Notes 10: Time Series Analysis

Nathaniel E. Helwig

Department of Statistics
University of Illinois at Urbana-Champaign

Stat 420: Methods of Applied Statistics
Section N1U/N1G – Spring 2014

Copyright

Outline of Notes

1) Stochastic Processes:
- Overview
- Stationarity

2) Autoregressive (AR) Model:
- Model Form & Assumptions
- Estimation & Inference

3) Moving Average (MA) Model:
- Model Form & Assumptions
- Estimation & Inference

4) ARMA Model:
- Model Form & Assumptions
- Estimation & Inference

5) ARIMA Model:
- Model Form & Assumptions
- Estimation & Inference

6) Model Selection:
- Tests of Model Fit
- Comparing Models

Introduction to Time Series Analysis

Time series data (as the name implies) are random variables collected over some interval of time.

Example: Tuberculosis incidence in US from 1953–2009:

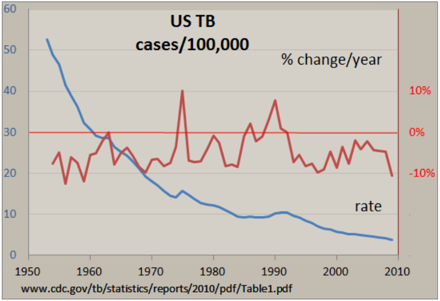


Figure from http://en.wikipedia.org/wiki/Time_series.

Stochastic Process Definition

A *stochastic process* is a sequence of random variables $\{Y_t : t \in \mathcal{T}\}$, where $\mathcal{T}$ is some time index (typically set of all integers).

- Complete structure of process is determined by all sets of distributions of all finite sequences of Y_t
- We typically focus on the mean, variance, and covariance structures of process

Time series analysis tries to uncover the stochastic process that generates the observed data.

Mean and Variance Functions

Given a stochastic process $\{Y_t : t \in \mathcal{T}\}$, the *mean function* is

$$\mu_t = E(Y_t) \quad \forall t \in \mathcal{T}$$

which is the expected value of Y at time point $t \in \mathcal{T}$.

- Given $\{y_t\}_{t=1}^n$ with $E(Y_t) = \mu \; \forall t$, we have $\hat{\mu} = \bar{y} = \frac{1}{n}\sum_{t=1}^n y_t$

The *variance function* of $\{Y_t : t \in \mathcal{T}\}$ is defined as

$$Var(Y_t) = E[(Y_t - \mu_t)^2] \quad \forall t \in \mathcal{T}$$

where μ_t is the mean of the process.

- Given $\{y_t\}_{t=1}^n$ with $E(Y_t) = \mu$ and $Var(Y_t) = \gamma_0 \; \forall t$, we have $\hat{\gamma}_0 = \frac{1}{n-1}\sum_{t=1}^n (y_t - \bar{y})^2$

Autocovariance and Autocorrelation Functions

The *autocovariance function* of $\{Y_t : t \in \mathcal{T}\}$ is defined as

$$\gamma_{t,s} = Cov(Y_t, Y_s) \quad \forall t, s \in \mathcal{T}$$

where $Cov(Y_t, Y_s) = E[(Y_t - \mu_t)(Y_s - \mu_s)] = E(Y_t Y_s) - \mu_t \mu_s$.

Similarly, the *autocorrelation function* (ACF) of $\{Y_t : t \in \mathcal{T}\}$ is

$$\rho_{t,s} = Cor(Y_t, Y_s) \quad \forall t, s \in \mathcal{T}$$

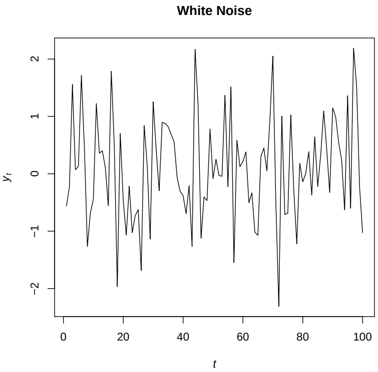
where $Cor(Y_t, Y_s) = \frac{Cov(Y_t, Y_s)}{\sqrt{Var(Y_t) Var(Y_s)}}$.

Example #1: White Noise

White Noise: $y_t = \epsilon_t$ where $\epsilon_t \overset{\text{iid}}{\sim} f(x)\ \forall t \in \mathcal{T}$ such that

- $E(\epsilon_t) = 0\ \forall t \in \mathcal{T}$
- $E(\epsilon_t, \epsilon_s) = \begin{cases} \sigma^2 & \text{if } t = s \\ 0 & \text{if } t \neq s \end{cases}$ for all $t, s \in \mathcal{T}$

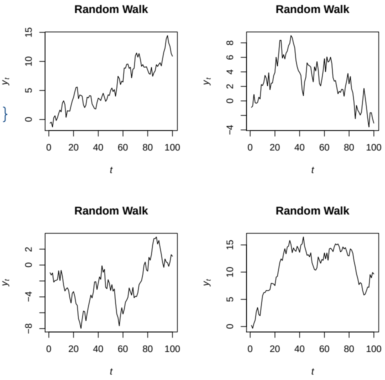
```
> set.seed(123)
> plot(rnorm(100),type="l",
       xlab=expression(italic(t)),
       ylab=expression(italic(y[t])),
       main="White Noise")
```



Example #2: Random Walk

Random Walk: $y_t = y_{t-1} + \epsilon_t$ where ϵ_t is white noise.

```
par(mfrow=c(2,2))
for(k in 1:4){
  set.seed(k)
  y=rep(NA,100)
  y[1]=rnorm(1)
  for(t in 2:100){y[t]=y[t-1]+rnorm(1)}
  plot(y,type="l",
       xlab=expression(italic(t)),
       ylab=expression(italic(y[t])),
       main="Random Walk")
}
```

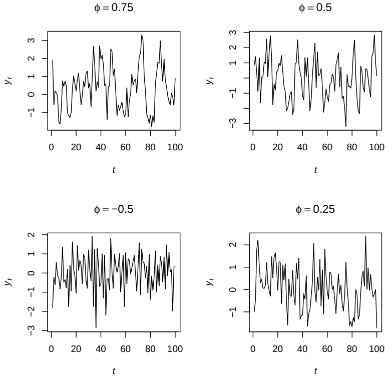


Example #3: AR(1)

AR(1): $y_t = \phi y_{t-1} + \epsilon_t$ where ϵ_t is white noise.

- Current Y relates to previous Y and contemporaneous error

```
par(mfrow=c(2,2))
phi=c(0.75,1/2,-1/2,0.25)
for(k in 1:4){
  set.seed(k)
  y=arima.sim(list(ar=phi[k]),n=100)
  plot(y,type="l",xlab=expression(italic(t)),
       ylab=expression(italic(y[t])),
       main=bquote(phi==.(phi[k])))
}
```



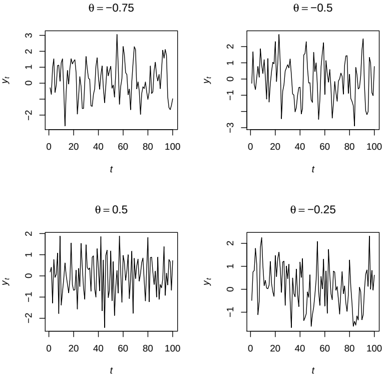
Example #4: MA(1)

MA(1): $y_t = \epsilon_t - \theta\epsilon_{t-1}$ where ϵ_t is white noise.

