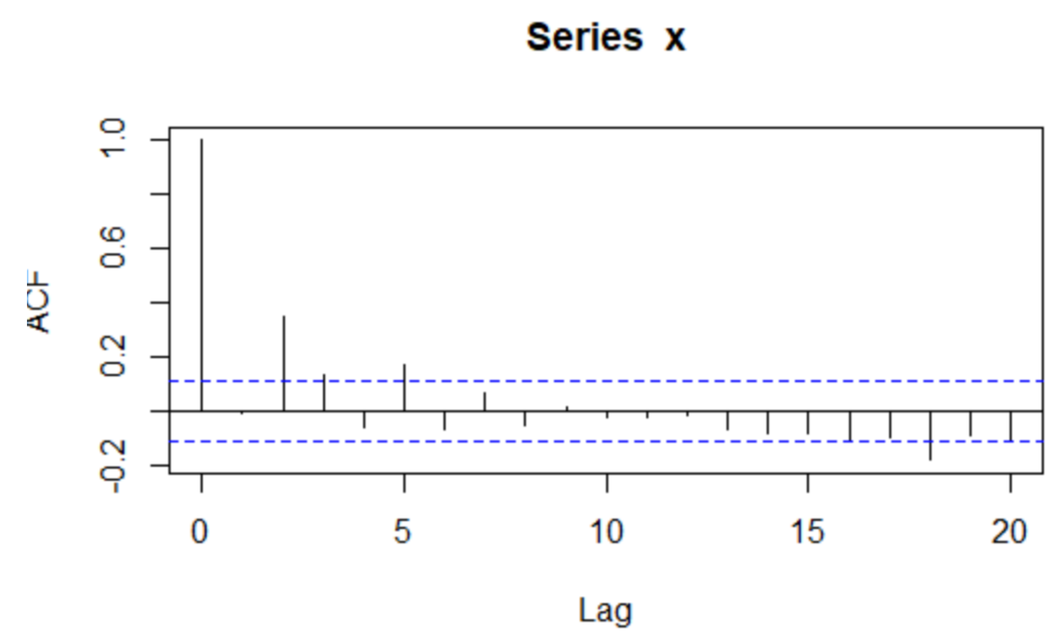
ACF

For an MA(q) process, $\gamma(h)$ is cut off after q lags.









ACF

2) The causal ARMA(p,q) model

$$\phi(B)\, X_t = \theta(B)\, w_t$$

$$\underset{\mu}{\underset{\parallel}{E(X_t)}} - \phi_1 \underset{\mu}{\underset{\parallel}{E(X_{t-1})}} - \dots - \phi_p \underset{\mu}{\underset{\parallel}{E(X_{t-p})}} = 0$$

$$\left.\begin{array}{l} \mu(1-\phi_1-\dots-\phi_p)=0 \\ \text{the zeros of } \phi(z) \text{ are outside the unit circle} \end{array}\right\} \Rightarrow \mu = 0$$

The process is causal $\Rightarrow X_t = \sum_{j=0}^{\infty} \psi_j \cdot w_{t-j}\,, \quad \psi_0 = 1$

$$\phi(B)\, X_t = \theta(B) \cdot w_t \quad \Big| \cdot X_{t-h}$$

ACF

$$x_t x_{t-h} - \phi_1 x_{t-1} x_{t-h} - \dots - \phi_p x_{t-p} x_{t-h} = w_t x_{t-h} + \theta_1 w_{t-1} x_{t-h} + \dots$$
$$+ \theta_q w_{t-q} x_{t-h} \quad \Big| \text{ apply } E$$

$$\gamma(h) - \phi_1 \gamma(h-1) - \dots - \phi_p \gamma(h-p) = E(w_t \cdot x_{t-h}) + \dots E(\theta_q w_{t-q} x_{t-h})$$
$$x_{t-h} = \sum_{j=0}^{\infty} \psi_j w_{t-h-j} = \sum_{j=h}^{\infty} \psi_{j-h} \cdot w_{t-j}$$

ACF

$$E(\theta_j \cdot w_{t-j} \cdot x_{t-h}) = \theta_j \, E\left(w_{t-j} \cdot \sum_{k=h}^{\infty} \psi_{k-h} \cdot w_{t-k}\right)$$

$$= \theta_j \cdot \psi_{j-h} \cdot \sigma_w^2 \,, \quad h \le j \le q$$

We get the difference equations

$$\gamma(h) - \phi_1 \gamma(h-1) - \dots - \phi_p \gamma(h-p) = 0 \,, \quad h \ge \max(p, q+1)$$

with initial conditions

$$\gamma(h) - \phi_1 \gamma(h-1) - \dots - \phi_p \gamma(h-p) = \sigma_w^2 \sum_{j=h}^{q} \theta_j \cdot \psi_{j-h} \,, \quad 0 \le h < \max(p, q+1).$$

