

Introduction to Time Series Analysis

Dr David A. Stephens

Department of Mathematics & Statistics
McGill University
Room 1225, Burnside Hall

`david.stephens@mcgill.ca`
`www.math.mcgill.ca/dstephens/`

Chapter 3

ARMA Models

3. ARMA Models

We study further properties of $\text{ARMA}(p, q)$ processes.

Consider the $\text{ARMA}(p, q)$ process $\{X_t\}$,

$$\Phi(B)X_t = \Theta(B)Z_t$$

where $\{Z_t\} \sim \text{WN}(0, \sigma_Z^2)$,

3.1 Basic Properties

Here

$$\Phi(z) = 1 - \phi_1 z - \cdots - \phi_p z^p = \prod_{j=1}^p (1 - \xi_j z)$$

$$\Theta(z) = 1 + \theta_1 z + \cdots + \theta_q z^q = \prod_{j=1}^q (1 - \omega_j z)$$

We assume that there are no common factors in $\Phi(z)$ in $\Theta(z)$.

- $\{X_t\}$ is **causal** and **stationary** if and only if $|\xi_j| < 1, j = 1, \dots, p$
- $\{X_t\}$ is **invertible** if and only if $|\omega_j| < 1, j = 1, \dots, q$.

3.1 Basic Properties

Note that by using the power series expansion of the generating function

$$\Psi(z) = \frac{\Theta(z)}{\Phi(z)}$$

we can construct the linear process ($\text{MA}(\infty)$) representation

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

3.1 Basic Properties

We must have

$$\begin{aligned}(1 - \phi_1 z - \cdots - \phi_p z^p)(\psi_0 + \psi_1 z + \cdots) \\ = (1 + \theta_1 z + \cdots + \theta_q z^q)\end{aligned}$$

so equating coefficients, we have $\psi_0 = 1$, and for $1, 2, \dots$, we have the recursion

$$\psi_j = \theta_j + \sum_{k=1}^p \phi_k \psi_{j-k}$$

3.1 Basic Properties

By a similar argument, under invertibility, we can construct an $AR(\infty)$ representation

$$Z_t = \sum_{j=0}^{\infty} \pi_j X_{t-j}$$

where $\pi_0 = 1$, and $j = 0, 1, 2, \dots$

$$\pi_j = -\phi_j - \sum_{k=1}^q \theta_k \pi_{j-k}$$

3.1 Basic Properties

ACF: The ACF $\gamma_X(h)$ is given by previous results for linear processes; for integer h

$$\gamma_X(h) = \mathbb{E}[X_t X_{t+h}] = \sigma_Z^2 \sum_{j=0}^{\infty} \psi_j \psi_{j+|h|}$$

Example: MA(q)

By the linear process representation, $\psi_0 = 1$, $\psi_j = \theta_j$ for $j = 1, \dots, q$, and $\psi_j = 0$ for $j > q$, so therefore

$$\gamma_X(h) = \sigma_Z^2 \sum_{j=0}^{q-h} \theta_j \theta_{j+h} \quad 0 \leq h \leq q$$

and zero otherwise.

3.1 Basic Properties

Example: ARMA(1, 1)

Consider the ARMA(1, 1) process based on a $WN(0, \sigma_Z^2)$ process, with $|\phi| < 1$. Then recall that

$$\psi_0 = 1$$

$$\psi_j = (\theta + \phi)\phi^{j-1}, j \geq 1.$$

gives the MA(∞) representation. Note that for $j \geq 2$,

$$\psi_j = \phi\psi_{j-1}.$$

3.1 Basic Properties

Example: ARMA(1, 1) stationary and causal

$h = 0$:

$$\begin{aligned}\gamma_X(0) &= \sigma_Z^2 \sum_{j=0}^{\infty} \psi_j^2 \\ &= \sigma_Z^2 \left[\psi_0^2 + \sum_{j=1}^{\infty} \psi_j^2 \right] \\ &= \sigma_Z^2 \left[1 + (\theta + \phi)^2 \sum_{j=0}^{\infty} \phi^{2j} \right] \\ &= \sigma_Z^2 \left[1 + \frac{(\theta + \phi)^2}{1 - \phi^2} \right]\end{aligned}$$

3.1 Basic Properties

Example: ARMA(1, 1) stationary and causal

$h = 1$:

$$\begin{aligned}\gamma_X(1) &= \sigma_Z^2 \sum_{j=0}^{\infty} \psi_j \psi_{j+1} \\&= \sigma_Z^2 \left[(\theta + \phi) + (\theta + \phi)^2 \sum_{j=1}^{\infty} \phi^{2j-1} \right] \\&= \sigma_Z^2 \left[(\theta + \phi) + (\theta + \phi)^2 \phi \sum_{j=0}^{\infty} \phi^{2j} \right] \\&= \sigma_Z^2 \left[(\theta + \phi) + \frac{\phi(\theta + \phi)^2}{1 - \phi^2} \right]\end{aligned}$$

3.1 Basic Properties

Example: ARMA(1, 1) stationary and causal

For $h \geq 2$

$$\begin{aligned}\gamma_X(h) &= \sigma_Z^2 \sum_{j=0}^{\infty} \psi_j \psi_{j+h} \\ &= \sigma_Z^2 \phi \sum_{j=0}^{\infty} \psi_j \psi_{j+h-1} & \psi_{j+h} &= \phi \psi_{j+h-1} \\ &= \phi \gamma_X(h-1)\end{aligned}$$

3.1 Basic Properties

For a direct computation for the general ARMA case, note that

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

Multiplying through by X_{t-h} , $h = 0, 1, 2, \dots$:

$$\text{LHS} \quad X_t X_{t-h} - \phi_1 X_{t-1} X_{t-h} - \cdots - \phi_p X_{t-p} X_{t-h}$$

$$\text{RHS} \quad Z_t X_{t-h} + \theta_1 Z_{t-1} X_{t-h} + \cdots + \theta_q Z_{t-q} X_{t-h}.$$

3.1 Basic Properties

Taking expectations we have for $j = 0, 1, \dots, p$

$$\mathbb{E}[X_{t-j}X_{t-h}] = \mathbb{E}[X_{t+h-j}X_t] = \gamma_X(h-j)$$

so for the left-hand side we have

$$\gamma_X(h) - \phi_1\gamma_X(h-1) - \dots - \phi_p\gamma_X(h-p).$$

3.1 Basic Properties

For the right hand side, using the $\text{MA}(\infty)$ form

$$X_{t-h} = \sum_{l=0}^{\infty} \psi_l Z_{t-h-l}.$$

Thus for $j = 0, 1, \dots, q$,

$$\mathbb{E}[X_{t-h} Z_{t-j}] = \mathbb{E} \left[\sum_{l=0}^{\infty} \psi_l Z_{t-h-l} Z_{t-j} \right] = \sum_{l=0}^{\infty} \psi_l \mathbb{E}[Z_{t-h-l} Z_{t-j}]$$

Now

$$\mathbb{E}[Z_{t-h-l} Z_{t-j}] = \begin{cases} \sigma_Z^2 & l = j - h \\ 0 & \text{otherwise} \end{cases}$$

3.1 Basic Properties

Therefore, for fixed h , as $\psi_j = 0$ if $j < 0$, we have

$$\mathbb{E}[X_{t-h}Z_{t-j}] = \sigma_Z^2\psi_{j-h}$$

for $j = h, \dots, q$, and zero otherwise. Hence, writing $\theta_0 = 1$, we have

$$\mathbb{E}\left[\sum_{j=0}^q \theta_j Z_{t-j} X_{t-h}\right] = \sigma_Z^2 \sum_{j=h}^q \theta_j \psi_{j-h} = \sigma_Z^2 \sum_{j=0}^{q-h} \theta_{j+h} \psi_j$$

Note that if $h > q$, $\theta_{j+h} = 0$ for all $j \geq 0$, so in this case the right hand side is zero.

3.1 Basic Properties

If $m = \max\{p, q + 1\}$, we therefore have the infinite system of equations

$$\gamma_X(h) - \sum_{j=1}^p \phi_j \gamma_X(h-j) = \begin{cases} \sigma_Z^2 \sum_{j=0}^m \theta_{j+h} \psi_j & 0 \leq h < m \\ 0 & h \geq m \end{cases}$$

This system in the unknowns

$$\gamma_X(0), \gamma_X(1), \gamma_X(2), \dots$$

for fixed parameters

$$\phi_1, \dots, \phi_p, \theta_1, \dots, \theta_q$$

can be solved using two approaches.

3.1 Basic Properties

- **Direct approach:** the equations form a system known as **linear difference equations** which has an analytic solution, expressed via the roots $\zeta_1, \dots, \zeta_p$ of the AR polynomial equation $\Phi(z) = 0$

$$\gamma_X(h) = \sum_{j=1}^p \alpha_j \zeta_j^{-h} \quad h \geq m - p$$

where the initial conditions determine $\alpha_1, \dots, \alpha_p$ and

$$\gamma_X(0), \dots, \gamma_X(h - p - 1).$$

3.1 Basic Properties

- **Recursive approach:** The first $p+1$ equations form a linear system in the autocovariances $\gamma_X(0), \dots, \gamma_X(p)$, which can be solved using matrix inversion, Levinson-Durbin, or innovations algorithm.

The remaining autocovariances $\gamma_X(p+1), \gamma_X(p+2), \dots$ can be computed using the recursion.

3.1 Basic Properties

In R, the function `ARMAacf` computes the ACF for given parameter settings

```
ARMAacf(ar = numeric(0), ma = numeric(0), lag.max = r, pacf = FALSE)
```

The function `ARMAtoMA` computes the $MA(\infty)$ coefficients.

3.2 Autocovariance Generating Function

The **autocovariance generating function** (ACVGF) can also be used for computing the ACV sequences.

For a real sequence $\{a_n\}$, the **generating function**

$$G(z) = \sum_n a_n z^n \quad z \in \mathbb{C}$$

is defined for $|z|$ in a neighbourhood of 1, that is

$$\frac{1}{r} < |z| < r$$

for $r > 1$.

3.2 Autocovariance Generating Function

Suppose that $\{Z_t\} \sim \text{WN}(0, \sigma_Z^2)$, and

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

Suppose that there exists an $r > 1$ such that

$$\sum_{j=-\infty}^{\infty} |\psi_j| z^j \quad \frac{1}{r} < |z| < r$$

is convergent. Then define the **ACVGF** for $\{X_t\}$ by

$$G_X(z) = \sum_{h=-\infty}^{\infty} \gamma_X(h) z^h \quad \frac{1}{r} < |z| < r$$

3.2 Autocovariance Generating Function

As polynomials are defined by their coefficients, there is a 1-1 correspondence between $\gamma_X(h)$ and $G_X(z)$.