- Current Y relates to previous and contemporaneous "shock"

```
par(mfrow=c(2,2))
theta=c(0.75,1/2,-1/2,0.25)
for(k in 1:4){
  set.seed(k)
  y=arima.sim(list(ma=theta[k]),n=100)
  plot(y,type="l",xlab=expression(italic(t)),
       ylab=expression(italic(y[t])),
       main=bquote(theta==.(-theta[k])))
}
```

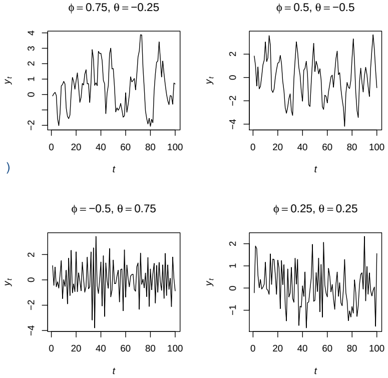
R parameterizes in terms of $-\theta$



Example #5: ARMA(1,1)

ARMA(1,1): $y_t = \phi y_{t-1} + \epsilon_t - \theta\epsilon_{t-1}$ where ϵ_t is white noise.

```
par(mfrow=c(2,2))
phi=c(0.75,1/2,-1/2,0.25)
theta=c(1/4,1/2,-3/4,-1/4)
for(k in 1:4){
  set.seed(k)
  y=arima.sim(list(ar=phi[k],ma=theta[k]),n=100)
  atitle=bquote(phi==.(phi[k]))
  btitle=bquote(theta==.(-theta[k]))
  plot(y,type="l",xlab=expression(italic(t)),
       ylab=expression(italic(y[t])),
       main=bquote(paste(.(atitle),", ",.(btitle))))
}
```



R parameterizes in terms of $-\theta$

Need for Stationarity

To make statistical inferences about time series data, we typically make simplifying assumptions about the process.

A stochastic process is *weakly stationary* if it satisfies:

- $\mu_t = \mu \; \forall t \in \mathcal{T}$
- $\gamma_{t,t-k} = \gamma_{0,k} \; \forall t \in \mathcal{T}$

Stationary processes are simpler to work with, so we will focus on (weakly) stationary processes.

ACF at Lag k

For weakly stationary process, $\gamma_{t,s}$ only depends on the lag $k = |t - s|$, so we can define the autocovariance and autocorrelation functions as

$$\gamma_k = \gamma_{t,t-k} \qquad \text{and} \qquad \rho_k = \rho_{t,t-k} = \frac{\gamma_k}{\gamma_0}$$

Note that $\gamma_0 = Cov(y_t, y_{t-0}) = Var(y_t)$.

Given $\{y_t\}_{t=1}^n$ with $E(Y_t) = \mu$ and $Var(Y_t) = \gamma_0 \ \forall t$, we have

$$r_k = \frac{\sum_{t=1}^{n-k}(y_t - \bar{y})(y_{t+k} - \bar{y})}{\sum_{t=1}^{n}(y_t - \bar{y})^2}$$

is the sample estimate of the ACF for lag $k \in \{0, 1, 2, \ldots\}$

PACF at Lag k

The partial autocorrelation function (PACF) between y_t and y_{t-k} eliminates the effects of the intervening values of y_{t-1} through y_{t-k+1}.

If we have a model of the form: $y_t - \mu = \sum_{j=1}^{k} \phi_{kj}(y_{t-j} - \mu) + \epsilon_t$, then ϕ_{kk} is the PACF between y_t and y_{t-k}.

In general we can compute PACFs from ACFs:

$$\phi_{kk} = \begin{cases} \rho_1 & \text{if } k = 1 \\ \frac{\rho_2 - \rho_1^2}{1 - \rho_1^2} & \text{if } k = 2 \\ \frac{\rho_k - \sum_{j=1}^{k-1} \phi_{k-1,j}\rho_{k-j}}{1 - \sum_{j=1}^{k-1} \phi_{k-1,j}\rho_j} & \text{if } k > 2 \end{cases}$$

where $\phi_{k,j} = \phi_{k-1,j} - \phi_{kk}\phi_{k-1,k-j}$ for $j \in \{1, \ldots, k-1\}$.

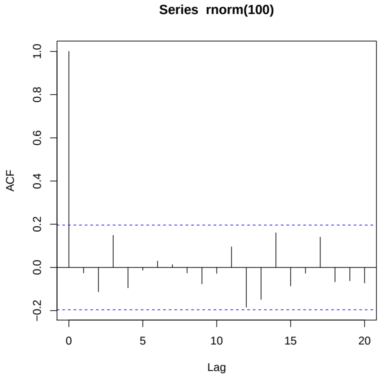
Correlogram

Plot of r_k versus lag k is called *correlogram*.

- Useful for visualizing autocorrelation pattern.
- Can also plot ϕ_{kk} versus lag k

ACF for white noise.

```
set.seed(123)
acf(rnorm(100))
```



Example #1: White Noise Stationarity

White Noise: $y_t = \epsilon_t$ where $\epsilon_t \overset{\text{iid}}{\sim} f(x)\ \forall t \in \mathcal{T}$ such that

- $E(\epsilon_t) = 0\ \forall t \in \mathcal{T}$
- $E(\epsilon_t, \epsilon_s) = \begin{cases} \sigma^2 & \text{if } t = s \\ 0 & \text{if } t \neq s \end{cases}$ for all $t, s \in \mathcal{T}$

White noise satisfies weak stationarity:

- $\mu_t = E(\epsilon_t) = 0\ \forall t \in \mathcal{T}$
- $Var(y_t) = \sigma^2 \implies \gamma_0 = \sigma^2$
- $\gamma_k = 0\ \forall k > 0$
- $\rho_k = \begin{cases} 1 & \text{if } k = 0 \\ 0 & \text{if } k \geq 1 \end{cases}$

Example #2: Random Walk Stationarity

Random Walk: $y_t = y_{t-1} + \epsilon_t$ where ϵ_t is white noise.

Random walk violates weak stationarity:

- $\mu_t = E(\epsilon_t) = 0 \ \forall t \in \mathcal{T}$
- $Var(y_t) = Cov(\sum_{i=0}^{t-1} \epsilon_i, \sum_{i=0}^{t-1} \epsilon_i) = t\sigma^2$
- $\gamma_{t,s} = Cov(\sum_{i=0}^{t-1} \epsilon_i, \sum_{i=0}^{s-1} \epsilon_i) = t\sigma^2 \ \ \forall t \leq s$
- $\rho_{t,s} = \sqrt{t/s} \ \ \forall t \leq s$

Random walk is mean stationary but not autocovariance stationary.

Example #3: AR(1) Stationarity

AR(1): $y_t = \phi y_{t-1} + \epsilon_t$ where ϵ_t is white noise.

