ARIMA models (3.6 - 3.9)

$\boxed{\text{Definition 3.11}}$ A process X_t is said to be ARIMA(p,d,q) if

$$\nabla^d X_t = (1-B)^d X_t \quad \text{is ARMA}(p,q).$$

We will write as $\phi(B)(1-B)^d X_t = \Theta(B) W_t$

If $E(\nabla^d X_t) = \mu$, $\quad \phi(B)(1-B)^d X_t = \delta + \Theta(B) W_t, \quad \delta = \mu(1-\phi_1-\dots-\phi_p)$.

Let $y_t = \nabla^d X_t$, the forecasts of y_t can lead to the forecasts of X_t

For example $y_t = \nabla X_t = X_t - X_{t-1}$, then we have

$$E(y_{n+m} \mid X_n, X_{n-1}, \dots, X_1) = E(y_{n+m} \mid y_n, y_{n-1}, \dots, y_2, X_1) \overset{\text{let}}{=} y^n_{n+m}$$

And hence $y^n_{n+m} = E(X_{n+m} \mid X_n, \dots, X_1) - E(X_{n+m-1} \mid X_n, \dots, X_1)$

$$= X^n_{n+m} - X^n_{n+m-1}$$

$$\Rightarrow X^n_{n+m} = y^n_{n+m} + X^n_{n+m-1} \quad \text{with initial condition } X^n_{n+1} = y^n_{n+1} + X_n$$

As in (3.86), we approximate $Var(X_{n+m} - X^n_{n+m})$ by $Var(X_{n+m} - \tilde{X}^n_{n+m})$

$= \sigma_w^2 \sum_{j=0}^{m-1} \psi_j^{*2}$, where $\psi^*(z) = \Theta(z)/[\phi(z)(1-z)^d] = \sum_{j=0}^{\infty} \psi_j^* z^j$

$\boxed{\text{Example 3.37}}$ ARIMA$(0,1,0)$, ie $\phi(B)=1$, $d=1$, $\Theta(B)=1$

$$\Rightarrow \phi(B)(1-B) X_t = \delta + \Theta(B) W_t$$

$$\Rightarrow y_t = X_t - X_{t-1} = \delta + W_t \text{ which is a random walk with drift}$$

For $W_t \sim \text{iid } N(0, \sigma_w^2)$ $\quad y^n_{n+m} = E(\delta + W_{n+m} \mid y_n, y_{n-1}, \dots, y_2, X_1) = \delta$

$$\therefore X^n_{n+m} = X^n_{n+m-1} + y^n_{n+m} = X^n_{n+m-1} + \delta = X_n + m\delta$$

m-step-ahead prediction error is given by

$$P^n_{n+m} = \sigma_w^2 \sum_{j=0}^{m-1} \psi_j^{*2} = m\sigma_w^2$$

$$\because \psi^*(z) = \Theta(z)/[\phi(z)(1-z)^d] = \frac{1}{1-z} = \sum_{j=0} z^j \Rightarrow \psi_j^* = 1 \quad \forall j$$

Example 3.38 A frequently used, and abused, forecasting method called exponentially weighted moving averages (EWMA) is of the form

$$\tilde{X}_{n+1}^{n} = (1-\lambda)X_n + \lambda \tilde{X}_n^{n-1},$$

which only require us to retain the previous forecast value and the current observation to forecast the next time period. It is indeed based on ARIMA(0,1,1) model : $X_t = X_{t-1} + W_t - \lambda W_{t-1}$, $|\lambda|<1$

Let $y_t = X_t - X_{t-1} = (1-B)X_t = (1-\lambda B)W_t$, a MA(1) model.