- This means that $\gamma_X(h)$ is **uniquely identified** by $G_X(z)$

3.2 Autocovariance Generating Function

Recall that

$$\gamma_X(h) = \sigma_Z^2 \sum_{j=-\infty}^{\infty} \psi_j \psi_{j+|h|}$$

Therefore

$$\begin{aligned} G_X(z) &= \sigma_Z^2 \sum_{h=-\infty}^{\infty} \sum_{j=-\infty}^{\infty} \psi_j \psi_{j+|h|} z^h \\ &= \sigma_Z^2 \left[\sum_{j=-\infty}^{\infty} \psi_j^2 + \sum_{h=1}^{\infty} \sum_{j=-\infty}^{\infty} \psi_j \psi_{j+h} (z^h + z^{-h}) \right] \\ &= \sigma_Z^2 \left[\sum_{j=-\infty}^{\infty} \psi_j z^j \right] \left[\sum_{h=-\infty}^{\infty} \psi_h z^{-h} \right] \\ &= \sigma_Z^2 \Psi(z) \Psi(z^{-1}), \text{ say.} \end{aligned}$$

3.2 Autocovariance Generating Function

For the stationary ARMA(p, q) model

$$\Phi(B)X_t = \Theta(B)Z_t$$

where $\Phi(z) \neq 0$ for $|z| = 1$, we have

$$X_t = \{\Phi(B)\}^{-1} \Theta(B)Z_t = \Psi(B)Z_t$$

say, so

$$G_X(z) = \sigma_Z^2 \Psi(z) \Psi(z^{-1}) = \sigma_Z^2 \frac{\Theta(z) \Theta(z^{-1})}{\Phi(z) \Phi(z^{-1})}$$

which is well defined and finite as the denominator is non zero.

3.2 Autocovariance Generating Function

Example: $\text{WN}(0, \sigma_Z^2)$

Note that if $\{Z_t\} \sim \text{WN}(0, \sigma_Z^2)$,

$$G_Z(z) = \sigma_Z^2$$

and that this is a unique characteristic of white noise processes.

In general, the ACVF and ACVGF are not unique: in general, different ARMA processes can have the same ACVF and ACVGF.

3.2 Autocovariance Generating Function

We have seen that the $\text{ARMA}(p, q)$ process

$$\Phi(B)X_t = \Theta(B)Z_t$$

is

- **causal** if the AR reciprocal roots satisfy

$$|\xi_j| < 1 \quad j = 1, 2, \dots, p$$

- **invertible** if the MA reciprocal roots satisfy

$$|\omega_j| < 1 \quad j = 1, 2, \dots, q$$

3.2 Autocovariance Generating Function

If the process is non-causal or non-invertible, we can derive a causal and invertible process with the **same** autocovariance sequence.

The new process is defined with respect to a different white noise process.

3.2 Autocovariance Generating Function

Example: MA(1)

If $\{Z_t\} \sim \text{WN}(0, \sigma_Z^2)$, and

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

so that

$$\Theta(z) = 1 - 2z,$$

as $|2| > 1$, the process is non-invertible. Define

$$Z_t^* = (1 - B/2)^{-1}(1 - 2B)Z_t.$$

Then

$$(1 - B/2)Z_t^* = (1 - 2B)Z_t = X_t.$$

3.2 Autocovariance Generating Function

Example: MA(1)

That is

$$\Phi(B)Z_t^* = \Theta(B)Z_t$$

where

$$\Phi(z) = 1 - \frac{1}{2}z \quad \Theta(z) = 1 - 2z.$$

Thus

$$Z_t^* = \Psi(B)Z_t$$

where

$$\Psi(z) = \frac{\Theta(z)}{\Phi(z)} = \frac{(1 - 2z)}{(1 - z/2)}$$

3.2 Autocovariance Generating Function

Example: MA(1)

We have

$$Z_t^* = (1 - B/2)^{-1} X_t = \sum_{j=0}^{\infty} \left(\frac{1}{2}\right)^j X_{t-j}$$

and the ACVGF for $\{Z_t^*\}$ is

$$\begin{aligned} G_{Z^*}(z) &= \sigma_Z^2 \Psi(z) \Psi(z^{-1}) \\ &= \sigma_Z^2 \frac{(1 - 2z)(1 - 2z^{-1})}{(1 - z/2)(1 - 1/(2z))} \\ &= 4\sigma_Z^2 \frac{(1 - 2z)(z - 2)}{(2 - z)(2z - 1)} \\ &= 4\sigma_Z^2 \end{aligned}$$

3.2 Autocovariance Generating Function

Example: MA(1)

Therefore $\{Z_t^*\} \sim \text{WN}(0, 4\sigma_Z^2)$, and

$$X_t = \Phi(B)Z_t^* = Z_t^* - \frac{1}{2}Z_{t-1}^*$$

has an **invertible** representation in terms of $\{Z_t^*\}$.

3.2 Autocovariance Generating Function

Consider the arbitrary stationary ARMA(p, q) process

$$\Phi(B)X_t = \Theta(B)Z_t$$

where $\Phi(z) \neq 0$, $\Theta \neq 0$ when $|z| = 1$

$$\Phi(z) = \prod_{j=1}^p (1 - \xi_j z)$$

$$\Theta(z) = \prod_{j=1}^q (1 - \omega_j z)$$

and

$$G_X(z) = \sigma_Z^2 \Psi(z) \Psi(z^{-1}) = \sigma_Z^2 \frac{\Theta(z) \Theta(z^{-1})}{\Phi(z) \Phi(z^{-1})}.$$

3.2 Autocovariance Generating Function

Then there exist polynomials

$$\Phi^*(z) \quad \Theta^*(z)$$

and a white noise process $\{Z_t^*\}$ such that

$$\Phi^*(B)X_t = \Theta^*(B)Z_t^*$$

is a causal and invertible representation.

3.2 Autocovariance Generating Function

Suppose that the reciprocal roots are ordered such that

$$\xi_{r+1}, \dots, \xi_p \quad \omega_{s+1}, \dots, \omega_q$$

lie outside of the unit circle. Then let

$$\Phi^*(z) = \Phi(z) \prod_{j=r+1}^p \left(\frac{1 - \xi_j z}{1 - \xi_j^{-1} z} \right)^{-1}$$

$$\Theta^*(z) = \Theta(z) \prod_{j=s+1}^q \left(\frac{1 - \omega_j z}{1 - \omega_j^{-1} z} \right)^{-1}$$

and consider

$$Z_t^* = \{\Theta^*(B)\}^{-1} \Phi^*(B) X_t$$

3.2 Autocovariance Generating Function

After some algebra, it follows that

$$G_{Z^*}(z) = \sigma_Z^2 \left\{ \prod_{j=r+1}^p |\xi_j|^{-2} \right\} \left\{ \prod_{j=s+1}^q |\omega_j|^2 \right\}$$

and hence that $\{Z_t^*\}$ is a white noise process and

$$\Phi^*(B)X_t = \Theta^*(B)Z_t^*$$

By construction, this is a causal and invertible representation of $\{X_t\}$.

3.3 Partial Autocorrelation

The PACF, $\alpha_X(h)$, defined by

$$\alpha_X(0) = 1$$

$$\alpha_X(h) = \phi_{h,h}, \quad h \geq 1.$$

$\phi_{h,h}$ is the last component of

$$\Phi_h = \Gamma_h^{-1} \gamma_h$$

We can carry out testing for the estimated PACF sequence as for ACF sequence.

3.4 Forecasting

The innovations algorithm can be used for the ARMA(p, q) model. If $m = \max\{p, q\}$, the one-step ahead forecasts are

$$\hat{X}_{n+1} = \begin{cases} \sum_{j=1}^n \theta_{n,j} (X_{n+1-j} - \hat{X}_{n+1-j}) & 1 \leq n \leq m \\ \sum_{j=1}^p \phi_j X_{n+1-j} + \sum_{j=1}^q \theta_{n,j} (X_{n+1-j} - \hat{X}_{n+1-j}) & n > m \end{cases}$$

where the $\theta_{n,j}$ are the quantities computed via the innovations algorithm.

3.5 Estimation

Estimation of the parameters of ARMA models proceeds using one of a number of different approaches

- Yule-Walker approach
- Burg's algorithm
- Maximum likelihood

Previous approaches use moment-based approaches in a non-parametric fashion.

3.5 Estimation

Yule-Walker for AR(p) models: If

$$\Phi(B)X_t = Z_t$$

then using the previous technique, we premultiply by X_{t-h} for $h = 0, \dots, p$ and take expectations, to yield

$$\gamma_X(h) = \phi_1 \gamma_X(h-1) + \dots + \phi_p \gamma_X(h-p) \quad h = 0, \dots, (p-1)$$

or equivalently, if $\phi_p = (\phi_1, \dots, \phi_p)^\top$,

$$\Gamma_p \phi_p = \gamma_p$$

and

$$\sigma_Z^2 = \gamma_X(0) - \phi_p^\top \gamma_p$$

3.5 Estimation

Replacing the theoretical quantities by sample quantities, the equations yield estimates of ϕ_p and σ_Z^2 ; using the sample correlations, we have

$$\hat{\phi}_p = \hat{\mathbf{R}}_p^{-1} \hat{\rho}_p$$

As $n \longrightarrow \infty$,

$$\sqrt{n}(\hat{\phi}_p - \phi_p) \sim \mathcal{AN}(0, \sigma_Z^2 \mathbf{\Gamma}_p)$$

The estimated parameters yield a stationary and causal procedure.

Note that to construct a stationary and causal $\text{AR}(p)$ process we may use the relationship to define ϕ_p from $\gamma_X(h)$.

3.5 Estimation

Note: In the $\text{AR}(p)$ case, if the models

$$\text{AR}(1), \text{AR}(2), \dots, \text{AR}(p), \dots$$

are fitted, then the sequence of estimates

- $\hat{\phi}_{1,1}$ for $\text{AR}(1)$,
- $\hat{\phi}_{2,1}, \hat{\phi}_{2,2}$ for $\text{AR}(2)$,
- ...
- $\hat{\phi}_{p,1}, \dots, \hat{\phi}_{p,p}$ for $\text{AR}(p)$

coincide with the Levinson-Durbin coefficients.