Assume that the AR(1) model is weak stationary:

- $\mu_t = \phi \mu_{t-1} + E(\epsilon_t) = 0 \; \forall t \in \mathcal{T}$
- $Var(y_t) = \phi^2 Var(y_{t-1}) + \sigma^2 \Longrightarrow \gamma_0 = \frac{\sigma^2}{1-\phi^2} \Longrightarrow |\phi| < 1$
- $E(y_{t-k} y_t) = \phi E(y_{t-k} y_{t-1}) \Longrightarrow \gamma_k = \phi \gamma_{k-1} \Longrightarrow \gamma_k = \phi^k \frac{\sigma^2}{1-\phi^2}$
- $\rho_k = \begin{cases} 1 & \text{if } k = 0 \\ \phi^k & \text{if } k \geq 1 \end{cases}$

AR(1) model satisfies weak stationarity if $|\phi| < 1$.

Example #4: MA(1) Stationarity

MA(1): $y_t = \epsilon_t - \theta\epsilon_{t-1}$ where ϵ_t is white noise.

Note that MA(1) model satisfies weak stationarity:

- $\mu_t = E(\epsilon_t) - \theta E(\epsilon_{t-1}) = 0 \ \forall t \in \mathcal{T}$
- $Var(y_t) = \sigma^2 + \theta^2\sigma^2 \Longrightarrow \gamma_0 = \sigma^2(1+\theta^2)$
- $E(y_{t-k}y_t) = -\theta E(\epsilon_{t-k}\epsilon_{t-1}) \Longrightarrow \gamma_k = \begin{cases} -\theta\sigma^2 & \text{if } k = 1 \\ 0 & \text{if } k > 1 \end{cases}$
- $\rho_k = \begin{cases} 1 & \text{if } k = 0 \\ \frac{-\theta}{1+\theta^2} & \text{if } k = 1 \\ 0 & \text{if } k \geq 2 \end{cases}$

Example #5: ARMA(1,1) Stationarity

ARMA(1,1): $y_t = \phi y_{t-1} + \epsilon_t - \theta\epsilon_{t-1}$ where ϵ_t is white noise.

Assume that the ARMA(1,1) model is weak stationary:

- $\mu_t = \phi\mu_{t-1} + E(\epsilon_t) - \theta E(\epsilon_{t-1}) = 0 \ \forall t \in \mathcal{T}$
- $\gamma_0 = \phi^2 Var(y_{t-1}) + \sigma^2(1 - 2\phi\theta + \theta^2) \Longrightarrow \gamma_0 = \frac{\sigma^2(1-2\phi\theta+\theta^2)}{1-\phi^2}$
- $\gamma_1 = E(y_{t-1}y_t) = E[y_{t-1}(\phi y_{t-1} + \epsilon_t - \theta\epsilon_{t-1})] = \phi\gamma_0 - \theta\sigma^2$
- $\gamma_k = E(y_{t-k}y_t) = E[y_{t-k}(\phi y_{t-1} + \epsilon_t - \theta\epsilon_{t-1})] = \phi\gamma_{k-1}$ for $k \geq 2$
- $\rho_k = \begin{cases} 1 & \text{if } k = 0 \\ \frac{(1-\phi\theta)(\phi-\theta)}{1-2\phi\theta+\theta^2}\phi^{k-1} & \text{if } k \geq 1 \end{cases}$

ARMA(1,1) model satisfies weak stationarity if $|\phi| < 1$.

AR(p) Model

AR(p) denotes p-th order autoregressive model, which has the form

$$y_t - \mu = \sum_{j=1}^{p} \phi_j (y_{t-j} - \mu) + \epsilon_t$$

where

- μ is unknown constant (mean)
- ϕ_j is slope parameter for lag j response
- ϵ_t is white noise

AR(p) Model and Lag Operators

Let L be a lag (or backshift) operator, such that

$$L^j y_t = y_{t-j}$$

for all $j \in \{1, \ldots, p\}$.

Can write the AR(p) model using lag operator:

$$\left(1 - \sum_{j=1}^{p} \phi_j L^j\right)(y_t - \mu) = \epsilon_t$$

Zero Mean AR(1) Model

Assume that $\mu = 0$, so the AR(1) has the form

$$\begin{aligned} y_t &= \phi y_{t-1} + \epsilon_t \\ &= \phi(\phi y_{t-2} + \epsilon_{t-1}) + \epsilon_t \\ &= \phi[\phi(\phi y_{t-3} + \epsilon_{t-2}) + \epsilon_{t-1}] + \epsilon_t \\ &\vdots \\ &= \sum_{j=0}^{\infty} \phi^j \epsilon_{t-j} \end{aligned}$$

AR(p) Stationarity

$\phi(L) = \left(1 - \sum_{j=1}^{p} \phi_j L^j\right)$ is *AR characteristic equation*.

For stationarity, all p roots of $\phi(L)$ need to exceed 1 in absolute value.

- All roots lie outside unit circle in complex plane
- AR(1): $\phi(L) = 1 - \phi L$ root: $1/\phi$
  stationarity condition: $|\phi| < 1$
- AR(2): $\phi(L) = 1 - \phi_1 L - \phi_2 L^2$ roots: $\frac{\phi_1 \pm \sqrt{\phi_1^2 + 4\phi_2}}{-2\phi_2}$
  stationarity conditions: $\phi_1 + \phi_2 < 1$, $\phi_2 - \phi_1 < 1$, $|\phi_2| < 1$

AR(p) Autocorrelation

Assuming stationarity, we can see that

$$
\begin{aligned}
\rho_k &= \frac{E[(y_{t-k}-\mu)(y_t-\mu)]}{\gamma_0} \\
&= \frac{E[(\sum_{j=1}^{p}\phi_j(y_{t-k-j}-\mu)+\epsilon_{t-k})(\sum_{j=1}^{p}\phi_j(y_{t-j}-\mu)+\epsilon_t)]}{\gamma_0} \\
&= \frac{\gamma_0(\phi_1\rho_{k-1}+\phi_2\rho_{k-2}+\cdots+\phi_p\rho_{k-p})}{\gamma_0} \\
&= \phi_1\rho_{k-1}+\phi_2\rho_{k-2}+\cdots+\phi_p\rho_{k-p}
\end{aligned}
$$

where $\gamma_0 = \phi_1\gamma_1 + \phi_2\gamma_2 + \cdots + \phi_p\gamma_p + \sigma^2 = \frac{\sigma^2}{1-\phi_1\rho_1-\phi_2\rho_2-\cdots-\phi_p\rho_p}$

AR(1) Parameter Estimation

To estimate AR(1) parameters minimize *conditional sum of squares*

$$S_c(\mu, \phi) = \sum_{t=2}^{n} [y_t - \mu - \phi(y_{t-1} - \mu)]^2$$

If we assume process is stationary, $\sum_{t=2}^{n} y_t \approx \sum_{t=2}^{n} y_{t-1}$, so we have

$$\begin{aligned} \hat{\mu} &\approx \bar{y} \\ \hat{\phi} &= \frac{\sum_{t=2}^{n}(y_t - \bar{y})(y_{t-1} - \bar{y})}{\sum_{t=2}^{n}(y_{t-1} - \bar{y})^2} \approx \hat{\rho}_1 \\ \hat{\sigma}^2 &= \frac{\sum_{t=2}^{n} \hat{\epsilon}_t^2}{n-3} \end{aligned}$$