[for the closed form solution, see Peter Brockwell, Richard Davis. Time Series: Theory and Methods, Springer-Verlag, 1987 – pag 91 –]

ACF

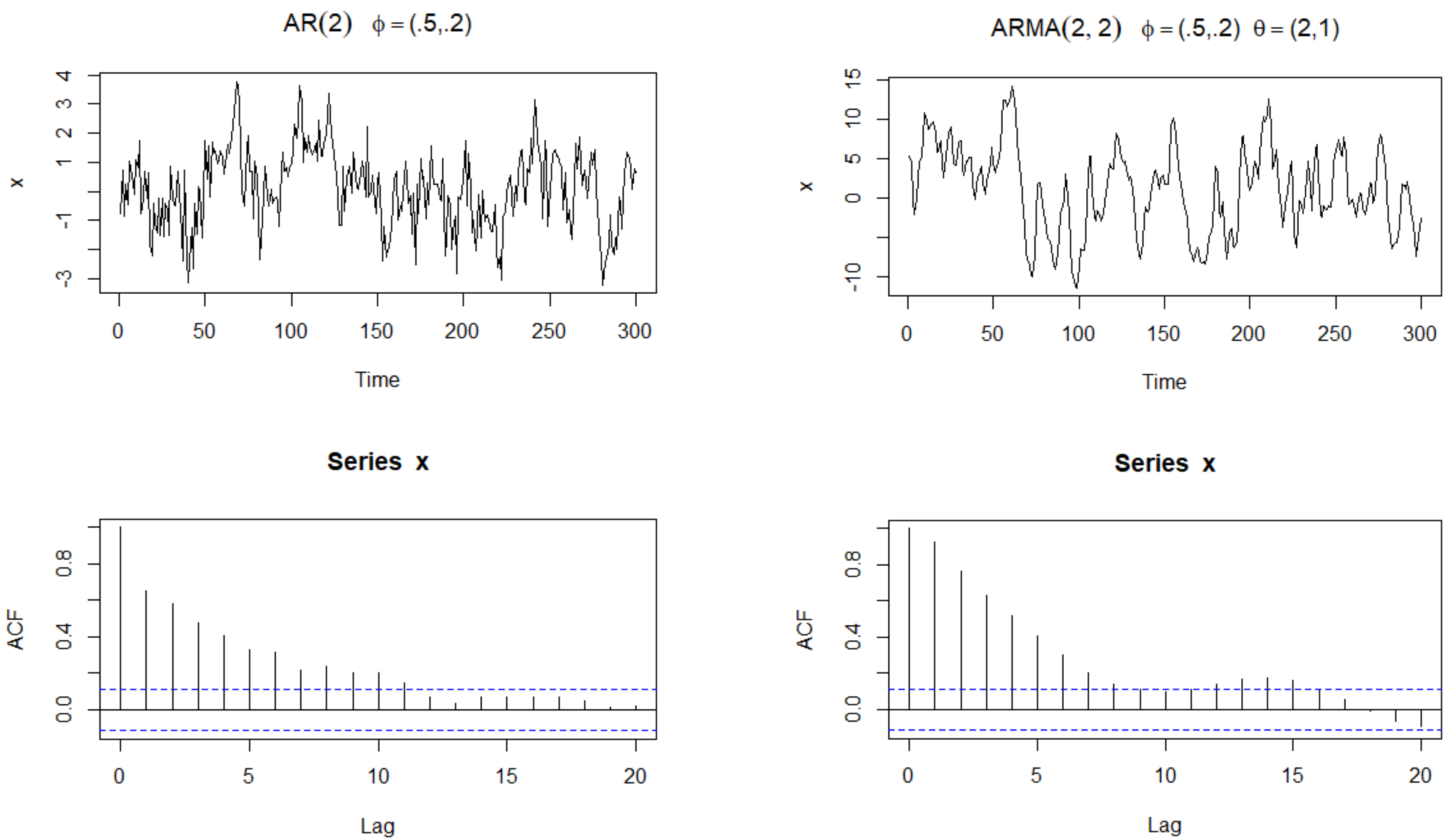
3) The causal AR(p) model

$$\gamma(h) = \sum_{i=1}^{p} \phi_i \cdot \gamma(h-i) \quad , h > 0$$

$$\gamma(0) = \sum_{i=1}^{p} \phi_i \cdot \gamma(i) + \sigma_w^2$$

With the ACF it is unlikely to see the difference between an ARMA model and an AR model.