The constants $\hat{\phi}_{h,h}$, $h = 0, 1, 2, \dots$ yield the PACF $\alpha_X(h)$.

3.5 Estimation

If $\{Z_t\} \sim \text{WN}(0, \sigma_Z^2)$, and $\{X_t\}$ is an $\text{ARMA}(p, q)$ model

$$\Phi(B)X_t = \Theta(B)Z_t$$

with a causal linear process representation

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

the equivalent calculation yields, for $0 \leq h \leq p + q$,

$$\gamma_X(h) - \phi_1 \gamma_X(h-1) - \cdots - \phi_p \gamma_X(h-p) = \sigma_Z^2 \sum_{j=h}^q \theta_j \psi_{j-h}.$$

These equations are **non-linear** in the parameters.

3.5 Estimation

Burg's Algorithm for AR(p) models: This algorithm proceeds via direct computation of the PACF series.

For $0 \leq i < n$, define

$$U_i(t) = X_{n+1+i-t} - \hat{X}_{n+1+i-t}^F$$

$$V_i(t) = X_{n+1+i-t} - \hat{X}_{n+1+i-t}^B$$

where $\hat{X}_{n+1+i-t}^F$ and $\hat{X}_{n+1+i-t}^B$ are the forward and backward forecasts of $X_{n+1+i-t}$ based on the preceding/following i observations.

3.5 Estimation

Then $U_0(t) = V_0(t) = X_{n+1-t}$, and

$$U_i(t) = U_{i-1}(t-1) - \phi_{i,i} V_{i-1}(t)$$

$$V_i(t) = V_{i-1}(t) - \phi_{i,i} U_{i-1}(t-1).$$

We then choose $\phi_{i,i}$ to minimize

$$\frac{1}{2(n-i)} \sum_{t=i+1}^n (u_i(t)^2 + v_i(t)^2)$$

that is, the sums of forward and backward prediction errors.

This minimization can be achieved analytically, with the computation performed recursively.

3.5 Estimation

Maximum Likelihood Estimation: In most cases, maximum likelihood (ML) estimation provides the most efficient approach, and usually produces estimators have good small sample properties.

However, outside of the Normal case, the likelihood function is not often tractable.

3.5 Estimation

We consider first Gaussian time series $\{X_t\}$, presumed to be zero mean, with autocovariance

$$\mathbb{E}[X_i X_j] = \kappa(i, j)$$

Let

$$\hat{X}_{t+1} = \mathbb{E}[X_{t+1} | X_1, \dots, X_t] = \hat{P}_t X_{t+1} \quad t = 1, 2, \dots$$

which is always available for the Gaussian case.

3.5 Estimation

If $X_{1:t} = (X_1, \dots, X_t)^\top$, then

$$\hat{X}_{t+1} = \gamma_{t+1,t:1}^\top \Gamma_t^{-1} X_{1:t}$$

where

$$\gamma_{t+1,t:1} = \mathbb{E}[X_{t+1} X_{t:1}] \quad (t \times 1)$$

$$\Gamma_t = \mathbb{E}[X_{1:t} X_{1:t}^\top] \quad (t \times t)$$

are derived from the properties of the general multivariate normal distribution.

3.5 Estimation

Let

$$\mathbf{\Gamma}_n = \mathbb{E}[\mathbf{X}_{1:n}\mathbf{X}_{1:n}^\top]$$

then the likelihood for the parameters of $\mathbf{\Gamma}_n$ is

$$\mathcal{L}_n(\mathbf{\Gamma}_n) = \frac{1}{(2\pi)^{n/2}} |\mathbf{\Gamma}_n|^{-1/2} \exp \left\{ -\frac{1}{2} \mathbf{x}_{1:n}^\top \mathbf{\Gamma}_n^{-1} \mathbf{x}_{1:n} \right\}.$$

If n is large, computation of

$$|\mathbf{\Gamma}_n| = \det \mathbf{\Gamma}_n \qquad \mathbf{\Gamma}_n^{-1}$$

may not be straightforward numerically. However, the innovations algorithm can be used to compute the determinant and inverse.

3.5 Estimation

The innovations algorithm returns the quantities

$$\theta_{i,j}, \quad i = 1, 2, \dots, n-1, \quad j = 1, \dots, i.$$

If

$$\mathbf{C}_n = \begin{bmatrix} 1 & 0 & 0 & 0 & \cdots & 0 \\ \theta_{1,1} & 1 & 0 & 0 & \cdots & 0 \\ \theta_{2,2} & \theta_{2,1} & 1 & 0 & \cdots & 0 \\ \theta_{3,3} & \theta_{3,2} & \theta_{3,1} & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ \theta_{n-1,n-1} & \theta_{n-1,n-2} & \theta_{n-1,n-3} & \theta_{n-1,n-4} & \cdots & 1 \end{bmatrix}$$

Then, as before $X_{1:n} = \mathbf{C}_n(X_{1:n} - \hat{X}_{1:n})$, with the elements of the $n \times 1$ vector $(X_{1:n} - \hat{X}_{1:n})$ **uncorrelated**, with covariance matrix $\mathbf{D}_n = \text{diag}(v_0, v_1, \dots, v_{n-1})$.

3.5 Estimation

Therefore

$$\mathbf{\Gamma}_n = \mathbf{C}_n \mathbf{D}_n \mathbf{C}_n^\top$$

so that

$$\begin{aligned} \mathbf{X}_{1:n}^\top \mathbf{\Gamma}_n^{-1} \mathbf{X}_{1:n} &= (\mathbf{X}_{1:n} - \hat{\mathbf{X}}_{1:n})^\top \mathbf{D}_n^{-1} (\mathbf{X}_{1:n} - \hat{\mathbf{X}}_{1:n}) \\ &= \sum_{j=1}^n (X_j - \hat{X}_j)^2 / v_{j-1} \end{aligned}$$

and

$$|\mathbf{\Gamma}_n| = \prod_{j=0}^{n-1} v_j.$$

and thus the innovations algorithm facilitates the Gaussian likelihood computation for any n .

3.5 Estimation

For $\text{ARMA}(p, q)$, we saw that

$$\hat{X}_{n+1} = \begin{cases} \sum_{j=1}^n \theta_{n,j} (X_{n+1-j} - \hat{X}_{n+1-j}) & 1 \leq n \leq m \\ \sum_{j=1}^p \phi_j X_{n+1-j} + \sum_{j=1}^q \theta_{n,j} (X_{n+1-j} - \hat{X}_{n+1-j}) & n > m \end{cases}$$

where $m = \max(p, q)$, and

$$v_n = \mathbb{E}[(X_{n+1} - \hat{X}_{n+1})^2] = \sigma_Z^2 r_n$$

say.

3.5 Estimation

ML estimation requires that the likelihood is maximized, so the innovations algorithm is run for each function evaluation in the iterative optimization algorithm (Newton-Raphson, Quasi-Newton etc.).

The properties of the ML estimators are well-established.

3.5 Estimation

We have that

$$\mathcal{L}(\theta, \phi, \sigma_Z^2) = \left(\frac{1}{2\pi}\right)^{n/2} \left(\frac{1}{\sigma_Z^2}\right)^{n/2} \{R(\phi, \theta)\}^{-1/2} \exp\left\{-\frac{1}{2\sigma_Z^2} S(\phi, \theta)\right\}$$

where

$$R(\phi, \theta) = \prod_{j=1}^n r_{j-1} \qquad S(\phi, \theta) = \sum_{j=1}^n \frac{(X_j - \hat{X}_j(\phi, \theta))^2}{r_{j-1}}$$

Note that estimation of σ_Z^2 and (ϕ, θ) can be separated.

3.5 Estimation

Notes:

1. There is no guarantee that the resulting estimates will lead to a causal or invertible process. These constraints can be imposed during the optimization, or a reparameterization to the reciprocal roots can be performed.
2. The identical optimization can be carried out for non-Gaussian processes, leading to maximum pseudo likelihood estimates.

3.5 Estimation

Conditional Likelihood for AR(p): A **conditional likelihood** approach conditions on the first p data points to facilitate ease of likelihood construction. We have that

$$\begin{aligned}f(y_{1:n}|\phi, \sigma_Z^2) &= f(y_{1:p}|\phi, \sigma_Z^2)f(y_{(p+1):n}|y_{1:p}, \phi, \sigma_Z^2) \\&= \mathcal{L}_1(\phi, \sigma_Z^2)\mathcal{L}_2(\phi, \sigma_Z^2)\end{aligned}$$

say, where

$$f(y_{(p+1):n}|\phi, \sigma_Z^2) = \prod_{j=p+1}^n f(y_j|y_{(j-p-1):(j-1)}, \phi, \sigma_Z^2).$$

The conditional likelihood approach uses only $\mathcal{L}_2(\phi, \sigma_Z^2)$ as the objective function.

3.5 Estimation

Least-squares: A **least-squares** procedure based on $S(\phi, \theta)$, that is

$$(\hat{\phi}, \hat{\theta}) = \arg \min S(\phi, \theta)$$

with

$$\hat{\sigma}_Z^2 = \frac{S(\hat{\phi}, \hat{\theta})}{n - p - q}$$

can be used.

A **conditional** least-squares approach can also be used for $\text{AR}(p)$ models.

3.5 Estimation

In R the main function for ARMA estimation is `arima`:
For example

```
arima(x,order=c(1,0,1))
```

fits an ARMA(1, 1) model to the data in `x`.

3.5 Estimation

The library `fArma` contains the `armaFit` function, which provides a number of alternative fitting approaches.

