UNIVERSITY of LIMERICK

OLLSCOIL LUIMNIGH

# Time Series Analysis

# MS 4218

joseph.lynch@ul.ie

# Outline

Model diagnostics

Having chosen a model and estimated its parameters, we need to see how well the model fits the data in question.

There are two modes of analysis

- Residual analysis
- Analysis of over-parameterised models

## Assessment of model fit

Residual analysis

- Residual plots
- Normality of residuals
- Auto-correlation of residuals
- Ljung-Box test

# Residuals

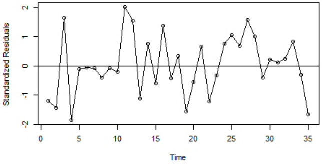
Residual $=$ Observed - Predicted.

If model is correctly specified, and the parameter estimates are close to their true values, then residuals should behave like white noise, i.e.,

independent, identically distributed Normal random variables with zero mean and common variance $\sigma_e^2$.

After a model has been fitted, we plot the residuals and hope to see a random pattern around 0.

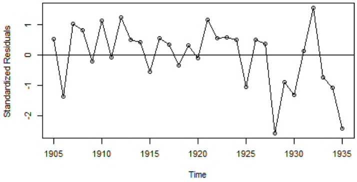
# Residual analysis from AR(1) model of industrial Colour property




```
data(color) ; # TSLecture6 Pages 42:43
m1.color=arima(color,order=c(1,0,0))#mles
plot(rstandard(m1.color),
ylab ='Standardized Residuals',type='o')
abline(h=0)
```

Standardised residuals look random and equally distributed around 0.

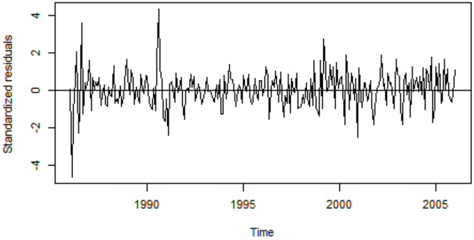
# Residual analysis from AR(3) model of sqrt(hare) data




```
data(hare); # TSLecture6 Pages 44:47
m1.hare=arima(sqrt(hare),order=c(3,0,0))
m2.hare=arima(sqrt(hare),order=c(3,0,0),
fixed=c(NA,0,NA,NA)); m2.hare #phi2=0
plot(rstandard(m2.hare), ylab='Standardized
Residuals',type='o'); abline(h=0)
```

Model suspect because of ↑ variation at end.

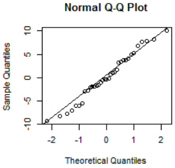
# Residual analysis from IMA(1,1) model of log(oil.price) data




```
data(oil.price); #TSLecture6 Pages 48:49
m1.oil=arima(log(oil.price),order=c(0,1,1))#mles
plot(rstandard(m1.oil),
ylab='Standardized residuals',type='l')
abline(h=0)
```

Model suspect because at least 3 residuals $> |3|$.

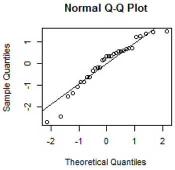
## Normality of residuals from AR(1) model of Colour property




```
win.graph(width=2.5,height=2.5,pointsize=8)
qqnorm(residuals(m1.color))
qqline(residuals(m1.color))
shapiro.test(residuals(m1.color))
#TSLecture3 Page 40
#W=0.9754, p-value = 0.6057
```

Residuals look Normally distributed, and high $p-$ value means we cannot reject Normality.

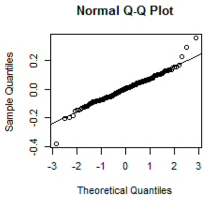
# Normality of residuals from AR(3) model of sqrt(hare) data




```
win.graph(width=2.5,height=2.5,pointsize=8)
qqnorm(residuals(m1.hare))
qqline(residuals(m1.hare))
shapiro.test(residuals(m1.hare))
#W = 0.9351, p-value = 0.06043
```

$p-$ value only just $> 0.05$, and, suspect looking extreme values but $n = 31$.

## Normality of residuals from IMA(1,1) model of log(oil.price) data




```
win.graph(width=2.5,height=2.5,pointsize=8)
qqnorm(residuals(m1.oil))
qqline(residuals(m1.oil))
shapiro.test(residuals(ml.oil))
#W = 0.9688, p-value = 3.937e-05
```

Low $p-$ value $\rightarrow$ Residuals not Normally distributed.