ACF



PACF

It removes the linear effects of $x_1, \dots, x_{h-1}$ in the correlation between x_h and x_0.

$$\dots x_0, \underbrace{x_1, \dots, x_{h-1}}_{\text{partialled out}}, x_h, \dots$$

Consider the AR(1) model

$$x_t = \phi x_{t-1} + w_t$$

$$\gamma(2) = \operatorname{cov}(x_t, x_{t-2}) = \operatorname{cov}(\phi x_{t-1} + w_t, x_{t-2})$$

$$= \operatorname{cov}(\phi^2 x_{t-2} + \phi w_{t-1} + w_t, x_{t-2}) = \phi^2 \gamma(0)$$

We can partial out the effect of x_{t-1} in the dependence of x_t on x_{t-2}

$$\operatorname{cov}(x_t - \phi x_{t-1}, x_{t-2} - \phi x_{t-1}) = \operatorname{cov}(w_t, x_{t-2} - \phi x_{t-1}) = 0$$

PACF

A Hilbert space $\mathcal{H}$ is an inner-product space which is complete, i.e. every Cauchy sequence $\{X_n\}$ (that is $\|X_n - X_m\| \to 0$ as $m, n \to \infty$) converges in norm to $X \in \mathcal{H}$:

$$\|X_n - X\| \to 0 \text{ as } n \to \infty$$

Example – the space $L^2(\Omega, \mathcal{K}, P)$ with the inner product $\langle X, Y \rangle = E(XY)$ and $E(X^2) < \infty$.

The closed span of $\{X_1, \ldots, X_k\}$ is the set of all linear combinations $\alpha_1 X_1 + \ldots + \alpha_k X_k$, $\alpha_1 \ldots \alpha_k \in \mathbb{R}$.

PACF

The best linear predictor of X in $\overline{sp}\{X_1, \dots, X_k\}$ is

$$P_{\overline{sp}\{X_1,\dots,X_k\}} X = \sum_{i=1}^{k} \alpha_i X_i .$$

where $\alpha_1, \dots \alpha_k$ satisfy

$$\sum_{i=1}^{n} \alpha_i E(X_i X_j) = E(X \cdot X_j) \qquad j = 1, \dots, n$$

Also, $\alpha_1 \dots \alpha_k$ minimize $E\left[\left(X - \sum_{i=1}^{n} \alpha_i X_i\right)^2\right]$ (by the projection theorem) – for more, see Chapter 2 in (Brockwell & Davis)

PACF

<u>Def</u>. The partial autocorrelation function (PACF) of a stationary process X_t, denoted ϕ_{hh}, is defined as:

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

$$\phi_{hh} = \mathrm{corr}\left(X_{t+h} - \hat{X}_{t+h},\ X_t - \hat{X}_t\right) \qquad h \geq 2$$

where $\hat{X}_{t+h} = P_{\overline{sp}\{X_{t+1}, \dots, X_{t+h-1}\}} X_{t+h}$

$$\hat{X}_t = P_{\overline{sp}\{X_{t+1}, \dots, X_{t+h-1}\}} X_t$$

PACF

<u>Examples</u> The PACF of an AR(1)

$$X_t = \phi X_{t-1} + w_t, \qquad |\phi| < 1$$

$$\phi_{11} = \rho(1) = \phi$$

$$\phi_{22} = \mathrm{corr}\left(X_{t+2} - \hat{X}_{t+2}, X_t - \hat{X}_t \right)$$

$$\hat{X}_{t+2} = \alpha X_{t+1}$$

α minimizes $E\left[\left(X_{t+2} - \hat{X}_{t+2}\right)^2\right] = E\left[(X_{t+2} - \alpha X_{t+1})^2\right]$

$$= \gamma(0) - 2\alpha\gamma(1) + \alpha^2\gamma(0)$$

$$\frac{\partial E}{\partial \alpha} = 0 \implies \alpha = \frac{\gamma(1)}{\gamma(0)} = \rho(1) = \phi$$