AR(1) Forecasting

If we iterate an AR(1) model forward j time points, we have

$$y_{t+j} = \mu + \phi^j (y_t - \mu) + \sum_{k=0}^{j-1} \phi^k \epsilon_{t+k+1}$$
$$\hat{y}_{t+j} = \mu + \phi^j (y_t - \mu)$$

Forecast error for j time points in future is

$$\epsilon_t(j) = y_{t+j} - \hat{y}_{t+j} = \sum_{k=0}^{j-1} \phi^k \epsilon_{t+k+1}$$

so the variance of a forecast j time points in future is

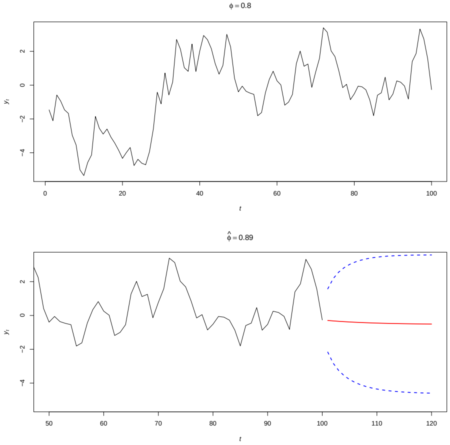
$$Var(\epsilon_t(j)) = \sigma^2 \sum_{k=0}^{j-1} \phi^{2k} = \sigma^2 \frac{1-\phi^{2j}}{1-\phi^2}$$

Can form $100(1-\alpha)\%$ prediction intervals using

$$\hat{y}_{t+j} \pm z_{\alpha/2} \sqrt{Var(\epsilon_t(j))}$$

AR(1) Example

```
set.seed(1234)
y=arima.sim(list(ar=0.8),n=100)
armod=arima(y,c(1,0,0))
arprd=predict(armod,20)  # predict j=20 points
par(mfrow=c(2,1))
plot(y,type="l",xlab=expression(italic(t)),
     ylab=expression(italic(y[t])),
     main=bquote(phi==0.8))
plot(y,type="l",xlab=expression(italic(t)),
     ylab=expression(italic(y[t])),
     main=bquote(hat(phi)==0.89),xlim=c(50,120))
lines(arprd$p,col="red",lwd=2)
lines(arprd$p-arprd$se*qnorm(.975),
      col="blue",lty=2,lwd=2)
lines(arprd$p+arprd$se*qnorm(.975),
      col="blue",lty=2,lwd=2)
```



AR(p) Parameter Estimation

For an AR(p) model, we can derive the Yule-Walker equations

$$\begin{aligned}
\rho_1 &= \phi_1 + \phi_2\rho_1 + \phi_3\rho_2 + \cdots + \phi_p\rho_{p-1} \\
\rho_2 &= \phi_1\rho_1 + \phi_2 + \phi_3\rho_1 + \cdots + \phi_p\rho_{p-2} \\
&\vdots \\
\rho_p &= \phi_1\rho_{p-1} + \phi_2\rho_{p-2} + \phi_3\rho_{p-3} + \cdots + \phi_p
\end{aligned}$$

which uses the ACF definition and the relation $\rho_0 = 1$.

Given $\hat{\boldsymbol{\rho}} = \{\hat{\rho}_j\}_{p\times 1}$ and $\hat{\mathbf{R}} = \{r_{jk}\}_{p\times p}$ with $r_{jk} = \hat{\rho}_{t-j,t-k}$ for $j,k \in \{0,\ldots,p-1\}$, we have

$$\hat{\boldsymbol{\rho}} = \hat{\mathbf{R}}\phi \quad \Longrightarrow \quad \hat{\phi} = \hat{\mathbf{R}}^{-1}\hat{\boldsymbol{\rho}}$$

is the method of moments estimator of ϕ.

AR(2) Example

For an AR(2) model, the Yule-Walker equations are

$$\begin{aligned}\rho_1 &= \phi_1 + \phi_2\rho_1 \\ \rho_2 &= \phi_1\rho_1 + \phi_2\end{aligned}$$

so the estimates are given by

$$\begin{aligned}\hat{\phi}_1 &= \frac{r_1(1 - r_2)}{1 - r_1^2} \\ \hat{\phi}_2 &= \frac{r_2 - r_1^2}{1 - r_1^2}\end{aligned}$$

where r_k is sample ACF at lag k.

MA(q) Model

MA(q) denotes q-th order moving average model, which has the form

$$y_t = \epsilon_t - \sum_{i=1}^{q} \theta_i \epsilon_{t-i}$$

where

- θ_i is slope parameter for lag i "shock"
- ϵ_t is white noise

Nonuniqueness

Consider the MA(1) model: $y_t = \epsilon_t - \theta\epsilon_{t-1}$

Previously we derived that $\rho_k = \begin{cases} 1 & \text{if } k = 0 \\ \frac{-\theta}{1+\theta^2} & \text{if } k = 1 \\ 0 & \text{if } k \geq 2 \end{cases}$

Note that θ and $1/\theta$ produce the same ACF:

$$\frac{-(1/\theta)}{1+(1/\theta)^2} = \frac{-1}{\theta + 1/\theta} = \frac{-\theta}{\theta^2 + 1}$$

Invertibility of MA(1)

Consider the MA(1) model: $y_t = \epsilon_t - \theta\epsilon_{t-1}$

Rewrite the MA(1) model: $\epsilon_t = y_t + \theta\epsilon_{t-1}$

If $|\theta| < 1$, we can iterate backwards: $\epsilon_t = \sum_{i=0}^{\infty} \theta^i y_{t-i}$

If $|\theta| < 1$, the MA(1) model can be inverted to an AR(∞) model:

$$y_t = \epsilon_t - \sum_{i=1}^{\infty} \theta^i y_{t-i}$$

Invertibility of MA(q)

Consider the MA(q) model: $y_t = \epsilon_t - \sum_{i=1}^{q} \theta_i \epsilon_{t-i}$

Using lag operators we can rewrite the model as

$$y_t = (1 - \sum_{i=1}^{q} \theta_i L^i)\epsilon_t$$

$\theta(L) = (1 - \sum_{i=1}^{q} \theta_i L^i)$ is the *MA characteristic equation*.

- For invertibility, need roots of $\theta(L)$ to exceed 1 in absolute value

MA(q) Autocorrelation

For the MA(q) model, we can see that

$$\begin{aligned}
\rho_k &= \frac{E[y_{t-k} y_t]}{\gamma_0} \\
&= \frac{\sigma^2(-\theta_k + \theta_1\theta_{k+1} + \theta_2\theta_{k+2} + \cdots + \theta_{q-k}\theta_q)}{\gamma_0} \\
&= \frac{-\theta_k + \theta_1\theta_{k+1} + \theta_2\theta_{k+2} + \cdots + \theta_{q-k}\theta_q}{1 + \theta_1^2 + \theta_2^2 + \cdots + \theta_q^2}
\end{aligned}$$

for $k \in \{1, \ldots, q\}$ where $\gamma_0 = (1 + \theta_1^2 + \theta_2^2 + \cdots + \theta_q^2)\sigma^2$.

Note that $\rho_k = 0$ for $k > q$.

MA(1) Method of Moments

Recall MOM equates sample statistics with population parameters.

