Autoregressive Models

<table>
<tr><th>Properties $\phi(B)(Y_t - \mu) = Z_t$</th><th>Regimes of $\mathrm{AR}(1)$</th></tr>
<tr><td>

- $Y_t = \phi_0 + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \cdots + \phi_p Y_{t-p} + Z_t$
- $\phi(z) = 1 - \phi_1 z - \phi_2 z^2 - \cdots - \phi_p z^p$

<u>Regimes for $\mathrm{AR}(p)$</u>:

- Causal, stationary: all roots of $\phi(z)$ have modulus > 1 (unique solution)
- Non-causal, stationary: at least one root has modulus < 1
- Non-stationary: one root has modulus $= 1$

<u>Other Properties</u>:

- $pacf(h) = \begin{cases} \phi_p, & h = p \\ 0, & h > p \end{cases}$
- Equivalent causal condition for $\mathrm{AR}(2)$
  - $\phi_1 + \phi_2 < 1$
  - $\phi_2 - \phi_1 < 1$
  - $|\phi_2| < 1$
- [ACF for $\mathrm{AR}(2)$]
  - $Y_t - \phi_1 Y_{t-1} - \phi_2 Y_{t-2} = Z_t$
  - $\mathrm{Cov}[Y_t - \phi_1 Y_{t-1} - \phi_2 Y_{t-2}, Y_{t-h}] = 0$
  - $\rho(h) - \phi_1 \rho(h-1) - \phi_2 \rho(h-2) = 0$
  - $\rho(0) = 1$
  - $\rho(1) = \rho(-1) = \frac{\phi_1}{1-\phi_2}$
  - $\rho(h) = c_1 z_1^{-h} + c_2 z_2^{-h}$
  - $\rho(h) = z_0^{-h}(c_1 + c_2 h)$
  - $\rho(h) = c_1 z_1^{-h} + \bar{c}_1 \bar{z}_1^{-h}$

</td><td>

- $Y_t = \phi_0(1 + \phi_1 + \cdots + \phi_1^{t-1}) + (Z_t + \phi_1 Z_{t-1} + \cdots + \phi_1^{t-1} Z_1) + \phi_1^t y_0$
- $Y_t^* = \frac{\phi_0}{1-\phi_1} + \sum_{j=0}^{\infty} \phi_1^j Z_{t-j}$

<u>Case #1</u>: $|\phi_1| < 1$, $\exists Y_t^*$ stationary, causal

- $\mathbb{E}[Y_t^*] = \frac{\phi_0}{1-\phi_1}$
- $\mathrm{Cov}[Y_t^*, Y_{t+h}^*] = \frac{\sigma^2 \phi_1^{|h|}}{1-\phi_1^2}$
- $\mathrm{Corr}[Y_t^*, Y_{t+h}^*] = \phi_1^{|h|}$

<u>Case #2</u>: $|\phi_1| > 1$, $\exists Y_t^*$ stationary, non-causal

- $Y_{t-1} = -\frac{\phi_0}{\phi_1} + \frac{Y_t}{\phi_1} - \frac{Z_t}{\phi_1}$
- $Y_t^* = \frac{\phi_0}{1-\phi_1} - \sum_{j=1}^{\infty} \frac{Z_{t+j}}{\phi_1^j}$
- If initialized in the past, non-stationary and explosive.

<u>Case #3</u>: $|\phi_1| = 1$: non-stationary

- When $\phi_1 = 1$, $Y_t = t\phi_0 + (Z_t + \cdots + Z_1) + y_0$ is non-stationary $\mathrm{Var}[Y_t] = t\sigma^2$
- $\phi_1 = -1$: also non-stationary

</td></tr>
<tr><th>Backshift Calculus</th><th></th></tr>
<tr><td>

- $\phi(B) Y_t = \phi_0 + Z_t$
- $Y_t = \frac{1}{\phi(B)}(\phi_0 + Z_t) = (\mathbb{I} - a_1 B)^{-1} \ldots \left(\mathbb{I} - a_p B\right)^{-1}(\phi_0 + Z_t)$
- $= (\mathbb{I} + a_1 B + a_1^2 B^2 + \cdots) \ldots \left(\mathbb{I} + a_p B + a_p^2 B^2 + \cdots\right)(\phi_0 + Z_t)$
- $\psi(z) = \frac{1}{\phi(z)}$
  - $\psi_0 = 1$
  - $\psi_1 = \phi_1$
  - $\psi_2 = \phi_1^2 + \phi_2$
- Can also set $Y_t = \psi(B) Z_t$ and match coefficients ($\phi(B)\psi(B) = \mathbb{I}$)