PACF

$$\hat{X}_t = \beta X_{t+1} \text{ so that } \beta \text{ minimizes } E\left[(X_t - \hat{X}_t)^2\right] \Rightarrow$$
$$\beta = \phi$$

$$\phi_{22} = \operatorname{corr}\left(X_{t+2} - \phi X_{t+1},\ X_t - \phi X_{t+1}\right)$$
$$= \operatorname{corr}\left(W_{t+2},\ X_t - \phi X_{t+1}\right) = 0$$

PACF

The PACF of an AR(p)

$$X_{t+h} = \sum_{j=1}^{p} \phi_j \cdot X_{t+h-j} + w_{t+h}$$

$$\phi_{hh} = \text{corr}\left(X_{t+h} - \hat{X}_{t+h},\ X_t - \hat{X}_t\right)$$

$$\hat{X}_{t+h} = P_{\overline{sp}\{X_{t+1}, \ldots, X_{t+h-1}\}} X_{t+h}$$

for $h > p$, it can be proven that $\hat{X}_{t+h} = \sum_{j=1}^{p} \phi_j \cdot X_{t+h-j}$

$\Rightarrow \phi_{hh} = \text{corr}\left(w_{t+h},\ X_t - \hat{X}_t\right) = 0$ because $X_t - \hat{X}_t$ depends only on $w_{t+h-1}, w_{t+h-2} \ldots$

PACF

The PACF of an invertible MA(q)

$$X_t = -\sum_{j=1}^{\infty} \pi_j X_{t-j} + w_t$$ - because of the infinite AR representation, the PACF will never cut off, as for AR(p) case.