For MA model, we consider $\tilde{y}_{n+m} = E(y_{n+m} \mid y_n, y_{n-1}, \dots)$

$$W_t = \lambda W_{t-1} + y_t = \lambda^2 W_{t-2} + \lambda y_{t-1} + y_t = \sum_{j=1}^{\infty} \lambda^j y_{t-j} + y_t$$

Note that $\tilde{y}_{n+m} = 0$ for $m \geq 2$

$$\tilde{y}_{n+1} = \sum_{j=1}^{\infty} -\lambda^j y_{n+1-j} \Rightarrow \tilde{X}_{n+1} - X_n = \sum_{j=1}^{\infty} (-\lambda^j)(X_{n+1-j} - X_{n-j})$$

$$= \sum_{j=1}^{\infty} (-\lambda^j) X_{n+1-j} + \sum_{j=1}^{\infty} \lambda^j X_{n-j}$$

$$= -\lambda X_n + \sum_{j=2}^{\infty} (-\lambda^j) X_{n+1-j} + \sum_{j=2}^{\infty} \lambda^{j-1} X_{n+1-j}$$

$$\Rightarrow \tilde{X}_{n+1} = (1-\lambda)X_n + \sum_{j=2}^{\infty} \lambda^{j-1}(1-\lambda) X_{n+1-j} = \sum_{j=1}^{\infty} (1-\lambda)\lambda^{j-1} X_{n+1-j}$$

or $$= (1-\lambda)X_n + \lambda \sum_{j=1}^{\infty} \lambda^{j-1}(1-\lambda) X_{n-j} = (1-\lambda)X_n + \lambda \tilde{X}_n$$

Recall that $\tilde{X}_{n+1}^{n}$ is formed by setting $\sum_{j=n+1}^{\infty} \pi_j X_{n+1-j} = 0$ $\quad (\pi_j = (1-\lambda)\lambda^{j-1})$

$\therefore$ We also have $\tilde{X}_{n+1}^{n} = (1-\lambda)X_n + \lambda \tilde{X}_n^{n-1}$

Now $$(1-B)X_t = (1-\lambda B)W_t \Rightarrow X_t = \psi(B)W_t = \sum_{j=0}^{\infty} \psi_j^* W_t$$

with $$\psi(z) = \frac{1-\lambda z}{1-z} = 1 + (1-\lambda)\frac{z}{1-z} = 1 + (1-\lambda)\sum_{j=1}^{\infty} z^j \quad \text{for } |z|<1$$

$\therefore \psi_0^* = 1 \qquad \psi_j^* = 1-\lambda$ for $j \geq 1$

$$\therefore P_{n+m}^{n} \approx E(X_{n+m} - \tilde{X}_{n+m}) = \sigma_w^2 \sum_{j=0}^{m-1} \psi_j^{*2} = \sigma_w^2 \left[1 + (m-1)(1-\lambda)^2\right]$$

Building ARIMA models

STEP 1: Plotting the data

STEP 2: Possibly transforming the data

See if the time series looks stationary. If not, we need to first transform the data. For example,

(i) increasing variance (e.g. $X_t = 2X_{t-1}$), we may take log (e.g. $y_t = \log X_t = \log 2 + \log X$

(ii) increasing trend (e.g. $X_t = t + W_t$), we may apply detrending (e.g. fitting a linear regression of X_t on $\alpha + \beta t$, and then consider $y_t = X_t - \hat{\alpha} - \hat{\beta} t$)

(iii) Random walk, we may apply differencing.
If differencing is called for, then start from $d=1$, and inspect the time plot of ∇X_t. If additional differencing is necessary, then try differencing again.
Avoid overdifferencing. For example, $X_t = W_t$ is uncorrelated by
$\nabla X_t = W_t - W_{t-1}$ is MA(1).

STEP 3: Identifying the dependence orders of the model (p,d,q)

Another way to determine d is by plotting the ACFs. Recall (3.50) for AR(p) model, $\rho(h) = z_1^{-h} P_1(h) + z_2^{-h} P_2(h) + \dots + z_r^{-h} P_r(h)$, $h \geq p$
where z_i's are the roots of $\phi(z) = 0$.
If differencing is needed, then $z_i = 1$ for some i.
$\Rightarrow$ $\hat{\rho}(h)$ will not decay to zero fast as h increases

After d is selected, we can check the sample ACF and PACF of $\nabla^d X_t$ to select a set of (p,d,q) candidates.