</td><td></td></tr>
</table>

Moving Average Models

Moving Average $\mathrm{MA}(q)$, $Y_t - \mu = \theta(B)Z_t$

- Summation of random noises
- $Y_t = \mu + Z_t + \theta_1 Z_{t-1} + \cdots + \theta_q Z_{t-q}$
- $\theta_0 = 1$, $Z_t \sim N(0, \sigma^2)$
- $q + 2$ parameters
- $\mathrm{Cov}[Y_t, Y_{t+h}] = \begin{cases} \sigma^2 \sum_{j=0}^{q-h} \theta_j \theta_{j+h}, & 0 \le h \le q \\ 0, & h > q \end{cases}$
- Joint density of $Y_1, \dots, Y_n$ is multivariate normal $N(\mu\mathbb{1}, \Sigma)$ where $\Sigma_{i,j} = \mathrm{Cov}[Y_i, Y_j]$
- Likelihood $\left(\frac{1}{\sqrt{2\pi}}\right)^n |\det\Sigma|^{\frac{1}{2}} e^{-\frac{1}{2}(y-\mu\mathbb{1})^T \Sigma^{-1}(y-\mu\mathbb{1})}$
- Always stationary $\forall q$

Properties

- For $\mathrm{MA}(q)$, $acf(h) = \begin{cases} \frac{\sum_{i=0}^{q-h} \theta_i \theta_{i+h}}{\sum_{i=0}^{q} \theta_i^2}, & h \le q \\ 0, & h > q \end{cases}$
- For $\mathrm{MA}(1)$, $pacf(h) = \frac{(-\theta)^h (1-\theta^2)}{1-\theta^{2(h+1)}}$, $h \ge 1$

Moving Average $\mathrm{MA}(1)$: $Y_t - \mu = Z_t + \theta Z_{t-1}$

- Likelihood $= \left(\frac{1}{\sqrt{2\pi}\sigma}\right)^n e^{-\frac{S(\mu,\theta)}{2\sigma^2}}$
- $S(\mu,\theta) = (y_1 - \mu)^2 + \sum_{t=2}^n (y_t - \mu(1 - \theta + \cdots + (-1)^{t-1}\theta^{t-1}) - \theta y_{t-1} + \theta^2 y_{t-2} - \cdots + (-1)^{t-1}\theta^{t-1} y_1)^2$
- Take $\theta, \mu, \log\sigma \sim \mathrm{Uniform}(-C, C)$ independent.
- Restrict θ to $[-1,1]$ for identifiability.

$$f_{\mu,\theta|\text{data}}(\mu,\sigma) \propto \left(\frac{1}{S(\mu,\theta)}\right)^{\frac{n}{2}} I\{-1 < \theta < 1, -C < \mu < C\}$$

Alternate Forms:

- $Z_t = -\mu(1 - \theta + \cdots + (-1)^{t-1}\theta^{t-1}) + Y_t - \theta Y_{t-1} + \cdots + (-1)^{t-1}\theta^{t-1} Y_1 + (-1)^t \theta^t Z_0$
- $Y_t = Z_t + \mu(1 - \theta + \cdots + (-1)^{t-1}\theta^{t-1}) + \theta Y_{t-1} - \cdots - (-1)^{t-1}\theta^{t-1} Y_1 - (-1)^t \theta^t Z_0$

Estimation and Uncertainty:

- $Y_t | Y_1 = y_1, \dots, Y_{t-1} = y_{t-1} \sim N\left(\mu', {\sigma'}^2\right)$
- $\mu' = \mu(1 - \theta + \cdots + (-1)^{t-1}\theta^{t-1}) + \theta y_{t-1} - \cdots - (-1)^{t-1}\theta^{t-1} y_1$
- $\sigma' = \sigma$
- $f_{Y_1,\dots,Y_n}(y_1, \dots, y_n) = f_{Y_1}(y_1) \cdot f_{Y_2|Y_1=y_1}(y_2) \cdot f_{Y_3|Y_1=y_1,Y_2=y_2}(y_3) \cdot \dots \cdot f_{Y_n|Y_1=y_1,\dots,Y_{n-1}=y_{n-1}}(y_n)$
- $S(\mu,\theta) = (y_1 - \mu)^2 + \sum_{t=2}^n (y_t - \mu(1 - \theta + \theta^2 - \cdots + (-1)^{t-1}\theta^{t-1}) - \theta y_{t-1} + \theta^2 y_{t-2} - \cdots + (-1)^{t-1}\theta^{t-1} y_1)^2$
- $f_{\mu,\theta|\text{data}}(\mu,\theta) \propto \left(\frac{1}{S(\mu,\theta)}\right)^{\frac{n}{2}}$
- $\mu, \theta | \text{data} \sim t_{n-2,2}\left((\hat\mu, \hat\theta), \frac{S(\hat\mu,\hat\theta)}{n-2}\left(\frac{1}{2}\nabla^2 S(\hat\mu,\hat\theta)\right)^{-1}\right)$

Prediction:

- $Y_{n+1} | Y_1 = y_1, \dots, Y_n = y_n, \theta = \hat\theta, \mu = \hat\mu, \sigma = \hat\sigma \sim N\left(\hat\mu + \hat\theta \hat Z_n, \hat\sigma^2\right)$
- $\hat Z_k = -\hat\mu\left(1 - \hat\theta + \cdots + (-1)^{k-1}\hat\theta^{k-1}\right) + y_k - \hat\theta y_{k-1} + \cdots + (-1)^{k-1}\theta^{k-1} y_1$
- $\hat Z_n = -\hat\mu\left(1 - \hat\theta + \cdots + (-1)^{n-1}\hat\theta^{n-1}\right) + y_n - \hat\theta y_{n-1} + \cdots + (-1)^{n-1}\theta^{n-1} y_1$
- $Y_{n+i} | Y_1 = y_1, \dots, Y_n = y_n, \theta = \hat\theta, \mu = \hat\mu, \sigma = \hat\sigma \sim N\left(\hat\mu, \hat\sigma^2(1 + \hat\theta^2)\right)$