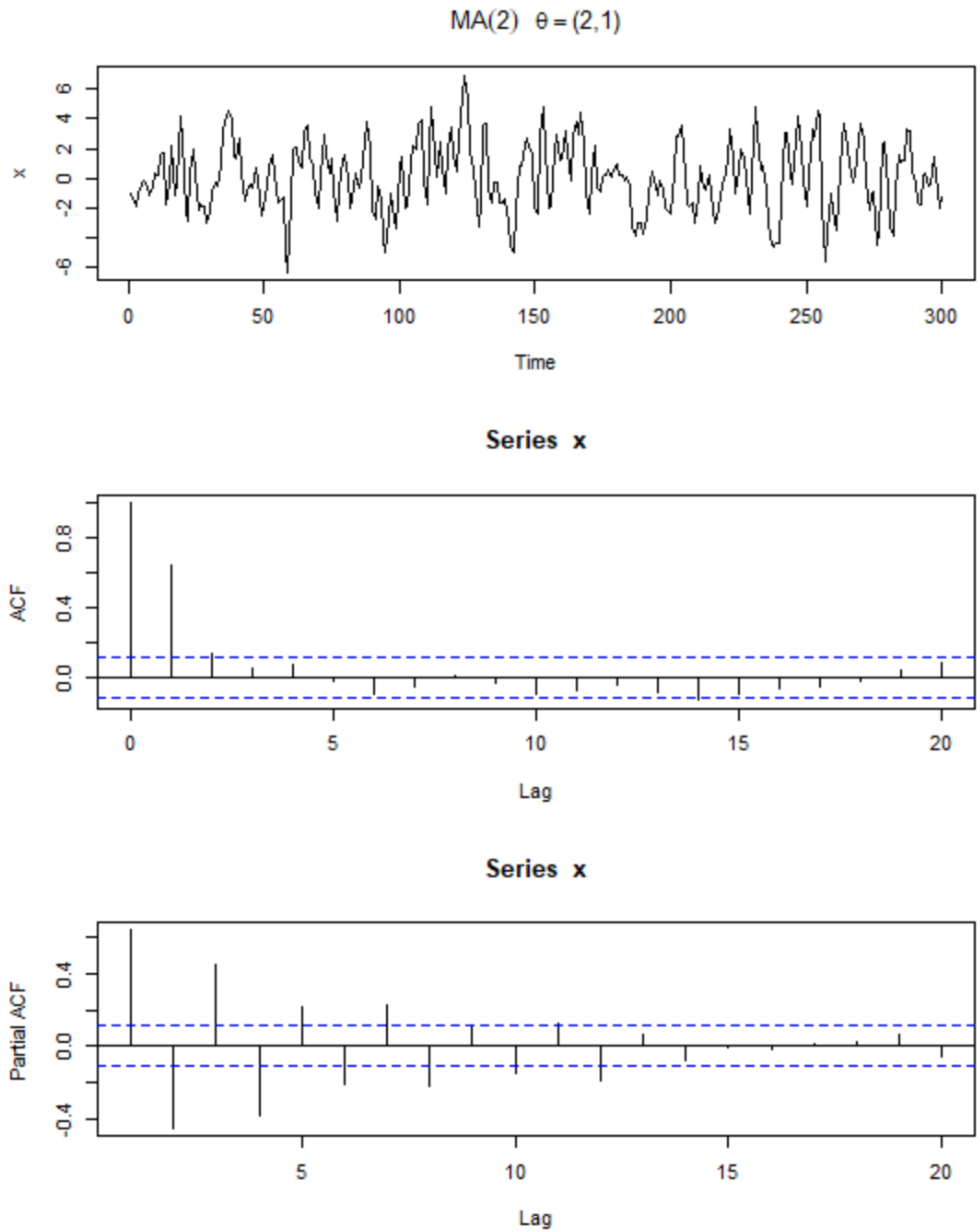
PACF

The PACF of an invertible MA(q)

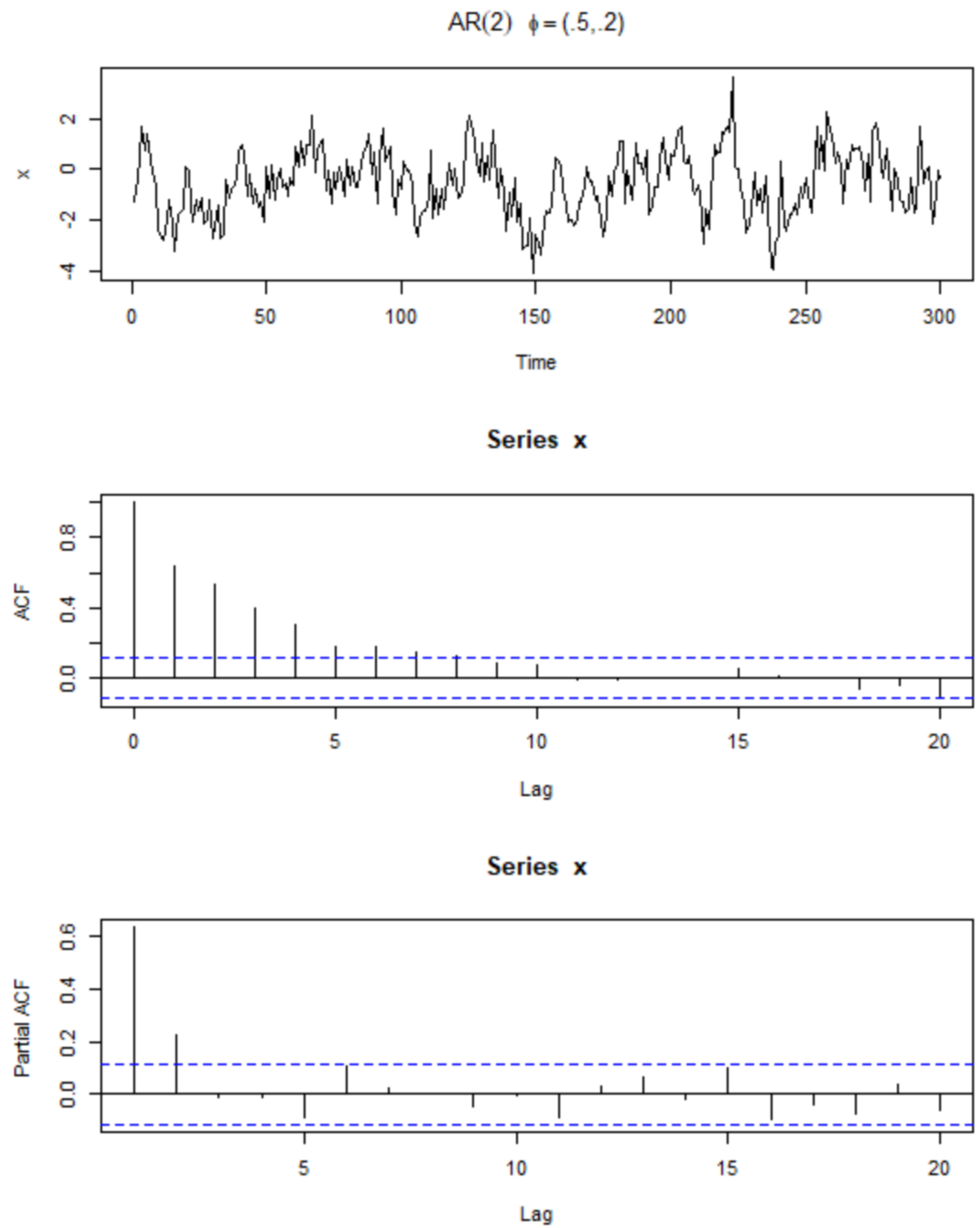
$$X_t = -\sum_{j=1}^{\infty} \pi_j X_{t-j} + w_t$$ - because of the infinite AR representation, the PACF will never cut off, as for AR(p) case.