STEP 4 : Parameter estimation

STEP 5 : Diagnostics

STEP 6 : Model choice

For each candidate (p, d, q), if the model is correct, then

$$X_{t+1} - \tilde{X}_{t+1} \sim N(0, P_{t+1}^{t}) \quad \text{if} \quad w_t \sim \text{iid}\, N(0, \sigma_w^2)$$

The normal distribution for w_t is useful for constructing prediction interval for forecasts. To check this assumption, we construct

$$e_t = \frac{X_t - \hat{X}_t^{t-1}}{\hat{P}_t^{t-1}},$$

where $\hat{X}_t^{t-1}$ and $\hat{P}_t^{t-1}$ are the estimate of $\tilde{X}_t^{t-1}$ (or $\tilde{X}_t$) and P_t^{t-1} with estimated parameters $\hat{\phi}, \hat{\theta}$ and $\hat{\sigma}_w^2$.

To check the normality, we can plot a Q-Q plot with normal distribution.

To check the independence, beside checking if the sample ACF $\hat{\rho}_e(h)$ are small in magnitude, we can apply the Ljung-Box test to test $H_0: \rho_e(1) = \cdots = \rho_e(H) = 0$ (Typically we choose $H = 20$)

The test statistic is $Q = n(n+2)\sum_{h=1}^{H} \frac{\hat{\rho}_e^2(h)}{n-h} \approx n \sum_{h=1}^{n} \hat{\rho}_e^2(h)$

(Recall from Property 1.2 that $\hat{\rho}_X(h) \xrightarrow{d} N(0, \frac{1}{\sqrt{n}})$ if $X_t \sim \text{iid}(0, \sigma_w^2)$)

Under H_0, $Q \xrightarrow{d} \chi^2_{H-p-q}$. $\therefore$ H_0 is rejected if the observed value of Q exceeds the $(1-\alpha)$-quantile of χ^2_{H-p-q}.

Note that if you use the R function Box.test on $\{e_t\}$, the function will assume $p = q = 0$.

Finally, for the remaining promising (p, d, q) models, we pick the one with minimum AIC, AICc or BIC value.

Regression with autocorrelated errors

Consider the regression model $y_t = \sum_{j=1}^{r} \beta_j z_{tj} + X_t$, $t=1,\dots,n$, and $\{X_1,\dots,X_n\}$ are correlated. Let $\vec{y} = (y_1,\dots,y_n)^T$, $\vec{Z} = (z_{tj})_{1\le t\le n, 1\le j\le r}$, $\vec{\beta} = (\beta_1,\dots,\beta_r)^T$ and $\vec{X} = (X_1,\dots,X_n)^T$. If we know that $\vec{X} \sim N(0, \vec{P}) \Rightarrow \vec{P}^{-1/2}\vec{X} \sim N(0, I)$ then $\vec{y}^* = \vec{P}^{-1/2}\vec{y} = \vec{P}^{-1/2}(\vec{Z}\vec{\beta} + \vec{X}) = (\vec{P}^{-1/2}\vec{Z})\vec{\beta} + \vec{P}^{-1/2}\vec{X} = \vec{Z}^*\vec{\beta} + \vec{\delta}$, where $\vec{\delta} \sim N(0, I)$ and hence we can estimate $\vec{\beta}$ by the usual OLS estimator $\hat{\beta} = (\vec{Z}^{*T}\vec{Z}^*)^{-1}\vec{Z}^{*T}\vec{y}^* = (\vec{Z}^T\vec{P}^{-1}\vec{Z})^{-1}\vec{Z}^T\vec{P}^{-1}\vec{y}$

However, if we don't have prior information about $\vec{P}$, the estimate of $\vec{P}$ based on stationary assumption of $\vec{X}$ can be very poor.

A better way is to assume X_t follows ARMA(p,q) model that is invertible so that $\pi(B)X_t = W_t$, where $W_t \sim wn(0, \sigma_w^2)$. Then we have

$$y_t^* = \pi(B)y_t = \sum_{j=1}^{r}\beta_j \pi(B) z_{tj} + \pi(B)X_t = \sum_{j=1}^{r}\beta_j z_{tj}^* + W_t$$

Note that $\pi(B)$ involves parameters $\vec{\phi} = (\phi_1,\dots,\phi_p)$ and $\vec{\theta} = (\theta_1,\dots,\theta_q)$.