Autoregressive Integrated Moving Average (ARIMA) Models

$\mathrm{ARMA}(p,q)$ Models: $\phi(B)X_t = \mu + \theta(B)Z_t$

  - $\phi(B)X_t = \mu + \theta(B)Z_t$, $Z_i \sim N(0,\sigma^2)$
  - $\phi(z) = 1 - \phi_1 z - \phi_2 z^2 - \cdots - \phi_p z^p$
  - $\theta(B) = 1 + \theta_1 z + \theta_2 z^2 + \cdots + \theta_q z^q$
  - $X_t - \phi_1 X_{t-1} - \cdots - \phi_p X_{t-p} = \mu + Z_t + \theta_1 Z_{t-1} + \cdots + \theta_q Z_{t-q}$
- Causal if all roots of $\phi(z)$ have modulus strictly greater than 1 i.e. can write $X_t = \sum_{i=0}^{\infty} \psi_i Z_{t-i}$
- $\mathrm{Cov}[X_t, X_{t+h}] = \sigma^2 \sum_{i=0}^{\infty} \psi_i \psi_{i+h}$
- $acf(h) = \frac{\sum_{i=0}^{\infty} \psi_i \psi_{i+h}}{\sum_{i=0}^{\infty} \psi_i^2}$
- In general, obtain ACF by solving the difference equation $\rho(h) - \phi_1 \rho(h-1) - \cdots - \phi_p \rho(h-p)$, $h \geq \max(p, q+1)$ with initial conditions

Properties

- A doubly infinite sequence of RVs $\{X_t\}_{t=-\infty}^{\infty}$ is stationary if $\mathbb{E}[X_t]$ is constant and $\mathrm{Cov}[X_t, X_{t+h}]$ only depends on h
- An $\mathrm{ARMA}(p,q)$ process is causal only when the roots of $\phi(z)$ lie outside the unit circle
- $\{X_t\}_t \sim \mathrm{ARMA}(p,q)$ is causal if X_t can be expressed as $\sum_{i=0}^{\infty} \psi_i Z_{t-i} = \psi(B)Z_t$ for $\{\psi_i\}_{i=0}^{\infty}$ satisfying $\sum_{i=0}^{\infty} |\psi_i| < \infty$, $\psi_0 = 1$ i.e. it does not depend on the future.
- An $\mathrm{ARMA}(p,q)$ process is invertible if and only if $\theta(z) \neq 0$ for $|z| \leq 1$ i.e. all roots must lie outside the unit circle
- $\{X_t\}_t \sim \mathrm{ARMA}(p,q)$ process is invertible if X_t can be written as $\pi(B)X_t = Z_t$ for $\pi(B) = \sum_{j=0}^{\infty} \pi_j B^j$, $\sum_{j=0}^{\infty} |\pi_j| < \infty$ and $\pi_0 = 1$
- $\phi(B)\psi(B) = \theta(B)$
- $\theta(B)\pi(B) = \phi(B)$
- Just match the coefficients

ARIMA Models $\mathrm{ARIMA}(p,d,q)$

- A process $\{Y_t\}_t$ is $\mathrm{ARIMA}(p,d,q)$ if $\{X_t\}_t$ is $\mathrm{ARMA}(p,q)$, where $X_t = \nabla^d Y_t$
- $\phi(B)(X_t - \mu) = \theta(B)Z_t$, $Z_t \sim N(0,\sigma^2)$

Seasonal ARMA Models $\mathrm{ARMA}(P,Q)_s$

- $\Phi(B^s) = 1 - \Phi_1 B^s - \Phi_2 B^s - \cdots - \Phi_P B^{Ps}$
- $\Theta(B^s) = 1 + \Theta_1 B^s + \Theta_2 B^{2s} + \cdots + \Theta_Q B^{Qs}$
- $P + Q + 1$ parameters; special cases of $\mathrm{ARMA}(Ps, Qs)$ model – sparser model
- ACF and PACF are non-zero only at seasonal lags $h = 0, s, 2s, 3s, \ldots$
- At seasonal lags, the ACF and PACF of the models behave just like the normal ARMA model $\Phi(B)X_t = \Theta(B)Z_t$

Multiplicative Seasonal ARMA Models

- $\Phi(B^s)\phi(B)X_t = \Theta(B^s)\theta(B)Z_t$
- Write $\mathrm{ARMA}(p,q) \times (P,Q)_s$
- $\Phi(z) = 1 - \Phi_1 z^s - \Phi_2 z^{2s} - \cdots - \Phi_p z^{ps}$
- $\Theta(z) = 1 + \Theta_1 z^s + \Theta_2 z^{2s} + \cdots + \Theta_q z^{qs}$

Seasonal ARIMA Models (SARIMA)

- $\mathrm{ARIMA}(p,d,q) \times (P,D,Q)_S$
- $\Phi(B^s)\phi(B)\nabla_s^D \nabla^d X_t = \Theta(B^s)\theta(B)Z_t$