	AR(p)	MA(q)	ARMA(p,q)
ACF	decays	cuts off after lag q	decays
PACF	cuts off after lag p	decays	decays

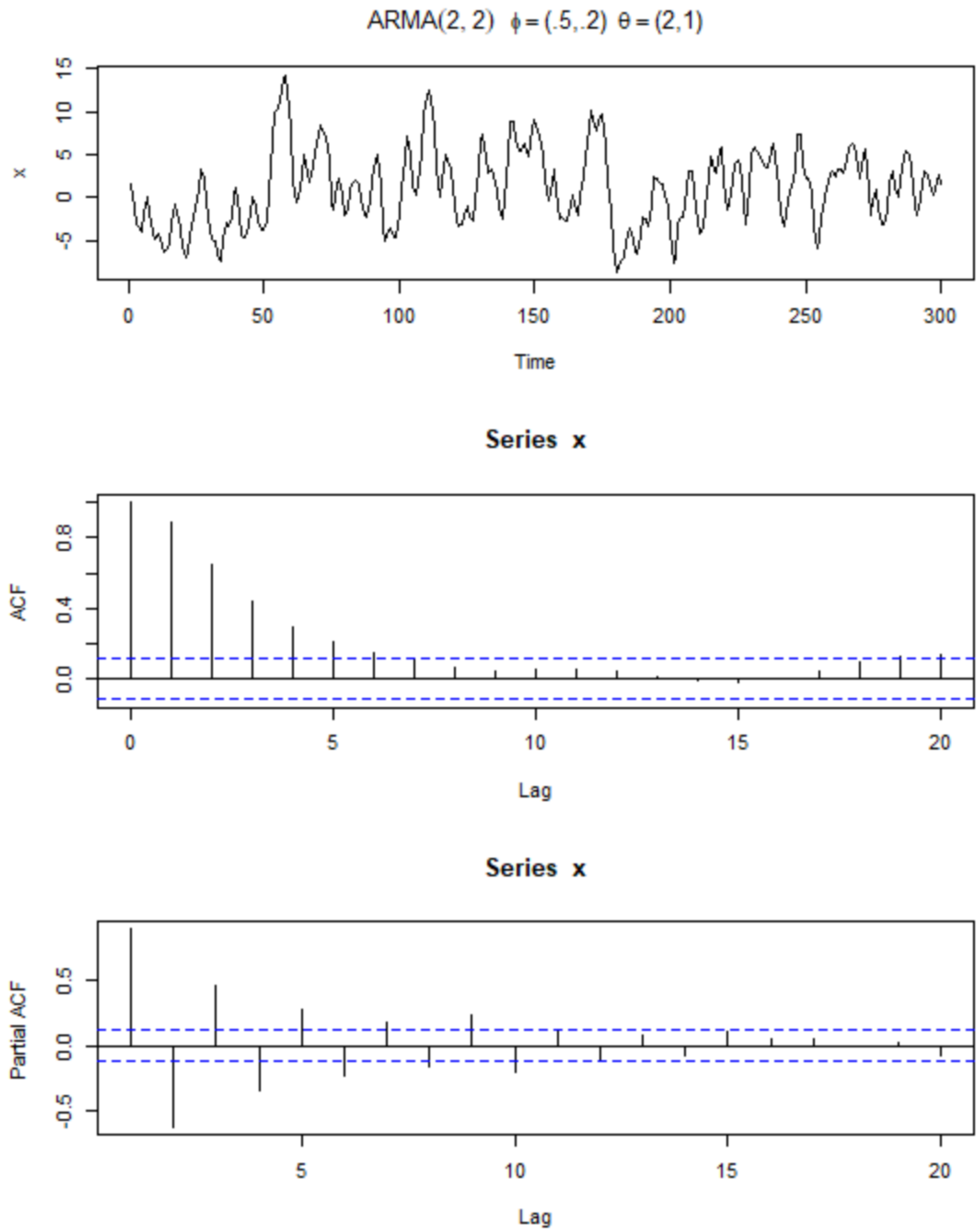
PACF



PACF



PACF


ARMA(2, 2) ϕ = (.5,.2) θ = (2,1)
x
Time
Series x
ACF
Lag
Series x
Partial ACF
Lag