For MA(1) model we can estimate θ using ACF:

$$\rho_1 = \frac{-\theta}{1+\theta^2} \quad \Longrightarrow \quad r_1 = \frac{-\theta}{1+\theta^2} \quad \Longrightarrow \quad r_1 + \theta + r_1\theta^2 = 0$$

Solving quadratic equation and applying invertibility restriction gives

$$\hat{\theta} = \frac{-1 + \sqrt{1 - 4r_1^2}}{2r_1}$$

and note that

- A real solution exists if $|r_1| < 0.5$
- MOM fails as estimator if $|r_1| > 0.5$

Invertible MA(q) Least Squares Estimation

For MA(1) we minimize conditional sum of squares

$$S_c(\theta) = \sum \epsilon_t^2 = \sum (\sum_{i=0}^{\infty} \theta^i y_{t-i})^2$$

using some numerical optimization technique.

With n observations we represent ϵ_t with recursive equations:

$$\epsilon_t = y_t + \theta \epsilon_{t-1}$$

for $t \in \{1, \ldots, n\}$ with $\epsilon_0 = 0$.

Same idea for MA(q) model, but the recursive equations are

$$\epsilon_t = y_t + \sum_{i=1}^{q} \theta_i \epsilon_{t-i}$$

for $t \in \{1, \ldots, n\}$ with $\epsilon_0 = \epsilon_{-1} = \cdots = \epsilon_{-q} = 0$.

MA(1) Forecasting

Consider the MA(1) model with nonzero mean: $y_t = \mu + \epsilon_t - \theta\epsilon_{t-1}$

If we iterate an MA(1) model forward 1 time point, we have

$$\begin{aligned} y_{t+1} &= \mu + \epsilon_{t+1} - \theta\epsilon_t \\ \hat{y}_{t+j} &= \mu - \theta E(\epsilon_t | y_1, \ldots, y_t) = \hat{\mu} - \theta\epsilon_t \end{aligned}$$

Forecast error for 1 time point in future is

$$\epsilon_t(1) = y_{t+1} - \hat{y}_{t+1} = \epsilon_{t+1}$$

so the variance of a forecast 1 time point in future is $Var(\epsilon_t(1)) = \sigma^2$.

Can form $100(1-\alpha)\%$ prediction intervals using

$$\hat{y}_{t+1} \pm z_{\alpha/2}\hat{\sigma}$$

MA(1) Forecasting (continued)

Consider the MA(1) model with nonzero mean: $y_t = \mu + \epsilon_t - \theta\epsilon_{t-1}$

If we iterate an MA(1) model forward $j > 1$ time points, we have

$$
\begin{aligned}
y_{t+j} &= \mu + \epsilon_{t+j} - \theta\epsilon_{t+j-1} \\
\hat{y}_{t+j} &= \hat{\mu} + E(\epsilon_{t+j}|y_1, \ldots, y_t) - \theta E(\epsilon_{t+j-1}|y_1, \ldots, y_t) = \hat{\mu}
\end{aligned}
$$

Forecast error for $j > 1$ time points in future is

$$\epsilon_t(j) = y_{t+1} - \hat{y}_{t+1} = \epsilon_{t+j} - \theta\epsilon_{t+j-1}$$

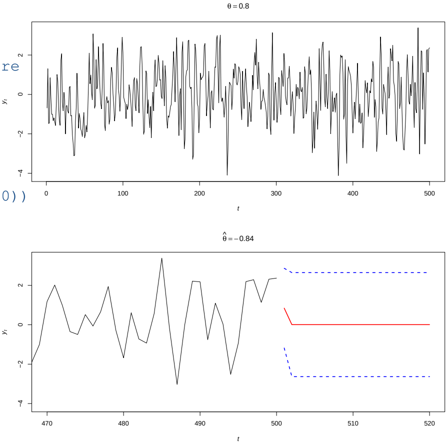
so variance of a forecast $Var(\epsilon_t(j)) = \sigma^2(1 + \theta^2) = \gamma_0$

Can form $100(1-\alpha)$% prediction intervals using

$$\hat{y}_{t+j} \pm z_{\alpha/2}\sqrt{\gamma_0}$$

MA(1) Example

```
set.seed(1234)
y=arima.sim(list(ma=0.8),n=500)
mamod=arima(y,c(0,0,1))
maprd=predict(mamod,20)  # predict j=20 into future
par(mfrow=c(2,1))
plot(y,type="l",xlab=expression(italic(t)),
     ylab=expression(italic(y[t])),
     main=bquote(theta==0.8))
plot(y,type="l",xlab=expression(italic(t)),
     ylab=expression(italic(y[t])),
     main=bquote(hat(theta)==-0.84),xlim=c(470,520))
lines(maprd$p,col="red",lwd=2)
lines(maprd$p-maprd$se*qnorm(.975),
     col="blue",lty=2,lwd=2)
lines(maprd$p+maprd$se*qnorm(.975),
     col="blue",lty=2,lwd=2)
```



ARMA(p,q) Model

ARMA(p,q) denotes (p, q)-th order autoregressive moving average model, which has the form

$$y_t - \mu = \sum_{j=1}^{p} \phi_j (y_{t-j} - \mu) + \epsilon_t - \sum_{i=1}^{q} \theta_i \epsilon_{t-i}$$

where

- μ is overall constant (mean)
- ϕ_j is slope parameter for lag j response
- θ_i is slope parameter for lag i “shock”
- ϵ_t is white noise

Zero Mean ARMA(1,1)

Assuming $\mu = 0$, we can write ARMA(1,1) as

$$\begin{aligned}
y_t &= \phi y_{t-1} + \epsilon_t - \theta\epsilon_{t-1} \\
&= \phi(\phi y_{t-2} + \epsilon_{t-1} - \theta\epsilon_{t-2}) + \epsilon_t - \theta\epsilon_{t-1} \\
&= \phi[\phi(\phi y_{t-3} + \epsilon_{t-2} - \theta\epsilon_{t-3}) + \epsilon_{t-1} - \theta\epsilon_{t-2}] + \epsilon_t - \theta\epsilon_{t-1} \\
&\vdots \\
&= \epsilon_t + (\phi - \theta)\sum_{j=1}^{\infty} \phi^{j-1}\epsilon_{t-j}
\end{aligned}$$

ARMA(p,q) Stationarity and Invertibility

For estimation purposes, we require stationarity and invertibility:

- Stationarity: roots of AR characteristic equation exceed 1 in absolute value
- Invertibility: roots of MA characteristic equation exceed 1 in absolute value

Note: ARMA(p,q) is combination of AR(p) and MA(q), so it requires both AR(p) and MA(q) conditions.

Method of Moments for ARMA(1,1)

ARMA(1,1): $y_t = \phi y_{t-1} + \epsilon_t - \theta\epsilon_{t-1}$ where ϵ_t is white noise.

ACF: $\rho_k = \frac{(1-\phi\theta)(\phi-\theta)}{1-2\phi\theta+\theta^2}\phi^{k-1}$

Using the ACF we have the MOM estimators: $\hat{\phi} = r_2/r_1$

Given $\hat{\phi}$ we find invertible root of

$$r_1 = \frac{(1-\hat{\phi}\theta)(\hat{\phi}-\theta)}{1-2\hat{\phi}\theta+\theta^2}$$

to estimate θ.