Box-Jenkins Method

Box-Jenkins Method

Set-up:

- Goal: apply $\text{ARMA}(p,q)$ or $\text{ARIMA}(p,d,q)$ models to time series
- Pre-process the data $y_1,\dots,y_n$ to transform it into $x_1,\dots,x_n$ which does not have any discernible trends.
- Fit an $\text{ARMA}(p,q)$ model for appropriate p,q to the transformed data x_t

Phase I – Pre-processing:

- (1) Parametric pre-processing: fit a parametric function f of t to $y_1,\dots,y_n$, then obtain residuals $x_i = y_i - f(t)$
- Use linear regression or frequency to remove linear and sinusoidal trends
- Differencing $\nabla y_t = y_t - y_{t-1}$
- $\nabla^2 y_t = y_t - 2y_{t-1} + y_{t-2}$
- If you take k order difference, you get time series of $t-k$.
- Differencing eliminates increasing decreasing trends.
- Seasonal differencing used to eliminate periodic trends. E.g. if have seasonal trend of s, do $\nabla_s y_t = y_t - y_{t-s}$: time series of length $n-s$

Phase II – Fit an ARIMA model:

- Use PACF and ACF to choose
- OR brute force all combinations of (p,q)
- Remember to transform your prediction back to original data

Final Checks

- Do not forget μ and σ^2 when counting parameters, especially μ in models.

Splines $Y_t = f(t) + \epsilon_t$, f smooth

- Way of pre-processing i.e. fitting a parametric function of t to $y_1,\dots,y_n$
- $f(t) = \beta_0 + \beta_1 t + \beta_2 (t-s_1)_+ + \beta_3 (t-s_2)_+ + \cdots + \beta_{k+1}(t-s_k)_+$
- $f(t) = \beta_0 + \beta_1 t + \beta_2 (t-2)_+ + \beta_3 (t-3)_+ + \cdots + \beta_{n-1}\big(t-(n-1)\big)_+$

Prior:

- $\beta_0, \beta_1 \sim N(0, C)$
- $\beta_2, \dots, \beta_{n-1} \sim N(0, \tau^2)$
- More flexible than linear regression since $\tau = \sqrt{C}$ reduces to uniform prior.

$$\beta \sim N\left(\begin{bmatrix}0\\0\\\vdots\\0\\0\end{bmatrix}, \begin{bmatrix}C & 0 & 0 & \cdots & 0\\ 0 & C & 0 & \cdots & 0\\ 0 & 0 & \tau^2 & \cdots & 0\\ \vdots & \vdots & \vdots & \ddots & \vdots\\ 0 & 0 & 0 & \cdots & \tau^2\end{bmatrix}\right)$$

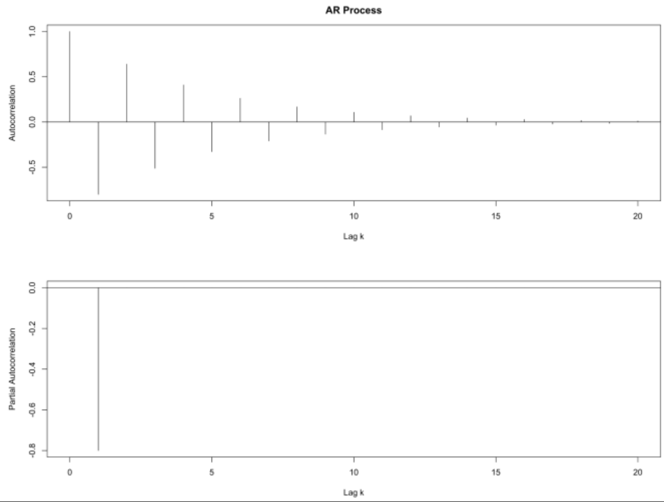
Fact:

- $\beta \sim N_p(m_0, Q_0)$, $Y|\beta \sim N_n(X\beta, \sigma^2 \mathbb{I}_n)$ means $\beta|Y \sim N_p(m_1, Q_1)$
- $m_1 = \left(Q_0^{-1} + \frac{1}{\sigma^2}X^TX\right)^{-1}\left(Q_0^{-1}m_0 + \frac{1}{\sigma^2}X^TY\right)$
- $Q_1 = \left(Q_0^{-1} + \frac{1}{\sigma^2}X^TX\right)^{-1}$
- $\beta|\text{data},\sigma \sim N\left(\left(Q_0^{-1} + \frac{1}{\sigma^2}X^TX\right)^{-1}\frac{1}{\sigma^2}X^TY, \left(Q_0^{-1} + \frac{1}{\sigma^2}X^TX\right)^{-1}\right)$
- If want smooth function fit, take small value of τ.
- $f_{\text{data}|\tau,\sigma}(\text{data}) = \left(\frac{1}{\sqrt{2\pi}}\right)^n (\det\Sigma)^{-\frac{1}{2}} e^{-\frac{1}{2}Y^T\Sigma^{-1}Y}$
- $\Sigma = XQ_0X^T + \sigma^2\mathbb{I}_n$
- $\log\tau, \log\sigma \sim \text{Uniform}(-C, C), C = 10^6$