We estimate the parameters by

$$(\hat{\phi}, \hat{\theta}, \hat{\beta}) = \arg\min_{\phi,\theta,\beta} \sum_{t=1}^{n}\left(\pi(B)y_t - \sum_{j=1}^{r}\beta_j \pi(B) z_{tj}\right)^2 \quad \text{------} \quad (1)$$

The optimization can be done by R function "sarima"; see Example 3.44.

General procedure for fitting $y_t = \sum_{j=1}^{r}\beta_j z_{tj} + X_t$

(i) Compute $\hat{\beta}_j^0$ by OLS as usual to get $\hat{X}_t = y_t - \sum_{j=1}^{r}\hat{\beta}_j^0 z_{tj}$

(ii) Identify ARMA model (choosing p and q) for $\hat{X}_t$. Let $\hat{X}_t$ follow $\phi(B)\hat{X}_t = \theta(B)W_t \Rightarrow W_t = \pi(B)\hat{X}_t$. If $p=q=0$, output $\hat{\beta} = \hat{\beta}^0$

(iii) Compute $(\hat{\phi}, \hat{\theta}, \hat{\beta})$ by solving (1)

(iv) Inspect the residuals $\hat{W}_t$ for whiteness, and adjust the model if necessary.

A pure seasonal ARMA model, ARMA$(P,Q)_s$, takes the form

$$\Phi_P(B^s)X_t = \Theta_Q(B^s)w_t,$$

where $\Phi_P(B^s) = 1 - \Phi_1 B^s - \Phi_2 B^{2s} - \dots - \Phi_P B^{Ps}$

$\Theta_Q(B^s) = 1 + \Theta_1 B^s + \Theta_2 B^{2s} + \dots + \Theta_Q B^{Qs}$

$\boxed{\text{Example 3.46}}$ ARMA$(1,0)_{12}$ $\quad (1-\Phi_1 B^{12})X_t = w_t$

$$\Rightarrow \quad X_t = \Phi_1 X_{t-12} + w_t$$

A multiplicative seasonal ARMA model, ARMA$(p,q)\times(P,Q)_s$, takes the form

$$\Phi_P(B^s)\phi(B)X_t = \Theta_Q(B^s)\theta(B)w_t$$

$\boxed{\text{Example 3.47}}$ Consider an ARMA$(0,1)\times(1,0)_{12}$

$$(1-\Phi B^{12})(1)X_t = (1)(1+\theta B)w_t \qquad |\Phi|<1 \text{ and } |\theta|<1$$

$$\Rightarrow \quad X_t - \Phi X_{t-12} = w_t + \theta w_{t-1}$$

$$\gamma(0) = Var(X_t) = Var(\Phi X_{t-12} + w_t + \theta w_{t-1}) = \Phi^2\gamma(0) + \sigma_w^2 + \theta^2\sigma_w^2$$

$$\Rightarrow \quad \gamma(0) = \frac{1+\theta^2}{1-\Phi^2}\sigma_w^2$$

$$\gamma(1) = Cov(X_t X_{t-1}) = \Phi\, Cov(X_{t-12}, X_{t-1}) + Cov(w_t, X_{t-1}) + \theta\, Cov(w_{t-1}, X_{t-1})$$

$$= \Phi\gamma(11) + \theta\sigma_w^2$$

$$\gamma(h) = \Phi\gamma(h-12) \quad \text{for } h\geq 2$$

$\therefore$ For $h = 12m+n$, $m\geq 0$, $0\leq n\leq 12$, $\gamma(12m+n) = \Phi^m\gamma(n)$