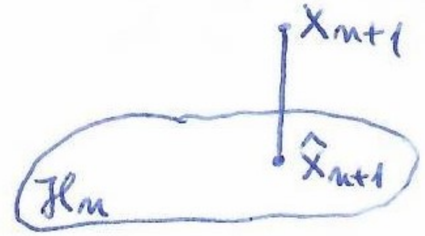
Forecasting[1],[2]

It's difficult to make predictions, especially about the future.

Danish proverb? Yogi Berra?

Let $\mathcal{H}_n = \overline{sp}\{X_1, \dots, X_n\}$

$\hat{X}_{n+1} = P_{\mathcal{H}_n} X_{n+1}$ – the Best Linear Predictor (BLP) of X_{n+1}



$$\hat{X}_{n+1} = \phi_{n1} X_n + \dots + \phi_{nn} X_1 \quad , \quad n \geq 1$$

where $E\left[\left(X_{n+1} - \sum_{j=1}^{n} \phi_{nj} X_{n+1-j}\right) X_{n+1-k}\right] = 0 \quad k = 1, \dots, n$

[1] Robert Shumway, David Stoffer. Time Series Analysis and its Applications with R Examples, 4th edition, Springer, 2017 (chapter 3)
[2] Peter Brockwell, Richard Davis. Time Series: Theory and Methods, Springer-Verlag, 1987 (chapter 5)

Forecasting

It can be rewritten as

$$\sum_{j=1}^{n} \phi_{nj}\,\gamma(k-j) = \gamma(k) \qquad k = 1, \dots, n$$

or, in matrix form

$$\Gamma_n \phi_n = \gamma_n$$

where

$$\Gamma_n = \begin{pmatrix} \gamma(0) & \gamma(1) & \cdots & \gamma(n-1) \\ \vdots & \ddots & & \vdots \\ \vdots & & \ddots & \vdots \\ \gamma(n-1) & \cdots & \cdots & \gamma(0) \end{pmatrix} \qquad \phi_n = \begin{pmatrix} \phi_{n1} \\ \vdots \\ \vdots \\ \phi_{nn} \end{pmatrix} \qquad \gamma_n = \begin{pmatrix} \gamma(1) \\ \vdots \\ \vdots \\ \gamma(n) \end{pmatrix}$$

Forecasting

<u>Proposition</u> If $\sigma_w^2 > 0$ and $\gamma(h) \longrightarrow 0$ as $h \to \infty$, then Γ_n is non-singular for every n.

$$\phi_n = \Gamma_n^{-1} \gamma_n \quad \text{the elements of } \phi_n \text{ are unique}$$

The predictor $\hat{X}_{n+1} = \sum_{i=1}^{n} \phi_{ni} X_{n+1-i} \qquad n = 1, 2, \ldots$

Forecasting

The mean squared error is

$$v_n = E\left[\left(X_{n+1} - \hat{X}_{n+1}\right)^2\right] = E\left[\left(X_{n+1} - \phi_n' X\right)^2\right]$$

where $X = (X_n, X_{n-1}, \dots, X_1)'$

$$\begin{aligned} v_n &= E\left[\left(X_{n+1} - \gamma_n' \Gamma_n^{-1} X\right)^2\right] \\ &= E\left(X_{n+1}^2 - 2\gamma_n' \Gamma_n^{-1} X X_{n+1} + \gamma_n' \Gamma_n^{-1} X X' \Gamma_n^{-1} \gamma_n\right) \\ &= \gamma(0) - 2\gamma_n' \Gamma_n^{-1} \gamma_n + \gamma_n' \Gamma_n^{-1} \Gamma_n \Gamma_n^{-1} \gamma_n \\ &= \gamma(0) - \gamma_n' \Gamma_n^{-1} \gamma_n \end{aligned}$$

ϕ_n and v_n can be computed iteratively, due to Levinson & Durbin.