Graphs

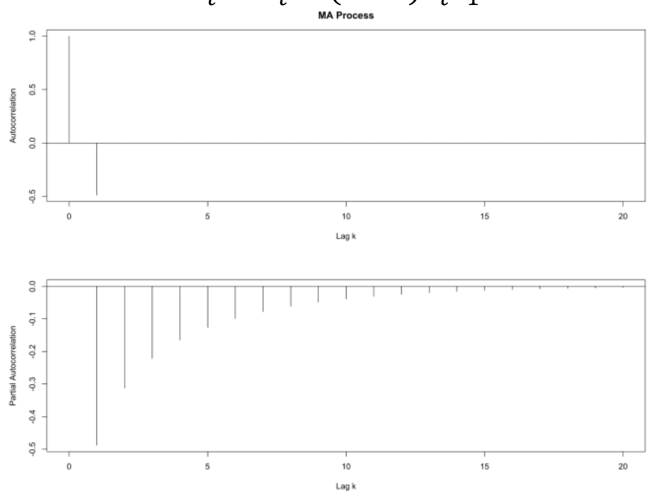
AR(1)

$$X_t + 0.8X_{t-1} = Z_t$$



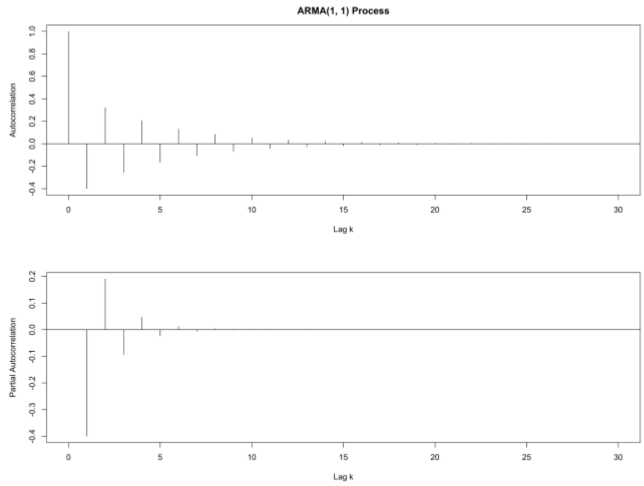
MA(1)

$$X_t = Z_t + (-0.8)Z_{t-1}$$



ARMA(1,1)

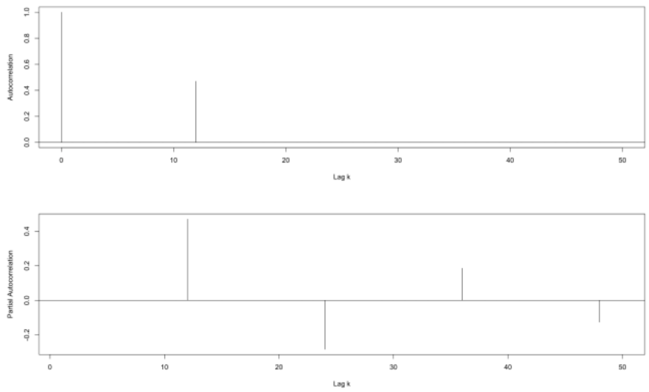
$$X_t + 0.8X_{t-1} = Z_t + 0.5Z_{t-1}$$



Seasonal MA(1)

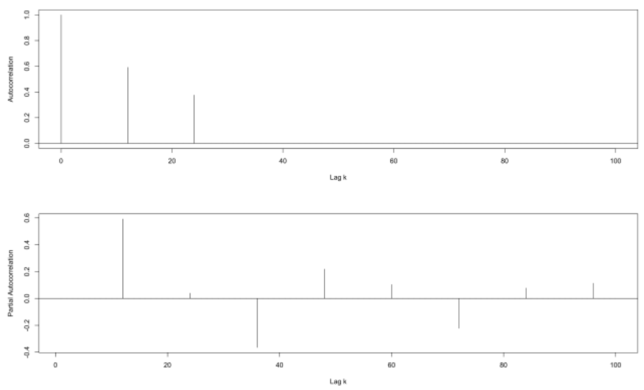
Large negative autocorrelation at lag 1 followed by some small autocorrelation coefficients, then large again at 11,12,13

$$X_t = Z_t + \Theta_1 Z_{t-12}$$



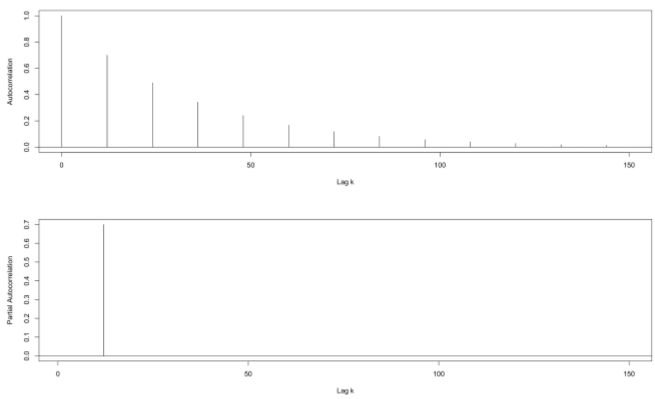
Seasonal MA(2)