Outliers are prominent!

## Auto-correlation of residuals

To check for independence of noise terms in model, consider sample acf of residuals, $\hat{r}_k$.

We have seen in TSLecture6 that for true white noise, $r_k \sim \left(0, \frac{1}{n}\right)$.

Even in well-fitted models, residuals can be correlated. (Durbin & Watson)

For small lags, $k$ and $j$, $Var(\hat{r}_k)$ can be well $< \frac{1}{n}$ and estimates of $\hat{r}_k$ and $\hat{r}_j$ can be highly correlated.

For large lags, the approx. variance of $\frac{1}{n}$ applies and $\hat{r}_k$ and $\hat{r}_j$ are approx. uncorrelated.

## Auto-correlation of residuals cont.

For an AR(1) process,

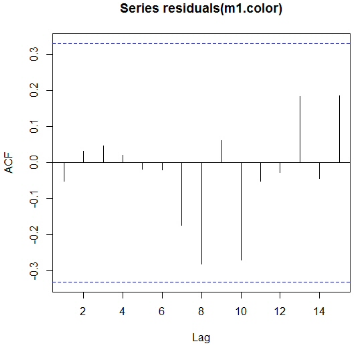
$$Var(\hat{r}_1) \approx \frac{\phi^2}{n}.$$

$$Var(\hat{r}_k) \approx \frac{1-(1-\phi^2)\phi^{2k-2}}{n} \qquad \text{for } k > 1.$$

In the AR(1) model for the color series, $\hat{\phi} = 0.57$ and $n = 35$, and the approximate standard deviations of the residual acf values are:

| Lag $k$ | 1 | 2 | 3 | 4 | 5 | $>5$ |
|---|---|---|---|---|---|---|
| $\sqrt{Var(\hat{r}_k)}$ | 0.096 | 0.149 | 0.163 | 0.167 | 0.168 | 0.169 |

# Sample acf of residuals from AR(1) model for Colour property




```
win.graph(width=4.875,height=3,pointsize=8)
acf(residuals(m1.color))
```

Dashed lines based on large lag standard error of $\pm\frac{2}{\sqrt{n}}$, and here there is no evidence of auto-correlation of the residuals.

## Auto-correlation of residuals cont.

For an AR(2) process,

$$Var(\hat{r}_1) \approx \frac{\phi_2^2}{n}.$$

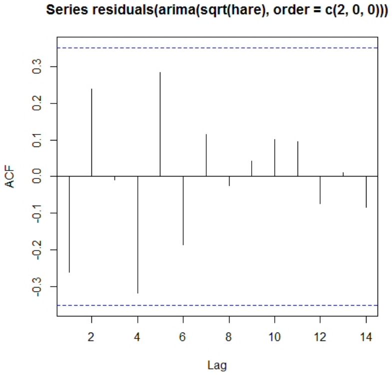
$$Var(\hat{r}_2) \approx \frac{\phi_2^2 + \phi_1^2(1+\phi_2)^2}{n}$$

$$Var(\hat{r}_k) \approx \frac{1}{n} \qquad \text{for } k \geq 3.$$

In the AR(2) model for the $\sqrt{hare}$ series, $\hat{\phi}_1 = 1.351$, $\hat{\phi}_2 = -0.776$ and $n = 31$.

Note here $\hat{r}_1 = -0.261$ and $2 \times \sqrt{Var(\hat{r}_1)} = 0.278$, so it is just within limits.

# Sample acf of residuals from AR(2) model for sqrt(hare) data




```
acf(residuals(arima(sqrt(hare),order=c(2,0,0))))
acf(residuals(arima(sqrt(hare),order=c(2,0,0))),
plot=F)$acf[1:6]
```

## Auto-correlation of residuals cont.

For MA(1) and MA(2) models, replace $\phi'$s with $\theta'$s and the same standard error limits apply.

ARMA standard error limits are complicated!

## Knowing the frequency helps

With monthly data, pay special attention to auto-correlation at lags 12, 24 36, etc.

With quarterly data, pay special attention to auto-correlation at lags 4, 8, 12, etc.

## Ljung-Box statistic

Accounts for size of auto-correlations summed over a number ($K$) of lags.

$$Q_* \sim \chi^2_{K-1},$$

$$= n(n+2)\left(\frac{\hat{r}_1^2}{n-1} + \frac{\hat{r}_2^2}{n-2} + \cdots + \frac{\hat{r}_K^2}{n-K}\right).$$

The Null Hypothesis is that the residual errors are uncorrelated.