Other functions for AR models include:

`ar`

`ar.burg`

`ar.yw`

`ar.mle`

3.5 Estimation

Example: Lake Huron data (de-trended)

Four potential models:

- AR(1) `arma(Lake.Yt,order=c(1,0,0))`
- AR(2) `arma(Lake.Yt,order=c(2,0,0))`
- MA(1) `arma(Lake.Yt,order=c(0,0,1))`
- ARMA(1,1) `arma(Lake.Yt,order=c(1,0,1))`

3.5 Estimation

Example: Lake Huron data (de-trended)

```
Huron<-scan("https://www.math.mcgill.ca//dstephens//545//Data//huron.txt")
Huron<-Huron+570
n.H<-length(Huron); t.H<-c(1:n.H); x<-c(1875:1972)
Lake.Yt<-residuals(lm(Huron~t.H))
arima(Lake.Yt,order=c(1,0,0))

##
## Call:
## arima(x = Lake.Yt, order = c(1, 0, 0))
##
## Coefficients:
##          ar1  intercept
##          0.7823    0.0794
## s.e.   0.0635    0.3174
##
## sigma^2 estimated as 0.4986:  log likelihood = -105.43,  aic = 216.86
```

3.5 Estimation

Example: Lake Huron data (de-trended)

```
arima(Lake.Yt,order=c(2,0,0))

##
## Call:
## arima(x = Lake.Yt, order = c(2, 0, 0))
##
## Coefficients:
##          ar1          ar2  intercept
##          1.0026  -0.2901    0.0198
## s.e.  0.0977    0.1005    0.2352
##
## sigma^2 estimated as 0.4589:  log likelihood = -101.44,  aic = 210.89
```

3.5 Estimation

Example: Lake Huron data (de-trended)

```
arima(Lake.Yt,order=c(0,0,1))

##
## Call:
## arima(x = Lake.Yt, order = c(0, 0, 1))
##
## Coefficients:
##          mal  intercept
##      0.7821   -0.0037
## s.e.  0.0646    0.1391
##
## sigma^2 estimated as 0.602:  log likelihood = -114.66,  aic = 235.33
```

3.5 Estimation

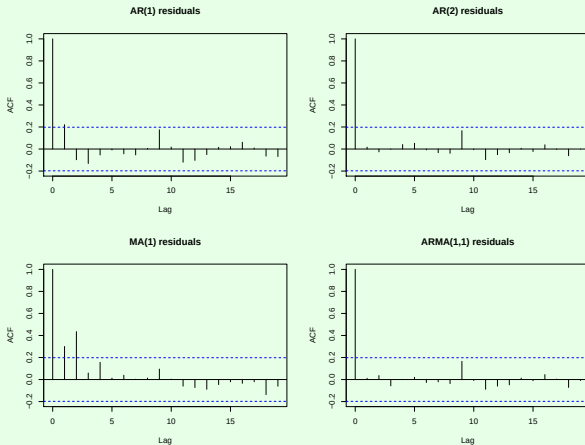
Example: Lake Huron data (de-trended)

```
arima(Lake.Yt,order=c(1,0,1))

##
## Call:
## arima(x = Lake.Yt, order = c(1, 0, 1))
##
## Coefficients:
##          ar1      ma1  intercept
##          0.6508  0.3567    0.0330
## s.e.    0.0948  0.1156    0.2607
##
## sigma^2 estimated as 0.4589:  log likelihood = -101.44,  aic = 210.87
```

3.5 Estimation

Example: Lake Huron data (de-trended)



3.6 Model selection

A series of ARMA models might be fitted, and then compared with reference to different aspects:

- fidelity to the data
- model complexity
- residuals
- forecasting performance

3.6 Model selection

Akaike Information Criterion (AIC)

The AIC model selection method is widely used; it trades off

- **fidelity to the observed data** (goodness of fit)
- **model complexity** (number of parameters fitted)

3.6 Model selection

For model M with parameters β yielding likelihood $L_M(\beta)$, the AIC is defined as

$$\text{AIC}(M, \hat{\beta}) = -2 \log L_M(\hat{\beta}) + 2d$$

where d is the dimension of β .

For the ARMA(p, q) model, there are $p + q + 1$ parameters $(\phi_p, \theta_q, \sigma_Z^2)$.

3.6 Model selection

For the ARMA(p, q) model,

$$\text{AIC}(M_{p,q}, \hat{\beta}) = -2\ell_M(\hat{\phi}_p, \hat{\theta}_q, S(\hat{\phi}_p, \hat{\theta}_q)/n) + 2(p + q + 1)$$

where ℓ_M is the log-likelihood, and

$$\hat{\sigma}_Z^2 = \frac{1}{n} S(\hat{\phi}_p, \hat{\theta}_q)$$

The model with the **smallest** (ie most negative) AIC is selected.

3.6 Model selection

Notes

- The AIC is a “within sample” criterion, and fits to the observed data. This might not always reflect the objective of the analysis we might be more interested in forecasting performance.
- The AIC is sometimes regarded to give an insufficient penalty for complexity, and chooses models that are too complex.
- As $n \rightarrow \infty$, the AIC will choose the best approximating model to the data generating model

3.6 Model selection

Corrected AIC

To fix the small sample issue, a correction was suggested:

$$AICC(M_{p,q}, \hat{\beta}) = -2\ell_M(\hat{\phi}_p, \hat{\theta}_q, S(\hat{\phi}_p, \hat{\theta}_q)/n) + \frac{2(p+q+1)n}{n-p-q-2}$$

that is the penalty $2d$ is replaced by $2dn/(n-p-q-2)$. This uses a bias-corrected estimator of σ_Z^2 .

For large n , the criterion reverts to the AIC.

3.6 Model selection

Bayesian Information Criterion (BIC)

The BIC (or Schwarz criterion) uses a different penalty

$$BIC(M_{p,q}, \hat{\beta}) = -2\ell_M(\hat{\phi}_p, \hat{\theta}_q, S(\hat{\phi}_p, \hat{\theta}_q)/n) + (p + q + 1) \log n$$

that is the penalty $2d$ is replaced by $d \log n$.

For large n , the criterion correctly selects the true model (provided the true model is one that has been fitted).

3.7 Residual checking

The residual sequence $\{\widehat{W}_t\}$ estimated by

$$\widehat{w}_t = \frac{x_t - \widehat{x}_t}{\sqrt{r_{t-1}}}$$

(that is, the one-step ahead prediction errors standardized by their standard deviation) should have the property that $\{\widehat{W}_t\}$ is an approximate white noise sequence.

3.7 Residual checking

In fact, if the model fitted is the correct model

$$\widehat{W}_t \xrightarrow{\text{m.s.}} Z_t \sim \text{WN}(0, \sigma_Z^2)$$

so the observed residual sequence should be approximately uncorrelated, zero mean and variance σ_Z^2 .

We can also compute

$$\widehat{R}_t = \frac{\widehat{W}_t}{\widehat{\sigma}_Z}$$

which should be a unit variance white noise process.

Chapter 4

Non-Stationary and Seasonal Models

4.1 Non-Stationary and Seasonal Models

Models with non-stationary components can be constructed as straightforward extensions to ARMA models

- Integrated Models: ARIMA
- Seasonal Models: SARIMA

It is also possible to specify time-varying or stochastic mean or variance models.

4.2 ARIMA Models

If d is a non-negative integer, then $\{X_t\}$ is an ARIMA(p, d, q) process if $\{Y_t\}$ defined by

$$Y_t = (1 - B)^d X_t$$

is a (causal) ARMA(p, q) process, that is

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

where $\{Z_t\} \sim \text{WN}(0, \sigma_Z^2)$; $\{X_t\}$ is stationary only if $d = 0$.

The ARIMA(p, d, q) process can be made stationary by d th order differencing.

4.2 ARIMA Models

Example: ARIMA(1, 1, 0) vs ARIMA(2, 0, 0)

If $\{Z_t\} \sim \text{WN}(0, \sigma_Z^2)$:

- **Stationary** ARIMA(2, 0, 0) $\equiv$ ARMA(2, 0) $\equiv$ AR(2)

$$(1 - 0.6B)(1 - 0.99B)X_t = Z_t$$

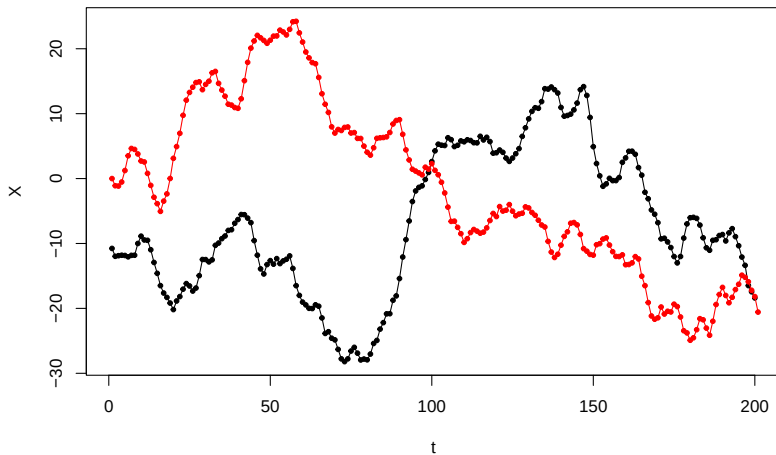
- **Non-stationary** ARIMA(1, 1, 0)

$$(1 - 0.6B)(1 - B)X_t = Z_t$$

4.2 ARIMA Models

```
n<-200;set.seed(164664)
r1<-0.6;r2<-0.99
x1<-arima.sim(n,model=list(ar=c(r1+r2,-r1*r2),order=c(2,0,0)))
x2<-arima.sim(n,model=list(ar=c(r1),order=c(1,1,0)))
par(mar=c(4,4,3,2))
plot(x1,xlab=expression(t),type='p',pch=19,cex=0.6,ylim=range(c(x1,x2)),ylab='X')
lines(x1);
points(x2,pch=19,cex=0.6,col='red');lines(x2,col="red")
```

4.2 ARIMA Models

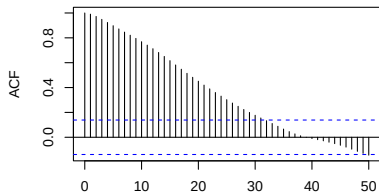


4.2 ARIMA Models

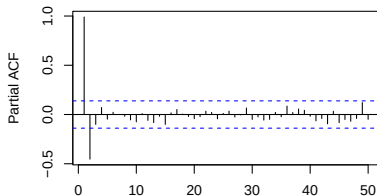
```
par(mfrow=c(2,2),mar=c(3,4,3,2))  
acf(x1,main="ACF",lag.max=50)  
pacf(x1,main="PACF",lag.max=50)  
acf(x2,main="ACF",lag.max=50)  
pacf(x2,main="PACF",lag.max=50)
```

4.2 ARIMA Models

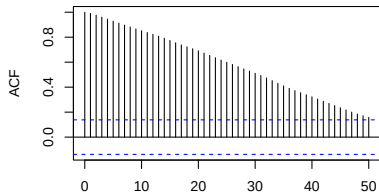
ACF



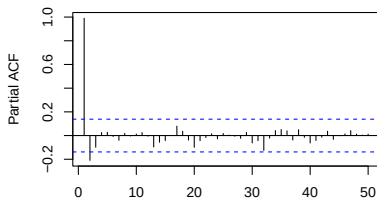
PACF



ACF



PACF



4.2 ARIMA Models

Consider the stationary AR(2) model with complex reciprocal roots

$$(1 - \xi B)(1 - \bar{\xi} B)X_t = Z_t$$

and the corresponding AR polynomial

$$\Phi(z) = (1 - \xi z)(1 - \bar{\xi} z) = 1 - 2\operatorname{Re}(\xi)z + |\xi|^2 z^2$$