$$X_t = Z_t + \Theta_1 Z_{t-12} + \Theta_2 Z_{t-24}$$



Seasonal AR(1)

$$X_t - \Phi X_{t-12} = Z_t$$



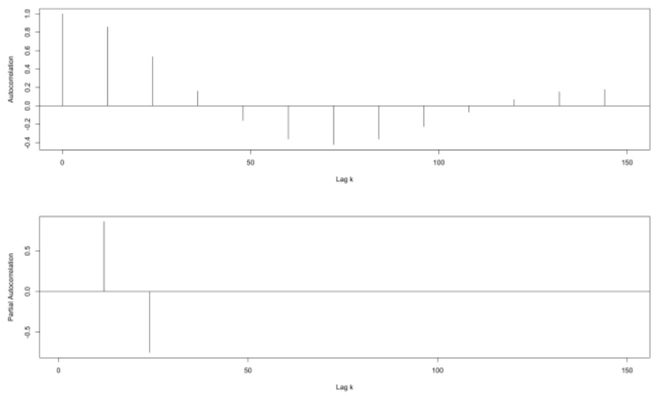
Seasonal AR(2)

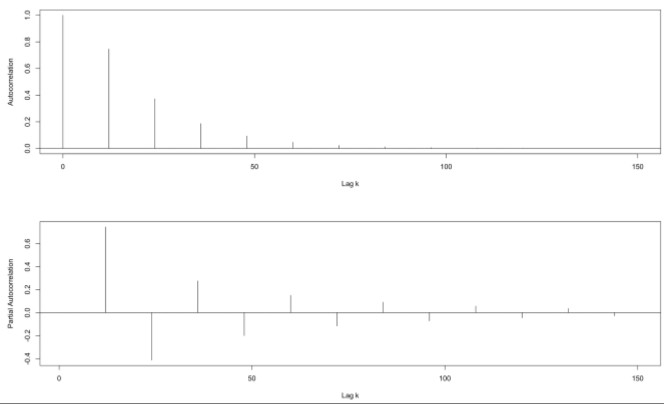
$$X_t - \Phi_1 X_{t-12} - \Phi_2 X_{t-24} = Z_t$$

Seasonal ARMA(1,1)

$$X_t - \Phi_1 X_{t-12} = Z_t + \Theta Z_{t-12}$$

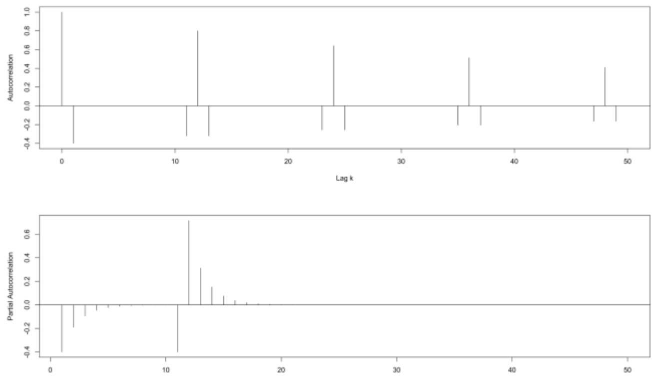
ACF and PACF of seasonal ARMA models just look like those of a usual ARMA models at seasonal lags.