If $H_0$ is true, the $p-$ value will be $> 0.05$.

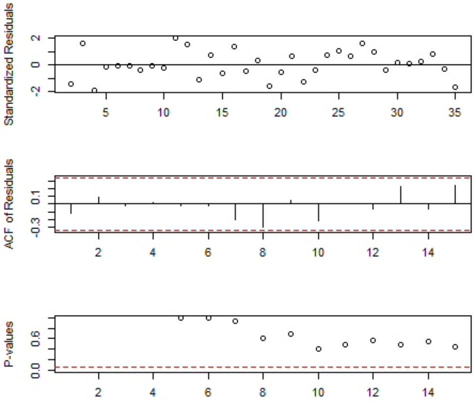
# Ljung-Box statistic for Colour property

```
round(acf(residuals(m1.color),plot=F)$acf[1:6],3)
LB.test(m1.color,lag=6)
#         Box-Ljung test
#data:  residuals from  m1.color
#X-squared = 0.2803, df = 5, p-value = 0.998
```

$$
\begin{aligned}
Q_* &\sim \chi^2_{K-1} \\
&= n(n+2)\left(\frac{\hat{r}_1^2}{n-1} + \frac{\hat{r}_2^2}{n-2} + \cdots + \frac{\hat{r}_K^2}{n-K}\right). \\
&= 35(35+2)\left\{\frac{-(0.051)^2}{35-1} + \frac{(0.032)^2}{35-2} + \cdots + \frac{(-0.019)^2}{35-6}\right\} \\
&\approx 0.28. \qquad (< CV : \chi^2_{(K-1=5,\alpha=0.05)} = 11.070)
\end{aligned}
$$

High $p-$ value means we cannot reject $H_0$ that residuals are uncorrelated.

# Diagnostic display for AR(1) model of Colour property




```
win.graph(width=4.875,height=4.5)
tsdiag(m1.color)
```

## Over-fitting and Parameter redundancy

Having fitted a supposedly adequate model, we fit a slightly more general model, e.g.,

if an AR(2) model seems appropriate, overfit with an AR(3) model.

Original AR(2) model confirmed if:

- $|\hat{\phi}_3|$ not significantly different from zero
- $\hat{\phi}_1$ and $\hat{\phi}_2$ do not change significantly from the AR(2) estimates

## Colour series AR(1)

```
m1.color=arima(color,order=c(1,0,0))
m1.color
#Coefficients:
#          ar1  intercept
#       0.5705    74.3293
#s.e.   0.1435     1.9151
#sigma^2 estimated as 24.83:
log likelihood = -106.07,
aic = 216.15
```

$\hat{\phi}_1$ is significant.

## Colour series AR(2) over-fit

```
arima(color,order=c(2,0,0))
#Coefficients:
#          ar1      ar2  intercept
#       0.5173   0.1005    74.1551
#s.e.   0.1717   0.1815     2.1463
#sigma^2 estimated as 24.6:
log likelihood = -105.92,  aic = 217.84 #(higher)
```

$\hat{\phi}_2$ is not significant and the $\hat{\phi}_1$ value has not changed much $\Rightarrow$ use an AR(1) model.

# Colour series ARMA(1,1) over-fit

```
arima(color,order=c(1,0,1))
#Coefficients:
#         ar1      ma1  intercept
#      0.6721  -0.1467    74.1730
#s.e.  0.2147   0.2742     2.1357
#sigma^2 estimated as 24.63:
log likelihood = -105.94,  aic = 217.88 #(higher)
```

$\uparrow$ standard errors. $\hat{\theta}$ not significant.

## Colour series ARMA(1,2) over-fit

```
arima(color,order=c(1,0,2))
#Coefficients:
#          ar1      ma1     ma2  intercept
#       0.6275  -0.1087  0.0603    74.1750
#s.e.   0.3090   0.3560  0.2231     2.1026
#sigma^2 estimated as 24.57:
#log likelihood = -105.9,  aic = 219.81  #(highest)
```

Do not ↑ the order of AR and MA parts simultaneously.

If, after fitting an MA(1), ∃ a lot of residual auto-correlation at lag 2, try an MA(2) not an ARMA(1,1) model.

Forecasting