ARMA(p,q) Least Squares Estimation

For ARMA(p,q) we minimize conditional sum of squares

$$S_c(\phi, \theta) = \sum \epsilon_t^2$$

using some numerical optimization technique.

With n observations we represent ϵ_t with recursive equations:

$$\epsilon_t = y_t - \sum_{j=1}^{p} \phi_j y_{t-j} + \sum_{i=1}^{q} \theta_i \epsilon_{t-i}$$

for $t \in \{1, \ldots, n\}$ with $\epsilon_p = \epsilon_{p-1} = \ldots = \epsilon_{p+1-q} = 0$.

ARMA(1,1) Forecasting

Consider the ARMA(1,1) model with nonzero mean:
$y_t - \mu = \phi(y_{t-1} - \mu) + \epsilon_t - \theta\epsilon_{t-1}$

If we iterate an ARMA(1,1) model forward 1 time point, we have

$$y_{t+1} = \mu + \phi(y_t - \mu) + \epsilon_{t+1} - \theta\epsilon_t$$
$$\hat{y}_{t+1} = \mu + \phi(y_t - \mu) - \theta\epsilon_t$$

Forecast error for 1 time point in future is

$$\epsilon_t(1) = y_{t+1} - \hat{y}_{t+1} = \epsilon_{t+1}$$

so the variance of a forecast 1 time point in future is $Var(\epsilon_t(1)) = \sigma^2$.

Can form $100(1-\alpha)$% prediction intervals using

$$\hat{y}_{t+1} \pm z_{\alpha/2}\hat{\sigma}$$

ARMA(1,1) Forecasting (continued)

Consider the ARMA(1,1) model with nonzero mean:
$y_t - \mu = \phi(y_{t-1} - \mu) + \epsilon_t - \theta\epsilon_{t-1}$

If we iterate an ARMA(1,1) model forward $j > 1$ time points, we have

$$\begin{aligned} y_{t+j} &= \mu + \phi(y_{t+j-1} - \mu) + \epsilon_{t+j} - \theta\epsilon_{t+j-1} \\ \hat{y}_{t+j} &= \mu + \phi(\hat{y}_{t+j-1} - \mu) \end{aligned}$$

Forecast error for $j > 1$ time points in future is

$$\epsilon_t(j) = y_{t+j} - \hat{y}_{t+j} = \phi(y_{t+j-1} - \hat{y}_{t+j-1}) + \epsilon_{t+j} - \theta\epsilon_{t+j-1}$$

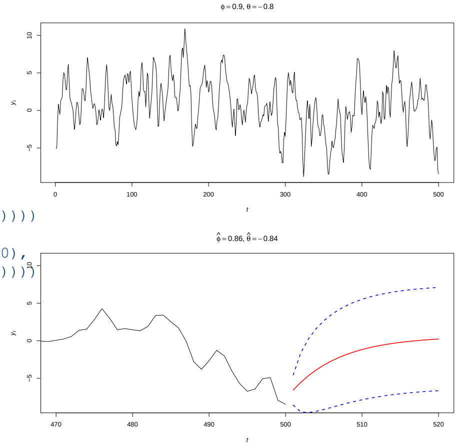
and for large j we have that $Var(\epsilon_t(j)) \approx \gamma_0$

Can form $100(1-\alpha)\%$ prediction intervals using

$$\hat{y}_{t+j} \pm z_{\alpha/2}\sqrt{Var(\epsilon_t(j))}$$

ARMA(1,1) Example

```
set.seed(1234)
y=arima.sim(list(ar=0.9,ma=0.8),n=500)
armamod=arima(y,c(1,0,1))
armaprd=predict(armamod,20)  # predict j=20
a1title=bquote(phi==0.9)
b1title=bquote(theta==-0.8)
a2title=bquote(hat(phi)==0.86)
b2title=bquote(hat(theta)==-0.84)
par(mfrow=c(2,1))
plot(y,type="l",xlab=expression(italic(t)),
     ylab=expression(italic(y[t])),
     main=bquote(paste(.(a1title),", ",.(b1title))))
plot(y,type="l",xlab=expression(italic(t)),
     ylab=expression(italic(y[t])),xlim=c(470,520),
     main=bquote(paste(.(a2title),", ",.(b2title))))
lines(armaprd$p,col="red",lwd=2)
lines(armaprd$p-armaprd$se*qnorm(.975),
     col="blue",lty=2,lwd=2)
lines(armaprd$p+armaprd$se*qnorm(.975),
     col="blue",lty=2,lwd=2)
```



ARIMA Motivation

Suppose that $\{Y_t : t \in \mathcal{T}\}$ is nonstationary

- Y_t has a non-constant mean (i.e., $\mu_t \neq \mu \; \forall t$).
- And/or Y_t has a non-constant variance (i.e., $Var(y_t) \neq \gamma_0$).

We can define the *first difference* of Y_t as

$$\Delta Y_t = Y_t - Y_{t-1}$$

which could be a stationary process:

- Random Walk: $y_t = y_{t-1} + \epsilon_t$ is nonstationary
- First Difference: $\Delta y_t = \epsilon_t$ is stationary

Higher-Order Differences

Similarly, we could define the second difference of Y_t as

$$\begin{aligned}\Delta^2 Y_t &= \Delta(\Delta Y_t) \\ &= \Delta(Y_t - Y_{t-1}) \\ &= \Delta Y_t - \Delta Y_{t-1} \\ &= (Y_t - Y_{t-1}) - (Y_{t-1} - Y_{t-2}) \\ &= Y_t - 2Y_{t-1} + Y_{t-2}\end{aligned}$$

More generally we could define the d-th difference as

$$\Delta^d Y_t = \Delta\Delta \cdots \Delta Y_t$$

ARIMA(p,d,q)

ARIMA(p,d,q) denotes (p, d, q)-th order autoregressive integrated moving average model, which has the form

$$w_t - \mu = \sum_{j=1}^{p} \phi_j (w_{t-j} - \mu) + \epsilon_t - \sum_{i=1}^{q} \theta_i \epsilon_{t-i}$$

where

- $w_t = \Delta^d y_t$ is the d-th difference of y_t
- μ is overall constant (mean)
- ϕ_j is slope parameter for lag j response
- θ_i is slope parameter for lag i "shock"
- ϵ_t is white noise

ARIMA(p,d,q) Stationarity and Invertibility

ARIMA(p,d,q) assumes d-th difference of y_t follows ARMA(p,q) model.

We require stationarity and invertibility of $w_t = \Delta^d y_t$.

- Stationarity: roots of AR characteristic equation exceed 1 in absolute value
- Invertibility: roots of MA characteristic equation exceed 1 in absolute value

ARIMA(p,d,q) Special Cases

If there is no AR component, we have IMA(d,q)

- IMA(1,1) important model in economics and business
- White noise influence does not die out

If there is no MA component, we have ARI(p,d)

- ARI(1,1) is most popular version
- Similar to stationary AR model, but has $|\phi_j| \geq 1$

ARIMA(p,d,q) Least Squares Estimation

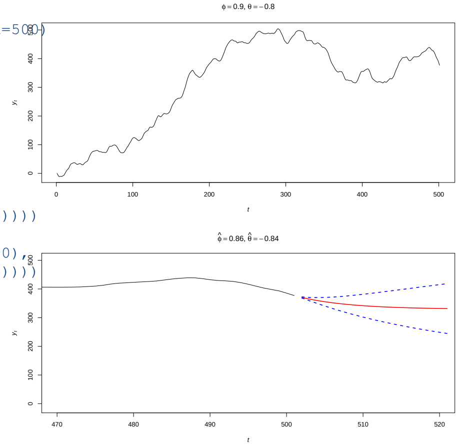
Same idea as before, but now using difference scores.