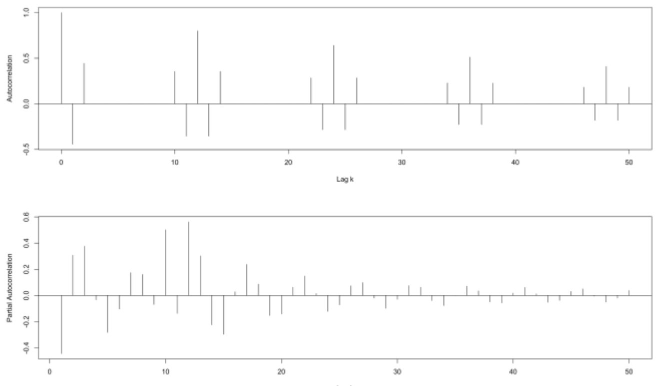
$\text{ARMA}(0,1) \times (1,0)_{12}$

$$X_t - \Phi_1 X_{t-12} = Z_t + \theta_1 Z_{t-1}$$



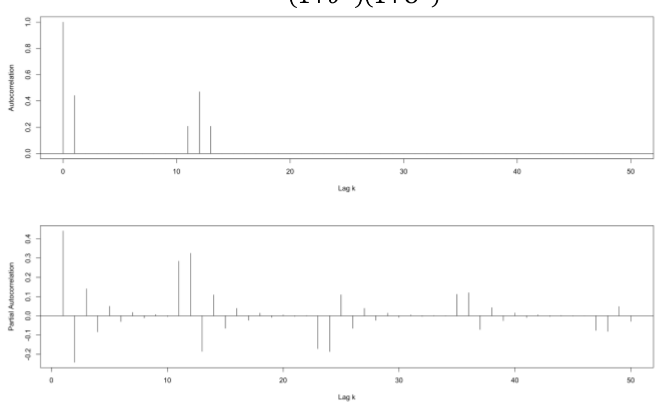
$\text{ARMA}(0,2) \times (1,0)_{12}$

$$X_t - \Phi_1 X_{t-12} = Z_t + \theta Z_{t-1} + \theta_2 Z_{t-2}$$



$\text{ARMA}(0,1) \times (0,1)_S$

- $X_t = (\mathbb{I} + \theta B)(\mathbb{I} + \Theta B^S) Z_t$
- $X_t = Z_t + \theta Z_{t-1} + \Theta Z_{t-S} + \theta\Theta Z_{t-S-1}$
- $acf(0) = 1$
- $acf(1) = \frac{\theta}{1+\theta^2}$
- $acf(S-1) = \frac{\theta\Theta}{(1+\theta^2)(1+\Theta^2)}$
- $acf(S) = \frac{\Theta}{1+\Theta^2}$
- $acf(S+1) = \frac{\theta\Theta}{(1+\theta^2)(1+\Theta^2)}$



State Space Models

Linear Least Squares Estimator (LLSE)

- $\mathbb{L}[Y|X_1,\dots,X_p] = \beta_0^* + \beta_1^* X_1 + \cdots + \beta_p^* X_p$
- $\mathbb{L}[Y|X_1,\dots,X_p] = \min_\beta \mathbb{E}\left[\left(Y - (\beta_0 + \beta_1 X_1 + \cdots + \beta_p X_p)\right)^2\right]$
- $\mathbb{E}\left[Y - \mathbb{L}[Y|X_1,\dots,X_p]\right] = 0$ (unbiased)
- $\mathbb{E}[X_i(Y - \mathbb{L}[Y|X_1,\dots,X_p])] = 0\ \forall i$ (uncorrelated)
- $\beta^* = \begin{bmatrix}\beta_1^* \\ \vdots \\ \beta_p^*\end{bmatrix} = \mathrm{Cov}[X]^{-1}\mathrm{Cov}[X,Y]$
- $\beta_0^* = \mathbb{E}[Y] - \mathrm{Cov}[Y,X]\mathrm{Cov}[X]^{-1}\mathbb{E}[X]$
- $\mathbb{L}[Y|X_1,\dots,X_p] = \mathbb{E}[Y] + \mathrm{Cov}[Y,X]\mathrm{Cov}[X]^{-1}(X - \mathbb{E}[X])$
- $[p=1]\ Y = \mathbb{E}[Y] + \rho_{X,Y}\sqrt{\frac{\mathrm{Var}[Y]}{\mathrm{Var}[X]}}(X - \mathbb{E}[X])$
- $r_{Y|X_1,\dots,X_p} = Y - \mathbb{L}[Y|X_1,\dots,X_p] = (Y - \mathbb{E}[Y]) - \mathrm{Cov}[Y,X]\mathrm{Cov}[X]^{-1}(X - \mathbb{E}[X])$
- $\mathrm{Var}\left[r_{Y|X_1,\dots,X_p}\right] = \mathrm{Var}[Y] - \mathrm{Cov}[Y,X]\mathrm{Cov}[X]^{-1}\mathrm{Cov}[X,Y]$
- Let $\Sigma = \mathrm{Cov}\begin{bmatrix}X_1 \\ \vdots \\ X_p \\ Y_1\end{bmatrix}$, then $\mathrm{Var}\left[r_{Y|X_1,\dots,X_p}\right] = Y_1^S$
- To prove best linear prediction, suffices to show unbiasedness and uncorrelatedness.