For $2\leq n\leq 12$, $\gamma(n) = \Phi\gamma(n-12) = \Phi\gamma(12-n)$

$$\Rightarrow \quad \gamma(11) = \Phi\gamma(1)$$

$$\therefore \quad \gamma(1) = \Phi(\Phi\gamma(1)) + \theta\sigma_w^2 \Rightarrow \gamma(1) = \frac{\theta}{1-\Phi^2}\sigma_w^2$$

$$\rho(1) = \frac{\gamma(1)}{\gamma(0)} = \frac{\theta}{1+\theta^2}$$

$$\rho(12m) = \frac{\gamma(12m)}{\gamma(0)} = \frac{\Phi^m\gamma(0)}{\gamma(0)} = \Phi^m, \qquad m=0,1,2,\dots$$

$$\rho(12m-1) = \rho(12(m-1)+11) = \Phi^{m-1}\rho(11) = \Phi^m\rho(1) \quad \text{for } m=1,2,\dots$$

$$\rho(12m+1) = \Phi^m\rho(1) = \frac{\theta}{1+\theta^2}\Phi^m$$

(also true for m=0)

Otherwise, for $2\leq n\leq 10$, so that $12m+n$ is not of the form $12m$ or $12m\pm1$

$$\rho(12m+n) = \Phi^m\rho(n) = \Phi^m\Phi\rho(12-n) = \Phi^{m+1}\Phi\rho(n) = \Phi^\infty\rho(n) = 0$$

Consider the model $X_t = S_t + W_t$, $S_t = S_{t-12} + V_t$

where $W_t \sim wn(0, \sigma_w^2)$ $V_t \sim wn(0, \sigma_v^2)$

Here S_t is a random walk and hence X_t is non-stationary. We can handle it by differencing.

$$(1-B^{12})X_t = X_t - X_{t-12} = S_t - S_{t-12} + W_t - W_{t-12} = V_t + W_t - W_{t-12}$$

Seasonal differencing can be indicated when ACF decays slowly at multiples of some season s, but is negligible between the periods.

A seasonal difference of order D is defined as $\nabla_s^D X_t = (1-B^s)^D X_t$ $D = 1, 2, \ldots$ Typically $D=1$ is sufficient to obtain seasonal stationarity.

Definition 3.12 The multiplicative seasonal ARIMA model, or SARIMA is given by $\Phi_P(B^s)\phi(B)\nabla_s^D\nabla^d X_t = \delta + \Theta_Q(B^s)\theta(B)W_t$, $W_t \sim N(0, \sigma$

which is denoted as $ARIMA(p,d,q) \times (P,D,Q)_s$

Example 3.48 Consider $ARIMA(0,1,1) \times (0,1,1)_{12}$ model with $EX_t = 0$

$$(1)(1)(1-B^{12})(1-B)X_t = (1+\Theta B^{12})(1+\theta B)W_t$$

$$\Rightarrow (1 - B - B^{12} + B^{13})X_t = (1 + \theta B + \Theta B^{12} + \Theta\theta B^{13})W_t$$

$$\Rightarrow X_t = X_{t-1} + X_{t-12} - X_{t-13} + W_t + \theta W_{t-1} + \Theta W_{t-12} + \Theta\theta W_{t-13}$$

Example 3.49 Consider the R data set "AirPassengers"

STEPs 1 and 2 (Plotting and transforming)

plot (X_t) $\longrightarrow$ increasing trend and increasing variance

$lx = \log X_t$ (to handle increasing variance) $\longrightarrow$ stabilized variance

$dlx = \nabla \log X_t$ $\longrightarrow$ trend removed. Obvious persistence in the seasons ($\nabla \log X_t \approx \nabla \log X_{t-12}$)

$ddlx = \nabla_{12} \nabla \log X_t$ $\longrightarrow$ seems stationary

STEP 3: Choose p, q, P and Q

Plot ACF and PACF of dd lx

- Seems to be cut-offs after lag 1 for both ACF and PACF
  $\Rightarrow$ try $(p,q) = (1,0)$, $(0,1)$ and $(1,1)$
- (Seasonal) ACF and PACF are both significant at $h=12$. The clear cut-off after lag 1s ($s=12$) for ACF (i.e. ACFs are very close to 0 for $h = 2s$, $3s$ and $4s$) suggests $(P,Q)=(0,1)$.

Then compare the AIC, AICc, BIC for various ARIMA$(p,d,q) \times (P,D,Q)_s$ models, all information criteria prefer the ARIMA$(0,1,1) \times (0,1,1)_{12}$ model

We can further compute the residuals to see if the assumption $w_t \sim N(0, \sigma_w^2)$ valid