Suppose that $\xi = \exp\{-i\omega\}/r$, where $r > 1$ and $-\pi < \omega < \pi$. Then

$$\Phi(z) = 1 - \left(\frac{2}{r} \cos \omega\right) z + \left(\frac{1}{r^2}\right) z^2.$$

4.2 ARIMA Models

It follows that

$$\rho_X(h) = \left(\frac{1}{r}\right)^h \frac{\sin(h\omega + \lambda)}{\sin \lambda} \quad h = 0, \pm 1, \pm 2, \dots$$

where

$$\lambda = \arctan \left[\left(\frac{r^2 + 1}{r^2 - 1} \right) \tan \omega \right].$$

In the limit as $r \downarrow 1$

$$\rho_X(h) \longrightarrow \cos(h\omega).$$

4.2 ARIMA Models

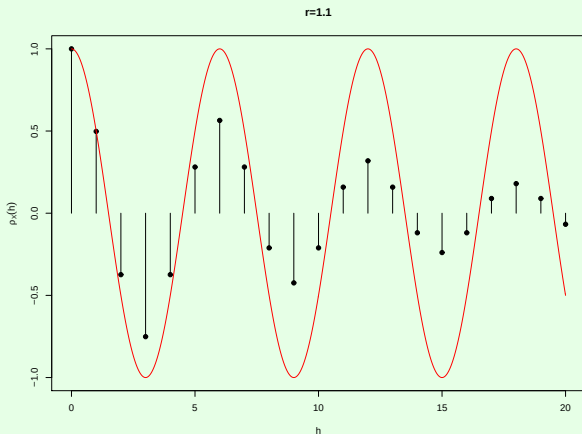
We examine this model as

- r takes values tending to 1 from above
- ω fixed.

and compare with the pure `cos` function.

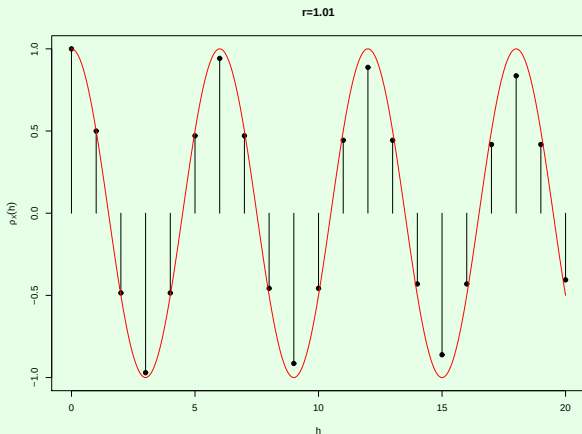
4.2 ARIMA Models

Example: $r = 1.1, \omega = \pi/3$



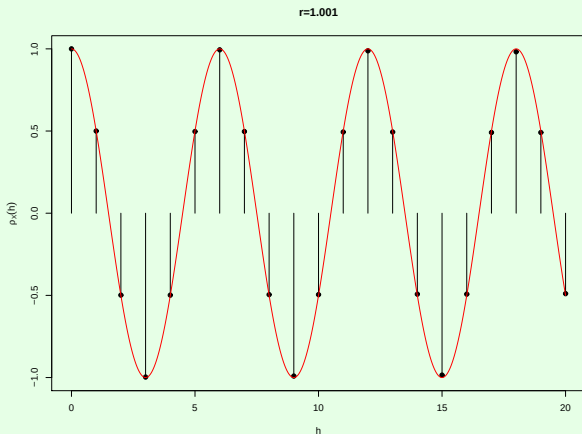
4.2 ARIMA Models

Example: $r = 1.01, \omega = \pi/3$



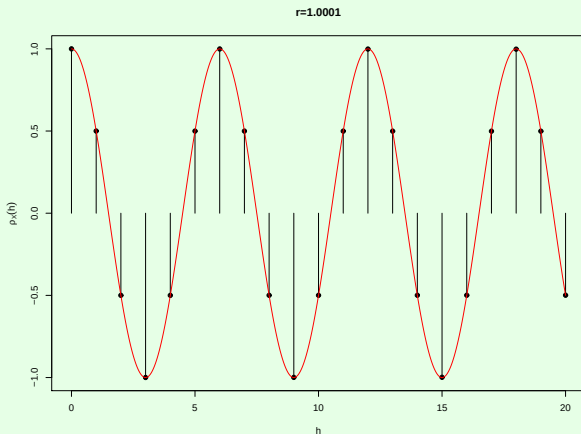
4.2 ARIMA Models

Example: $r = 1.001, \omega = \pi/3$



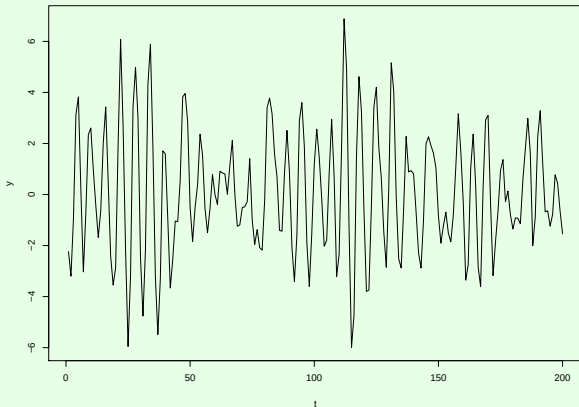
4.2 ARIMA Models

Example: $r = 1.0001, \omega = \pi/3$



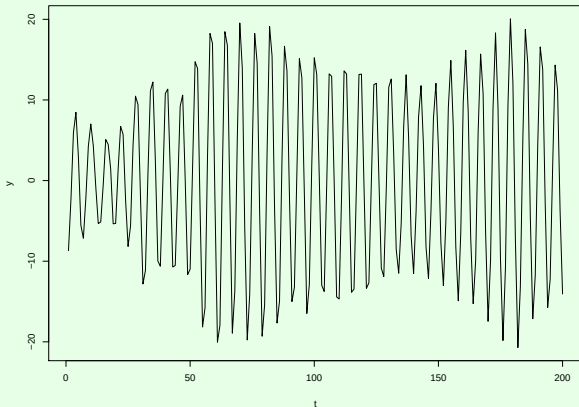
4.2 ARIMA Models

Example: $r = 1.1, \omega = \pi/3$



4.2 ARIMA Models

Example: $r = 1.001, \omega = \pi/3$



4.2 ARIMA Models

This demonstrates that we can construct non-stationary processes with different behaviour from the random walk model

$$(1 - B)X_t = Z_t$$

that is, we can construct a **periodic** non-stationary model by choosing

$$|\xi| \longrightarrow 1$$

with ξ having an imaginary part.

4.3 Unit Roots

The presence of a **unit root** of the AR polynomial equation

$$\Phi(z) = 0$$

that is, a solution where $|\xi| = 1$, renders the resulting AR or ARMA process non-stationary.

Therefore it is crucial to be able to identify processes where a unit root is present, and to distinguish this from a root near to, but not equal to, one in modulus.

4.3 Unit Roots

The Dickey-Fuller Test: Suppose $\{Z_t\} \sim \text{WN}(0, \sigma_Z^2)$, and

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

with $|\phi| < 1$. Then, by standard asymptotic theory, ML or moment-based estimators $\hat{\phi}$ satisfy

$$\sqrt{n}(\hat{\phi} - \phi) \xrightarrow{d} \mathcal{N}(0, 1 - \phi^2)$$

as $n \longrightarrow \infty$.

If $|\phi| = 1$, the asymptotic result does not hold.

4.3 Unit Roots

We may study the distribution of the statistic

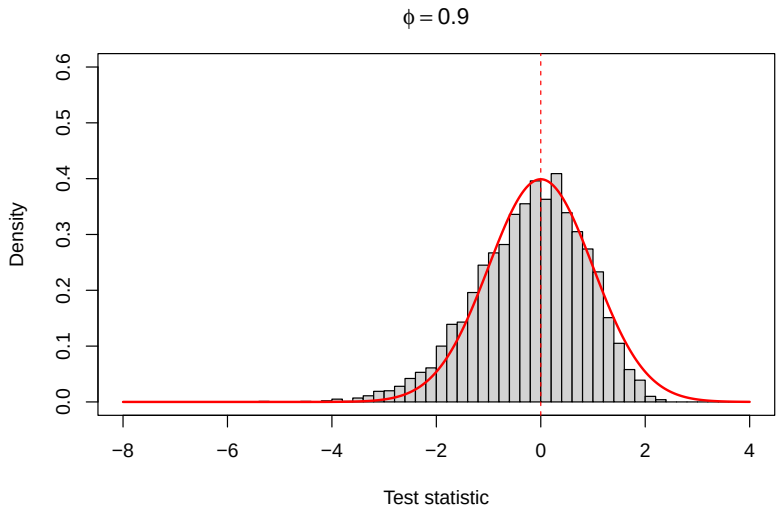
$$\frac{\sqrt{n}(\hat{\phi} - \phi)}{\sqrt{1 - \phi^2}}$$

in simulation for any value of ϕ , which should be standard Normally distributed when n is large.