Forecasting

The Durbin-Levinson algorithm

$$\phi_{11} = \frac{\gamma(1)}{\gamma(0)} \qquad v_0 = \gamma(0)$$

$$\phi_{nn} = \frac{\gamma(n) - \sum_{k=1}^{n-1} \phi_{n-1,k}\, \gamma(n-k)}{v_{n-1}} \quad , \quad n \geq 1$$

$$\phi_{nk} = \phi_{n-1,k} - \phi_{nn} \cdot \phi_{n-1,n-k} \qquad k = 1, \dots, n-1 \qquad n \geq 2$$

$$v_n = v_{n-1} \cdot \left[1 - \phi_{nn}^2\right]$$

Order of computation:

$\phi_{11}, v_1, \phi_{22}, \phi_{21}, v_2, \phi_{33}, \phi_{31}, \phi_{32}, v_3 \; \dots$

the predictor $\hat{X}_{n+1} = \phi_{n1} X_n + \phi_{n2} X_{n-1} + \dots + \phi_{nn} X_1$

Forecasting

<u>Proposition</u> The PACF of a stationary process $\{X_t\}$ can be obtained from the Durbin-Levinson algorithm, as ϕ_{nn} for $n=1,2\ldots$

Another useful algorithm for calculating the forecast $\hat{X}_{n+1}$ is based on the innovations $X_i - \hat{X}_i$, $i=\overline{1,n}$.

Forecasting

The innovations algorithm

$$\hat{X}_{n+1} = \sum_{j=1}^{n} \theta_{nj} \left(X_{n+1-j} - \hat{X}_{n+1-j} \right) \qquad n \geq 1$$

V_n - mean squared errors

$$V_0 = \gamma(0)$$

$$\hat{X}_1 = 0$$

$$\theta_{n,n-k} = \frac{\gamma(n-k) - \sum_{j=0}^{k-1} \theta_{k,k-j} \cdot \theta_{n,n-j} \cdot V_j}{V_k}, \quad k = 0, 1, \ldots, n-1$$

$$V_n = \gamma(0) - \sum_{j=0}^{n-1} \theta_{n,n-j}^2 \cdot V_j$$

Order of computation:

$V_0, \hat{X}_1;\ \theta_{11}, V_1, \hat{X}_2;\ \theta_{22}, \theta_{21}, V_2, \hat{X}_3;\ \theta_{33}, \theta_{32}, \theta_{31}, V_3, \hat{X}_4 \ldots$

Forecasting

Obs. The Durbin-Levinson algorithm and the innovations algorithm can be adapted to compute the h-step predictors

$$P_{\mathcal{H}_n} X_{n+h} = \phi_{n1}^{(h)} X_n + \ldots + \phi_{nn}^{(h)} X_1, \quad n, h \geq 1.$$

For details, see Brockwell & Davis, pages 161, 167.

Forecasting

1) Apply the Durbin-Levinson algorithm for an AR(2) process for the first three iterations $n=1,2,3$.
Indicate the first three values of the PACF.

Hint: for an AR(2), recall that

$$\rho(h) - \phi_1 \rho(h-1) - \phi_2 \rho(h-2) = 0 \quad \forall\, h \geq 1$$

and use it for $h=1,2,3$ in the D-L algorithm

Forecasting

2) Apply the innovations algorithm for an MA(1) process

$$X_t = \theta W_{t-1} + W_t$$

$$\gamma(0) = (1+\theta^2)\,\sigma_w^2$$

$$\gamma(1) = \theta\,\sigma_w^2$$

$$\gamma(h) = 0 \quad \text{for } h>1$$

$\Rightarrow \gamma(n-k) \neq 0$ only for $k=n-1$ in the formula for $\theta_{n,n-k}$

Forecasting

Thus, $\theta_{1,1} = \gamma(1)/v_0$

$\theta_{2,2} = 0, \quad \theta_{2,1} = \gamma(1)/v_1$

$\theta_{3,3} = 0, \quad \theta_{3,2} = 0, \quad \theta_{3,1} = \gamma(1)/v_2$

$- - - \quad - \quad -$

$\theta_{n,n} = \theta_{n,n-1} = \ldots = \theta_{n,2} = 0, \quad \theta_{n,1} = \gamma(1)/v_{n-1}$

$v_0 = (1 + \theta^2)\,\sigma_w^2$

$- - - \quad -$

$v_n = [1 + \theta^2 - \theta^2\sigma^2/v_{n-1}]\,\sigma_w^2$

$\hat{x}_{n+1} = \theta(x_n - \hat{x}_n) \cdot \sigma_w^2/v_{n-1}$

Forecasting

The D-L algorithm is convenient for AR(p) models.

The innovations algorithm is convenient for MA(q) models.