Define $w_t = \Delta^d y_t$ and minimize $\sum \epsilon_t^2$ numerically.

Recursive update of ϵ_t will depend on model form.

ARIMA(1,1,1) Example

```
set.seed(1234)
y=arima.sim(list(order=c(1,1,1),ar=0.9,ma=0.8),n=500)
arimamod=arima(y,c(1,1,1))
arimaprd=predict(arimamod,20)  # predict j=20
a1title=bquote(phi==0.9)
b1title=bquote(theta==-0.8)
a2title=bquote(hat(phi)==0.86)
b2title=bquote(hat(theta)==-0.84)
par(mfrow=c(2,1))
plot(y,type="l",xlab=expression(italic(t)),
     ylab=expression(italic(y[t])),
     main=bquote(paste(.(a1title),", ",.(b1title))))
plot(y,type="l",xlab=expression(italic(t)),
     ylab=expression(italic(y[t])),xlim=c(470,520),
     main=bquote(paste(.(a2title),", ",.(b2title))))
lines(arimaprd$p,col="red",lwd=2)
lines(arimaprd$p-arimaprd$se*qnorm(.975),
     col="blue",lty=2,lwd=2)
lines(arimaprd$p+arimaprd$se*qnorm(.975),
     col="blue",lty=2,lwd=2)
```



Test for Differencing (Test for Unit Root)

Suppose we are considering a model of the form

$$Y_t = \alpha Y_{t-1} + X_t$$

where $\{X_t\}$ is a stationary process.

- $\{Y_t\}$ is stationary if $|\alpha| < 1$ and nonstationary if $\alpha = 1$
- If $\{X_t\}$ is AR(k) process $X_t = \sum_{j=1}^{k} \phi_k X_{t-k} + \epsilon_t$

If $\alpha = 1$ we have $X_t = Y_t - Y_{t-1}$ and the model is

$$\begin{aligned} Y_t - Y_{t-1} &= \gamma Y_{t-1} + X_t \\ &= \gamma Y_{t-1} + \textstyle\sum_{j=1}^{k} \phi_j X_{t-j} + \epsilon_t \\ &= \gamma Y_{t-1} + \textstyle\sum_{j=1}^{k} \phi_j (Y_{t-j} - Y_{t-j-1}) + \epsilon_t \end{aligned}$$

where $\gamma = \alpha - 1$ is the parameter.

Augmented Dickey-Fuller Test

Dickey-Fuller (DF) Test considers situation on previous slide

- $H_0 : \gamma = 0$ ($\alpha = 1$) $\Longleftrightarrow \{Y_t\}$ is difference nonstationary.
  $\{Y_t\}$ is difference nonstationary $\Longleftrightarrow \{Y_t\}$ is nonstationary, but $\{\Delta Y_t\}$ is stationary.
- $H_1 : \gamma < 0$ ($|\alpha| < 1$) $\Longleftrightarrow \{Y_t\}$ is stationary.

Can test H_0 using regression approach

- ΔY_t is response
- Y_{t-1} and $(Y_{t-j} - Y_{t-j-1})$ are predictors for $j \in \{1, \ldots, k\}$

Augmented Dickey-Fuller (ADF) Test

- Test statistic: $T = \frac{\hat{\gamma}}{SE(\hat{\gamma})}$ where $\hat{\gamma}$ is OLS estimate
- Compare to ADF null distribution (see Fuller, 1996)

Augmented Dickey-Fuller Example #1

Check first difference:

```
> library(tseries)
> set.seed(1234)
> y=arima.sim(list(order=c(1,0,1),ar=0.9,ma=0.8),n=500)
> adf.test(y)

 Augmented Dickey-Fuller Test

data:  y
Dickey-Fuller = -6.7287, Lag order = 7, p-value = 0.01
alternative hypothesis: stationary

Warning message:
In adf.test(y) : p-value smaller than printed p-value
```

Reject $H_0 : \gamma = 0 \Longleftrightarrow \{Y_t\}$ is stationary (don't need first difference).

Augmented Dickey-Fuller Example #2

Check first difference:

```
> set.seed(1234)
> y=arima.sim(list(order=c(1,1,1),ar=0.9,ma=0.8),n=500)
> adf.test(y)

 Augmented Dickey-Fuller Test

data:  y
Dickey-Fuller = -0.8881, Lag order = 7, p-value = 0.9537
alternative hypothesis: stationary
```

Retain $H_0 : \gamma = 0 \iff \{Y_t\}$ is difference nonstationary (need first difference).

Check second difference:

```
> ydiff=y[2:500]-y[1:499]
> adf.test(ydiff)

 Augmented Dickey-Fuller Test

data:  ydiff
Dickey-Fuller = -6.6098, Lag order = 7, p-value = 0.01
alternative hypothesis: stationary

Warning message:
In adf.test(ydiff) : p-value smaller than printed p-value
```

Reject $H_0 : \gamma = 0 \iff \{\Delta Y_t\}$ is stationary (don't need second difference).

Augmented Dickey-Fuller Example #3

Check first difference:

```
> set.seed(1234)
> y=arima.sim(list(order=c(1,2,1),ar=0.9,ma=0.8),n=500)
> adf.test(y)

 Augmented Dickey-Fuller Test

data:  y
Dickey-Fuller = -2.917, Lag order = 7, p-value = 0.1902
alternative hypothesis: stationary
```

Retain $H_0 : \gamma = 0 \iff \{Y_t\}$ is difference nonstationary (need first difference).

Check second difference:

```
> ydiff=y[2:500]-y[1:499]
> adf.test(ydiff)

 Augmented Dickey-Fuller Test

data:  ydiff
Dickey-Fuller = -0.9944, Lag order = 7, p-value = 0.9396
alternative hypothesis: stationary
```

Retain $H_0 : \gamma = 0 \iff \{\Delta Y_t\}$ is difference nonstationary (need second difference).

Residual Diagnostics

Like in regression, we can examine residuals of fit model.

- If model fits the data, residuals should be white noise process.