Partial Autocorrelation

- Measures degree of association between two random variables, with the effect of a set of controlling variables removed.
- $\rho_{Y_1,Y_2|X_1,\dots,X_p} = \mathrm{Corr}\left[r_{Y_1|X_1,\dots,X_p}, r_{Y_2|X_1,\dots,X_p}\right]$
- $\mathrm{Cov}\left[r_{Y_1|X_1,\dots,X_p}, r_{Y_2|X_1,\dots,X_p}\right] = \mathrm{Cov}[Y_1,Y_2] - \mathrm{Cov}[Y_1,X](\mathrm{Cov}[X])^{-1}\mathrm{Cov}[X,Y_2]$
- $\rho_{Y_1,Y_2|X_1,\dots,X_p} = \frac{\mathrm{Cov}[Y_1,Y_2] - \mathrm{Cov}[Y_1,X](\mathrm{Cov}[X])^{-1}\mathrm{Cov}[X,Y_2]}{\sqrt{\mathrm{Var}[Y_1] - \mathrm{Cov}[Y_1,X]\mathrm{Cov}[X]^{-1}\mathrm{Cov}[X,Y_1]}\sqrt{\mathrm{Var}[Y_2] - \mathrm{Cov}[Y_2,X]\mathrm{Cov}[X]^{-1}\mathrm{Cov}[X,Y_2]}}$
- $[p=1]\ \rho_{Y_1,Y_2|X} = \frac{\rho_{Y_1,Y_2} - \rho_{Y_1,X}\rho_{Y_2,X}}{\sqrt{1-\rho_{Y_1,X}^2}\sqrt{1-\rho_{Y_2,X}^2}}$
- $R_{Y_1,Y_2|X_1,\dots,X_p} = \begin{bmatrix} r_{Y_1|X_1,\dots,X_p} \\ r_{Y_2|X_1,\dots,X_p}\end{bmatrix}$
- $\mathrm{Cov}\left[R_{Y_1,Y_2|X_1,\dots,X_p}\right] = \mathrm{Cov}[Y] - \mathrm{Cov}[Y,X](\mathrm{Cov}[X])^{-1}\mathrm{Cov}[X,Y]$ where $Y = \begin{bmatrix}Y_1 \\ Y_2\end{bmatrix}$ and $X = \begin{bmatrix}X_1 \\ \vdots \\ X_p\end{bmatrix}$
- $\mathrm{Cov}[Y] - \mathrm{Cov}[Y,X](\mathrm{Cov}[X])^{-1}\mathrm{Cov}[X,Y]$ is the Schur complement of $\mathrm{Cov}[Y]$ in $\mathrm{Cov}\begin{bmatrix}X \\ Y\end{bmatrix} = \Sigma$
- $\rho_{Y_1,Y_2|X_1,\dots,X_p} = \frac{-\Sigma_{n-1,n}^{-1}}{\sqrt{\Sigma_{n-1,n-1}^{-1}\Sigma_{n,n}^{-1}}}$
- If $Y_1,\dots,Y_n$ are random variables with $\mathrm{Cov}[Y] = \Sigma$, then $\rho_{Y_i,Y_j|Y_k,k\neq i,j} = \frac{-\Sigma_{i,j}^{-1}}{\sqrt{\Sigma_{i,i}^{-1}\Sigma_{j,j}^{-1}}}$
    - $\Sigma_{i,j}^{-1} = 0 \Leftrightarrow \rho_{Y_i,Y_j|Y_k,k\neq i,j} = 0$
    - $\Sigma_{i,j}^{-1} < 0 \Leftrightarrow \rho_{Y_i,Y_j|Y_k,k\neq i,j} > 0$
    - $\Sigma_{i,j}^{-1} > 0 \Leftrightarrow \rho_{Y_i,Y_j|Y_k,k\neq i,j} < 0$
- If $\mathbb{L}[Y|X_1,\dots,X_p] = \beta_0^* + \beta_1^* X_1 + \cdots + \beta_p^* X_p$, then $\beta_i^* = \rho_{Y,X_i|X_k,k\neq i}\sqrt{\frac{\mathrm{Var}\left[r_{Y|X_k,k\neq i}\right]}{\mathrm{Var}\left[r_{X_i|X_k,k\neq i}\right]}}$

- $[p = 1]$ $\beta_1^* = \rho_{Y,X_1}\sqrt{\frac{\mathrm{Var}[Y]}{\mathrm{Var}[X_1]}}$
- $\beta_i^* = 0 \Leftrightarrow \rho_{Y,X_i|X_k,k\neq i} = 0$

Schur's Complement

- $A = \begin{bmatrix} E & F \\ G & H \end{bmatrix}$
- $E^S = E - FH^{-1}G$
- $H^S = H - GE^{-1}F$
- $\det(A) = \det(E)\det(H^S) = \det(H)\det(E^S)$
- If A PSD, then so is E, E^S, H and H^S.
- $A^{-1} = \begin{bmatrix} (E^S)^{-1} & -E^{-1}F(H^S)^{-1} \\ -(H^S)^{-1}GE^{-1} & (H^S)^{-1} \end{bmatrix}$

Partial Autocorrelation Function (PACF)

- Let $\{Y_t\}$ be a stationary process.
- $pacf(h) \coloneqq \rho_{Y_t,Y_{t-h}|Y_{t-1},\dots,Y_{t-h+1}}$
- Note that $\rho_{Y_t,Y_{t-h}|Y_{t-1},\dots,Y_{t-h+1}}$ does not depend on t since $\{Y_t\}$ stationary.
- Alternative definition is $pacf(h)$ is the coefficient of Y_{t-h} in $\mathbb{L}[Y_t|Y_{t-1},\dots,Y_{t-h}]$
    - Use this to get sample partial autocorrelation.
    - Estimate $\mathrm{Cov}[X_{t-i}, X_{t-j}]$ for $(i,j) \in \{0,1,\dots,h\}^2$.
    - Find coefficient of X_{t-h} in the best linear predictor.
- $\beta_h^* = pacf(h)$ i.e. $pacf(h)$ is the coefficient of X_{t-h} in $\mathbb{L}[X_t|X_{t-1},\dots,X_{t-h}]$
    - $\mathrm{Var}[r_{Y_t|Y_{t-1},\dots,Y_{t-h+1}}] = \mathrm{Var}[r_{Y_{t-h}|Y_{t-1},\dots,Y_{t-h+1}}]$ since covariance matrix of $\begin{bmatrix} Y_t \\ \vdots \\ Y_{t-h+1} \end{bmatrix}$ is the same as $\begin{bmatrix} Y_{t-h} \\ \vdots \\ Y_{t-1} \end{bmatrix}$ due to stationarity.
- For causal stationary $\mathrm{AR}(p)$ model, $pacf(h) = \begin{cases} \phi_p, & h = p \\ 0, & h > p \end{cases}$. For $h < p$, it is an expression involving $\phi_1, \dots, \phi_p$.