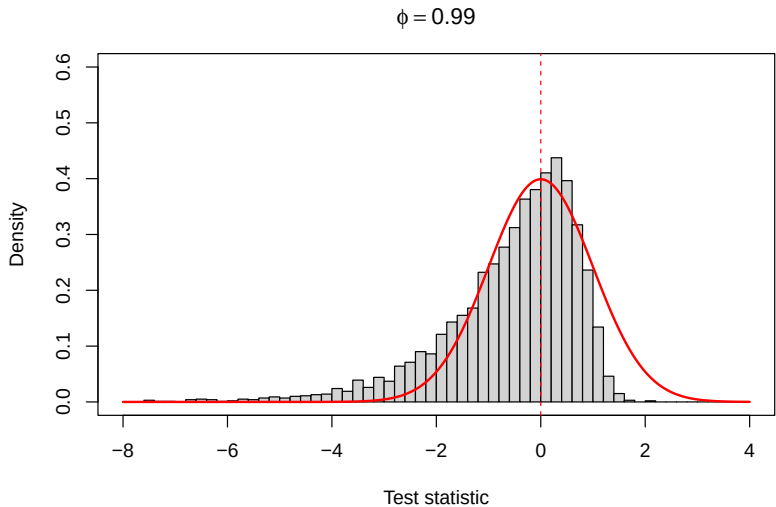
In the following simulations, we use $n = 500$ for

- $\phi = 0.9$
- $\phi = 0.99$
- $\phi = 0.999$

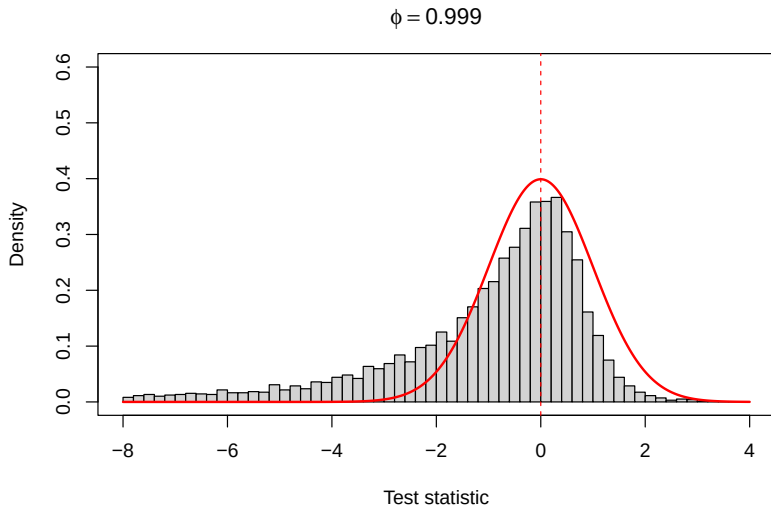
4.3 Unit Roots



4.3 Unit Roots



4.3 Unit Roots



4.3 Unit Roots

The Dickey-Fuller Test: Suppose $\{Z_t\} \sim \text{WN}(0, \sigma_Z^2)$, and

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

Then

$$\begin{aligned}(1 - B)X_t &= X_t - X_{t-1} = \mu(1 - \phi) + (\phi - 1)X_{t-1} + Z_t \\ &= \phi_0^* + \phi_1^*X_{t-1} + Z_t\end{aligned}$$

where

$$\phi_0^* = \mu(1 - \phi) \qquad \phi_1^* = \phi - 1.$$

can be estimated by OLS, regressing X_t on X_{t-1} .

4.3 Unit Roots

To test the hypothesis

$$H_0 : \phi_1^* = 0$$

which is equivalent to

$$H_0 : \phi_1 = 1$$

the **Dickey-Fuller** test statistic

$$t = \frac{\hat{\phi}_1^*}{\text{e.s.e}(\hat{\phi}_1^*)}$$

can be used. Numerator and denominator are available from the linear regression.

4.3 Unit Roots

However, if H_0 is **true**, and $\{X_t\}$ is non-stationary, the null distribution is non-standard.

In this simple case, the form of the null distribution can be computed, but it is not available in closed form. It is straightforward to compute using simulation.

4.3 Unit Roots

```
n<-10
Zt<-rnorm(n)
Xt<-cumsum(Zt)
X<-Xt[1:(n-1)]
tvals<-summary(lm(diff(Xt)~-1+X))$coef
tvals

##      Estimate Std. Error  t value Pr(>|t|)
## X 0.1527851   0.1101742  1.386759 0.2029315

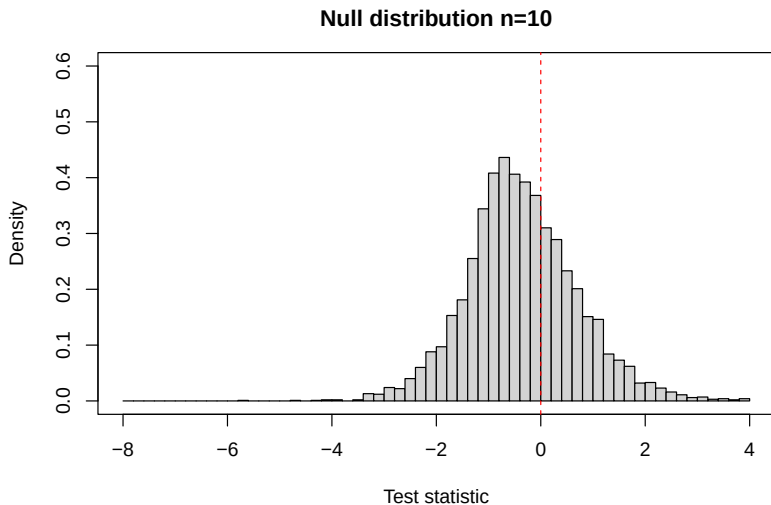
tstat<-(tvals[1,1]-1)/tvals[1,2]
tstat

## [1] -7.689774
```

4.3 Unit Roots

```
nreps<-5000 #5000 replications
n<-10
tstat10<-rep(0,nreps)
for(irep in 1:nreps){
  Zt<-rnorm(n)
  Xt<-cumsum(Zt)
  X<-Xt[1:(n-1)]
  tvals<-summary(lm(diff(Xt)~-1+X))$coef
  tstat<-tvals[1,1]/tvals[1,2]
  tstat10[irep]<-tstat
}
tstat10<-tstat10[tstat10 > -8 & tstat10 < 4]
par(mar=c(4,4,3,2))
hist(tstat10,freq=FALSE,breaks=seq(-8,4,by=0.2),ylim=range(0,0.6),
     main='Null distribution n=10',xlab='Test statistic')
box()
abline(v=0,lty=2,col='red')
```

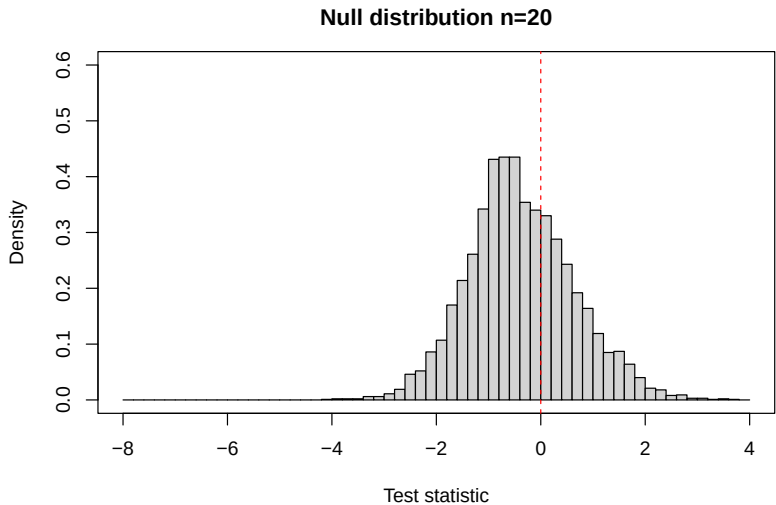
4.3 Unit Roots



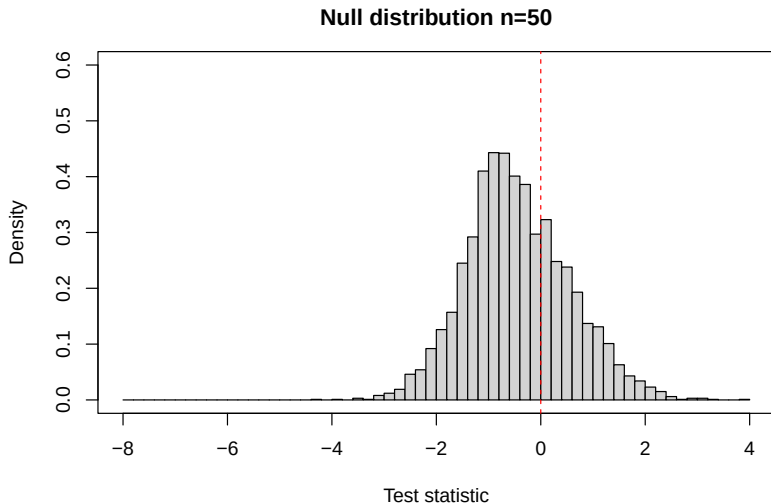
4.3 Unit Roots

```
nreps<-5000 #5000 replications
n<-20
tstat20<-rep(0,nreps)
for(irep in 1:nreps){
  Zt<-rnorm(n)
  Xt<-cumsum(Zt)
  X<-Xt[1:(n-1)]
  tvals<-summary(lm(diff(Xt)~-1+X))$coef
  tstat<-tvals[1,1]/tvals[1,2]
  tstat20[irep]<-tstat
}
tstat20<-tstat20[tstat20 > -8 & tstat20 < 4]
par(mar=c(4,4,3,2))
hist(tstat20,freq=FALSE,breaks=seq(-8,4,by=0.2),ylim=range(0,0.6),
     main='Null distribution n=20',xlab='Test statistic')
box()
abline(v=0,lty=2,col='red')
```