The Ljung-Box Q-statistic tests whether the residuals from the fitted model follow a white noise process:

$$Q = n(n+2)\sum_{k=1}^{K}\frac{\tilde{r}_k^2}{n-k}$$

where

- $\tilde{r}_k$ is residual autocorrelation at lag k
- Under $H_0 : \tilde{r}_1 = \cdots = \tilde{r}_K = 0$, we have $Q \sim \chi^2_{K-\nu}$ (ν is number of parameters in fit model)
- Set K large enough so $\psi_k^2 Var(\epsilon_{t-k}) \approx 0 \ \forall k > K$

Residual Diagnostics Example #1

Test $H_0 : \tilde{r}_1 = \cdots = \tilde{r}_{10} = 0$:

```
> set.seed(1234)
> y=arima.sim(list(order=c(1,0,0),ar=0.9),n=100)
> armod=arima(y,c(1,0,0))
> Box.test(armod$resid,lag=10,type="Ljung-Box",fitdf=1)

 Box-Ljung test

data:  armod$resid
X-squared = 10.7682, df = 9, p-value = 0.2919

> racf=acf(armod$resid,10,plot=FALSE)
> Qstat=100*(102)*sum((racf$acf[2:11]^2)/(100-1:10))
> Qstat
[1] 10.76817
> 1-pchisq(Qstat,10-1)
[1] 0.291935
```

Retain H_0 at $\alpha = 0.05$

Residual Diagnostics Example #2

Test $H_0 : \tilde{r}_1 = \cdots = \tilde{r}_{10} = 0$:

```
> set.seed(1234)
> y=arima.sim(list(order=c(2,0,0),ar=c(0.4,0.3)),n=100)
> armod=arima(y,c(1,0,0))
> Box.test(armod$resid,lag=10,type="Ljung-Box",fitdf=1)

 Box-Ljung test

data:  armod$resid
X-squared = 32.1002, df = 9, p-value = 0.0001913

> racf=acf(armod$resid,10,plot=FALSE)
> Qstat=100*(102)*sum((racf$acf[2:11]^2)/(100-1:10))
> Qstat
[1] 32.10023
> 1-pchisq(Qstat,10-1)
[1] 0.0001913386
```

Reject H_0 at $\alpha = 0.05$

Residual Diagnostics Example #3

Test $H_0 : \tilde{r}_1 = \cdots = \tilde{r}_{10} = 0$:

```
> set.seed(1234)
> y=arima.sim(list(order=c(2,0,1),ar=c(0.4,0.3),ma=0.9),n=100)
> armod=arima(y,c(2,0,0))
> Box.test(armod$resid,lag=10,type="Ljung-Box",fitdf=2)

 Box-Ljung test

data:  armod$resid
X-squared = 16.1814, df = 8, p-value = 0.03986

> racf=acf(armod$resid,10,plot=FALSE)
> Qstat=100*(102)*sum((racf$acf[2:11]^2)/(100-1:10))
> Qstat
[1] 16.18144
> 1-pchisq(Qstat,10-2)
[1] 0.0398555
```

Reject H_0 at $\alpha = 0.05$

Residual Diagnostics Example #4

Test $H_0 : \tilde{r}_1 = \cdots = \tilde{r}_{10} = 0$:

```
> set.seed(1234)
> y=arima.sim(list(order=c(2,0,1),ar=c(0.4,0.3),ma=0.9),n=100)
> armod=arima(y,c(2,0,1))
> Box.test(armod$resid,lag=10,type="Ljung-Box",fitdf=3)

 Box-Ljung test

data:  armod$resid
X-squared = 10.8586, df = 7, p-value = 0.1449

> racf=acf(armod$resid,10,plot=FALSE)
> Qstat=100*(102)*sum((racf$acf[2:11]^2)/(100-1:10))
> Qstat
[1] 10.85856
> 1-pchisq(Qstat,10-3)
[1] 0.1449087
```

Retain H_0 at $\alpha = 0.05$

Comparing ACFs/PACFs

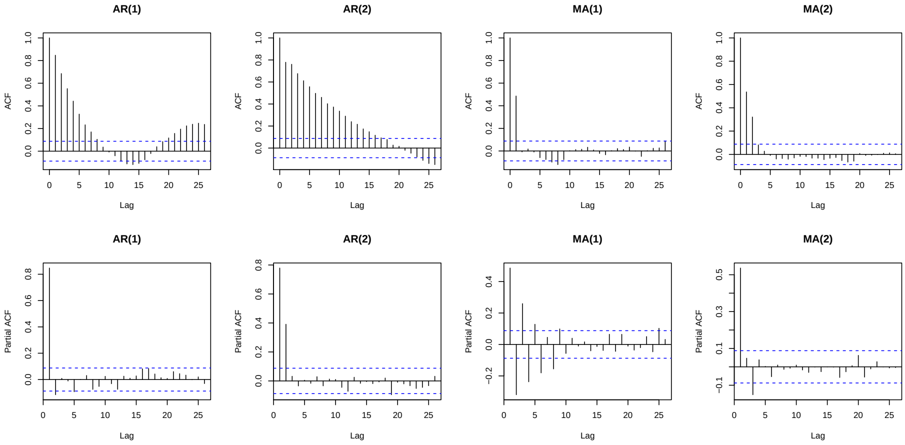
We can compare ACF and PACF properties for different processes:

	AR(p)	**MA(q)**	**ARMA(p,q)**
ACF	Tails off	Cuts off after lag q	Tails off
PACF	Cuts off after lag p	Tails off	Tails off

Note. From Cryer & Chan (2008); table assumes $p > 0$ and $q > 0$.

By examining correlograms we can get idea of which process is appropriate for a given sample of data.

Comparing ACFs/PACFs Example



Comparing ACFs/PACFs Example (R Code)

```
set.seed(1234)
par(mfcol=c(2,4))
y=arima.sim(list(order=c(1,0,0),ar=0.9),n=500)
acf(y,main="AR(1)")
acf(y,type="partial",main="AR(1)")
y=arima.sim(list(order=c(2,0,0),ar=c(0.5,0.4)),n=500)
acf(y,main="AR(2)")
acf(y,type="partial",main="AR(2)")
y=arima.sim(list(order=c(0,0,1),ma=0.9),n=500)
acf(y,main="MA(1)")
acf(y,type="partial",main="MA(1)")
y=arima.sim(list(order=c(0,0,2),ma=c(0.5,0.4)),n=500)
acf(y,main="MA(2)")
acf(y,type="partial",main="MA(2)")
```

Model Selection

We can compare several models and use fit indices (e.g., AIC) to choose best model.

Can use the `auto.arima` function in `forecast` package to fit a collection of models and return ARIMA(p,d,q) model with smallest AIC.

Note that automatic selection is NOT guaranteed to uncover true process; in practice compare results from AIC and Q-statistics.

- Can also use sample ACF/PACF to guide selection

Model Selection Example (Warning!)

```
> set.seed(1234)
> y=arima.sim(list(order=c(1,0,0),ar=0.9),n=500)
> auto.arima(y)
Series: y
ARIMA(0,1,0)

sigma^2 estimated as 1.127:  log likelihood=-737.82
AIC=1477.65   AICc=1477.65   BIC=1481.86
> y=arima.sim(list(order=c(1,0,0),ar=0.9),n=500)
> auto.arima(y)
Series: y
ARIMA(1,0,0) with zero mean

Coefficients:
         ar1
      0.8836
s.e.  0.0206

sigma^2 estimated as 0.8938:  log likelihood=-682.16
AIC=1368.33   AICc=1368.35   BIC=1376.75
> y=arima.sim(list(order=c(1,0,0),ar=0.9),n=500)
> auto.arima(y)
Series: y
ARIMA(1,1,2)

Coefficients:
         ar1      ma1      ma2
      0.8461  -0.9390  -0.0483
s.e.  0.0308   0.0532   0.0512

sigma^2 estimated as 0.9536:  log likelihood=-696.88
AIC=1401.76   AICc=1401.84   BIC=1418.61
```