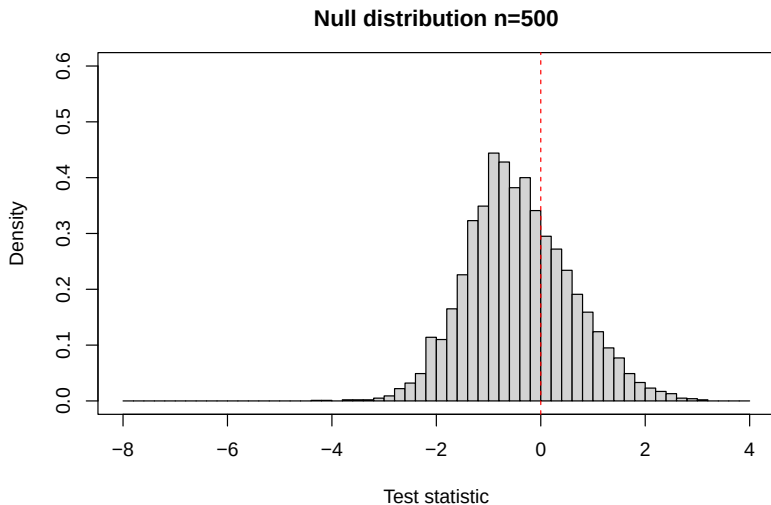
4.3 Unit Roots



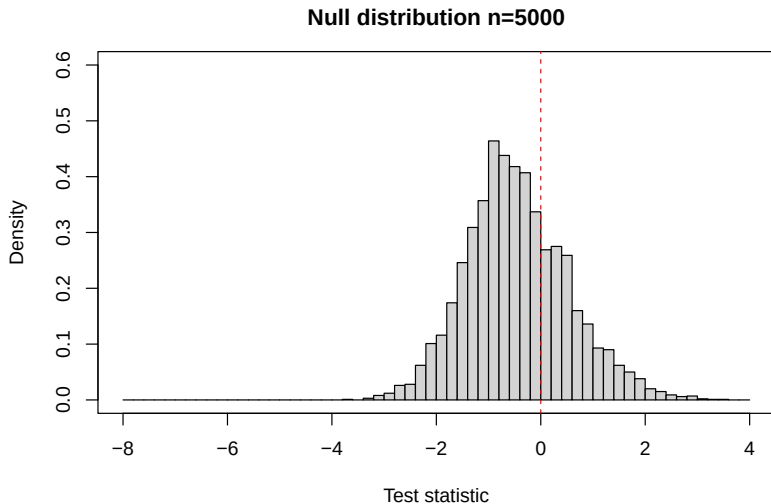
4.3 Unit Roots



4.3 Unit Roots



4.3 Unit Roots



4.3 Unit Roots

The test can be extended to AR(p) processes: if

$$X_t - \mu = \sum_{j=1}^p \phi_j (X_{t-j} - \mu) + Z_t$$

then for $p \geq 2$,

$$(1 - B)X_t = \phi_0^* + \phi_1^* X_{t-1} + \sum_{j=2}^p \phi_j^* (X_{t-j+1} - X_{t-j}) + Z_t$$

where

$$\phi_0^* = \mu(1 - \phi_1 - \phi_2 - \cdots - \phi_p)$$

and

$$\phi_1^* = \sum_{j=1}^p \phi_j - 1 \qquad \phi_j^* = - \sum_{k=j}^p \phi_k \qquad j = 2, \dots, p.$$

4.3 Unit Roots

If $\{X_t\} \sim \text{AR}(p)$ with one unit root, then

$$(1 - B)X_t \sim \text{AR}(p - 1)$$

therefore a unit root test can be carried out by examining the hypothesis

$$H_0 : \phi_1^* = 0$$

by fitting a regression of $(1 - B)X_t$ on

$$1, X_{t-1}, X_{t-1} - X_{t-2}, \dots, X_{t-p+1} - X_{t-p}$$

The null distribution of the test statistic changes with p .

4.3 Unit Roots

The general procedure is the **Augmented Dickey-Fuller** test.

R Code: The library `tseries` contains the function `adf.test`

4.3 Unit Roots

```
library(tseries)

## Registered S3 method overwritten by 'quantmod':
##   method              from
##   as.zoo.data.frame zoo

x <- rnorm(1000)
adf.test(x,k=0)

## Warning in adf.test(x, k = 0): p-value smaller than printed p-value

##
## Augmented Dickey-Fuller Test
##
## data:  x
## Dickey-Fuller = -33.086, Lag order = 0, p-value = 0.01
## alternative hypothesis: stationary

adf.test(cumsum(x),k=0)

##
## Augmented Dickey-Fuller Test
##
## data:  cumsum(x)
## Dickey-Fuller = -1.6625, Lag order = 0, p-value = 0.7212
## alternative hypothesis: stationary
```

4.3 Unit Roots

The series of length n , x , is a white noise process, but the vector constructed by `cumsum(x)` is a random walk.

The lag order, p , can be specified by using a `k=p+1` argument to the function.

4.3 Unit Roots

Example: S&P 500: Oct 03, 2001 - Oct 03, 2011



4.3 Unit Roots

```
sp500<-read.csv("https://www.math.mcgill.ca/dstephens//545//Data//sp500.csv")
head(sp500)
```

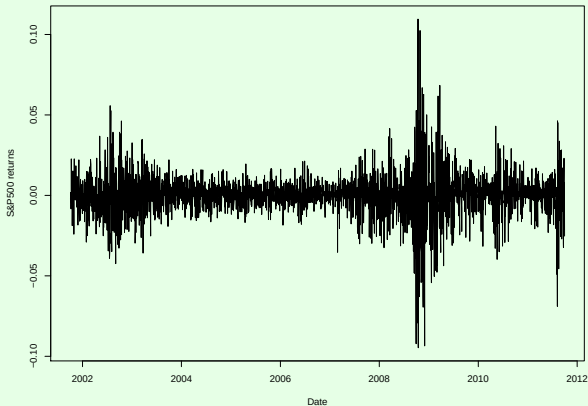
```
##           Date    Open    High    Low   Close      Volume Adj.Close
## 1 30/09/2011 1159.93 1159.93 1131.34 1131.42 4416790000    1131.42
## 2 29/09/2011 1151.74 1175.87 1139.93 1160.40 5285740000    1160.40
## 3 28/09/2011 1175.39 1184.71 1150.40 1151.06 4787920000    1151.06
## 4 27/09/2011 1163.32 1195.86 1163.32 1175.38 5548130000    1175.38
## 5 26/09/2011 1136.91 1164.19 1131.07 1162.95 4762830000    1162.95
## 6 23/09/2011 1128.82 1141.72 1121.36 1136.43 5639930000    1136.43
```

```
xt<-rev(sp500[,5])
adf.test(xt,k=0)
```

```
##
## Augmented Dickey-Fuller Test
##
## data: xt
## Dickey-Fuller = -2.0111, Lag order = 0, p-value = 0.5736
## alternative hypothesis: stationary
```

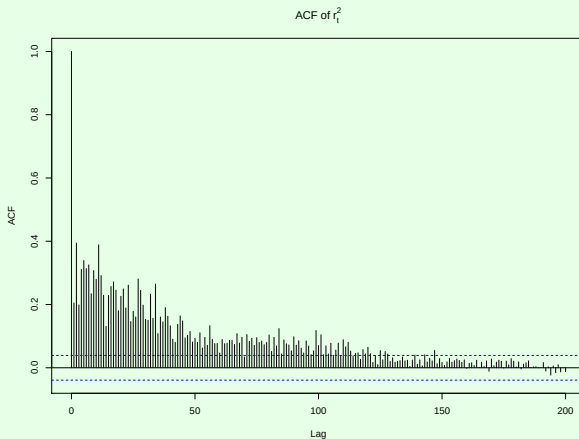
4.3 Unit Roots

Example: S&P 500: log returns $y_t = \log(x_{t+1}/x_t)$



4.3 Unit Roots

Example: S&P 500: ACF of y_t^2



4.3 Unit Roots

Unit Roots in MA processes: If $\{X_t\} \sim \text{ARMA}(p, q)$ is causal and invertible, then

$$Y_t = (1 - B)X_t$$

is an $\text{ARMA}(p, q + 1)$ process. This follows as

$$\Phi(B)X_t = \Theta(B)Z_t$$

with if $\{Z_t\} \sim \text{WN}(0, \sigma_Z^2)$, then

$$\Phi(B)Y_t = \Phi(B)(1 - B)X_t = (1 - B)\Theta(B)Z_t.$$

4.3 Unit Roots

The MA polynomial for $\{Y_t\}$ is

$$\Theta_Y(z) = (1 - z)\Theta(z)$$

Thus a test for a unit root in the MA polynomial for $\{Y_t\}$ is equivalent to a test for over-differencing of a stationary, causal and invertible ARMA process.

The asymptotic null distribution of the (likelihood-ratio) test is non-standard in this setting.

4.4 Forecasting ARIMA

For the general ARIMA(p, d, q) model

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

we perform first forecasting of

$$(1 - B)^d X_t = Y_t$$

where $\{Y_t\}$ is a causal ARMA(p, q) process. Then

$$X_t = Y_t - \sum_{j=1}^d \binom{d}{j} (-1)^j X_{t-j}$$

for $t > 0$.

4.4 Forecasting ARIMA

Therefore, for $h = 1, 2, \dots$, given $X_{1:n}$, we have that

$$\hat{P}_n X_{n+h} = \hat{P}_n Y_{n+h} - \sum_{j=1}^d \binom{d}{j} (-1)^j \hat{P}_n X_{n+h-j}$$

for which

- $\hat{P}_n Y_{n+h}$ is computed using the methods devised earlier for stationary processes
- $\hat{P}_n X_{n+h-j}$ is computed recursively, noting that

$$\hat{P}_n X_{n+1-j} = X_{n+1-j}$$

for $j \geq 1$.

4.5 Seasonal ARIMA Models

Recall the lag- s difference operator, B^s , $s \geq 1$,

$$(1 - B^s)X_t = X_t - X_{t-s}$$

A **seasonal ARIMA model** utilizes this operator: for example

$$(1 - B^s)\Phi(B)X_t = \Theta(B)Z_t$$

applies a seasonal differencing with period s .

4.5 Seasonal ARIMA Models

If d and D are non-negative integers, then $\{X_t\}$ is a seasonal ARIMA (SARIMA) process with period s

$$\{X_t\} \sim \text{ARIMA}(p, d, q) \times (P, D, Q)_s$$

of

$$Y_t = (1 - B)^d (1 - B^s)^D X_t$$

is a causal ARMA process determined by

$$\Phi(B)\Phi_S(B^s)Y_t = \Theta(B)\Theta_S(B^s)Z_t$$

where $\{Z_t\} \sim \text{WN}(0, \sigma_Z^2)$.

4.5 Seasonal ARIMA Models

In this formulation

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

$$\Phi_S(z) = 1 - \phi_1^{(S)} z - \cdots - \phi_P^{(S)} z^P$$

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

$$\Theta_S(z) = 1 + \theta_1^{(S)} z + \cdots + \theta_Q^{(S)} z^Q$$

Note that

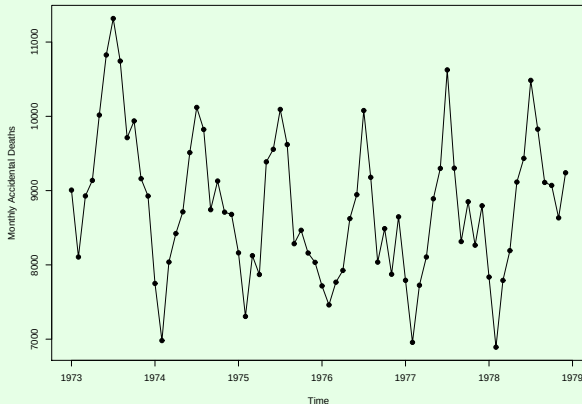
$$Y_t = (1 - B)^d (1 - B^S)^D X_t$$

is an ARMA($p + sP, q + sQ$) process with AR and MA polynomials

$$\Phi(B)\Phi_S(B^S) \qquad \Theta(B)\Theta_S(B^S)$$

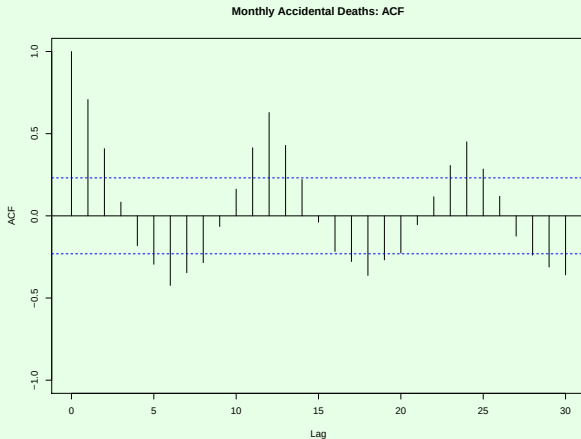
4.5 Seasonal ARIMA Models

Example: US Accidental Deaths: Jan 1973 - Dec 1978



4.5 Seasonal ARIMA Models

Example: US Accidental Deaths: Jan 1973 - Dec 1978



4.5 Seasonal ARIMA Models

Example: US Accidental Deaths: Jan 1973 - Dec 1978

Consider two SARIMA models:

- A. $\text{ARIMA}(1, 0, 1) \times (0, 1, 0)_{12}$: This model fits an $\text{ARMA}(1, 1)$ model to the seasonally-differenced data with $s = 12$.
- B. $\text{ARIMA}(1, 0, 1) \times (1, 0, 0)_{12}$ This model fits a SARIMA model with $p = q = 1$, $P = 1$, $Q = 0$, and $s = 12$.

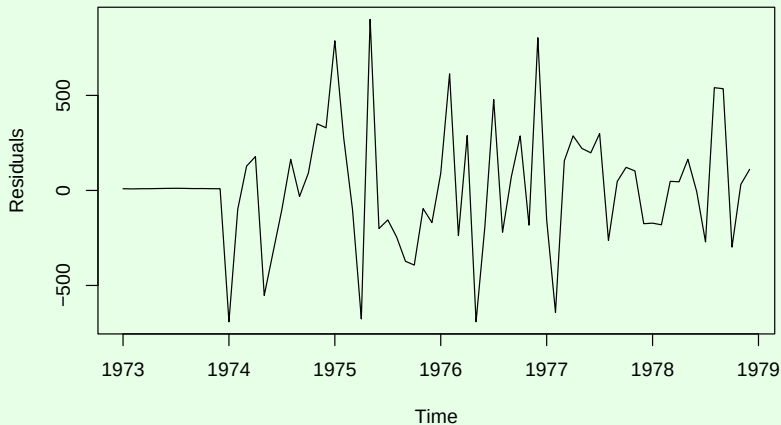
Model A posits a non-stationary model, and requires first differencing at lag 12. Model B posits a stationary model.

4.5 Seasonal ARIMA Models

```
model.a<-arima(USAccDeaths, order = c(1,0,1),  
              seasonal = list(order=c(0,1,0),period=12))  
model.b<-arima(USAccDeaths, order = c(1,0,1),  
              include.mean=FALSE, seasonal = list(order=c(1,0,0),period=12))  
res.a<-residuals(model.a)  
res.b<-residuals(model.b)
```

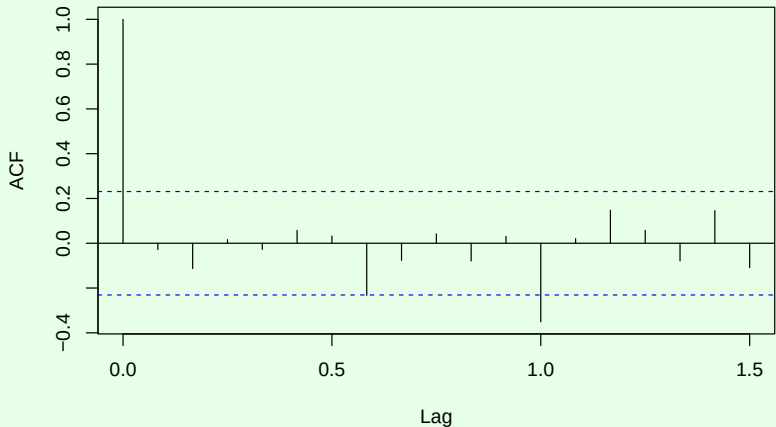
4.5 Seasonal ARIMA Models

Example: Model A: Residuals



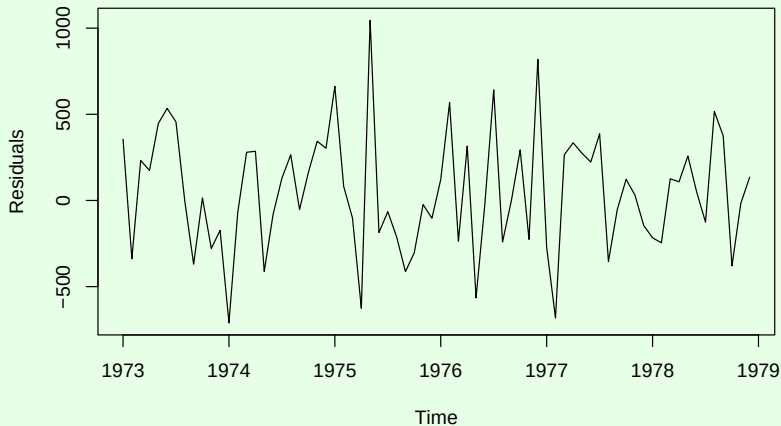
4.5 Seasonal ARIMA Models

Example: Model A: Residuals ACF



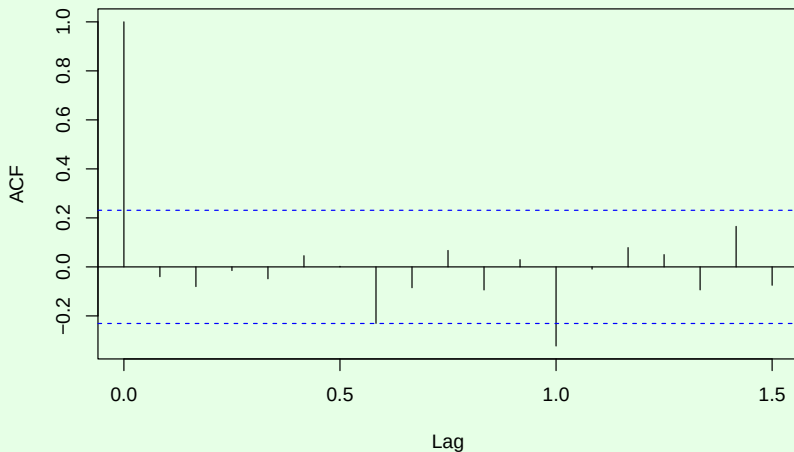
4.5 Seasonal ARIMA Models

Example: Model B: Residuals



4.5 Seasonal ARIMA Models

Example: Model B: Residuals ACF



4.5 Seasonal ARIMA Models

Model A:

```
model.a

##
## Call:
## arima(x = USAccDeaths, order = c(1, 0, 1), seasonal = list(order = c(0, 1, 0),
##      period = 12))
##
## Coefficients:
##          ar1          ma1
##      0.9421   -0.4328
## s.e.  0.0583   0.1445
##
## sigma^2 estimated as 123848:  log likelihood = -437.61,  aic = 881.23
```

4.5 Seasonal ARIMA Models

Model B:

```
model.b

##
## Call:
## arima(x = USAccDeaths, order = c(1, 0, 1), seasonal = list(order = c(1, 0, 0),
##      period = 12), include.mean = FALSE)
##
## Coefficients:
##          ar1          ma1          sar1
##          0.9852   -0.3342   0.9098
## s.e.    0.0127    0.1368   0.0395
##
## sigma^2 estimated as 121834:  log likelihood = -537.14,  aic = 1082.29
```

4.6 Non-stationarity in Mean

Non-stationarity in mean arises when there is a time-varying mean or trend structure, as described in earlier sections. Previously OLS was used to estimate parameters in a parametric trend, m_t ; OLS assumes that the residual errors are uncorrelated.

More efficient estimation can be used to take into account the correlation present if the residual error process is stationary, but exhibits autocorrelation.

4.6 Non-stationarity in Mean

Suppose

$$Y_t = m_t + X_t = \mathbf{w}_t^\top \beta + X_t$$

where

- $\mathbf{w}_t$ is a vector of (exogenous) covariates
- β is a vector of (fixed effects) coefficients
- $\{X_t\} \sim \text{ARMA}(p, q)$, with parameters (ϕ, θ)

We wish to estimate

$$\beta \quad (\phi_1, \dots, \phi_p)^\top \quad (\theta_1, \dots, \theta_q)^\top$$

4.6 Non-stationarity in Mean

Suppose

$$\text{Cov}[X_{1:n}] = \mathbb{E}[X_{1:n}X_{1:n}^\top] = \mathbf{\Gamma}_n(\phi, \theta).$$

If (ϕ, θ) are known, then we may estimate β using

Weighted Least Squares (WLS)

as

$$\hat{\beta} = (\mathbf{X}^\top \mathbf{\Gamma}_n(\phi, \theta)^{-1} \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{\Gamma}_n(\phi, \theta)^{-1} \mathbf{Y}$$

with variance

$$(\mathbf{X}^\top \mathbf{\Gamma}_n(\phi, \theta)^{-1} \mathbf{X})^{-1}$$

4.6 Non-stationarity in Mean

This suggests an iterative procedure

- Initialize $\phi^{(0)}, \theta^{(0)}$
- Compute $\beta^{(0)}$ using WLS setting $(\phi, \theta) = (\phi^{(0)}, \theta^{(0)})$.
- Compute the residuals

$$\hat{x}_t = y_t - \mathbf{w}_t^\top \beta^{(0)}$$

and then fit an ARMA(p, q) model to $\{\hat{x}_t\}$ to update estimates $(\phi^{(1)}, \theta^{(1)})$.

- Continue iterations $r = 1, 2, \dots$ until iterates at step r

$$(\beta^{(r)}, \phi^{(r)}, \theta^{(r)})$$

cease to change.