Lecture 9

ARIMA Models

10/03/2018

$MA(\infty)$

$MA(q)$

From last time,

$$MA(q): \qquad y_t = \delta + w_t + \theta_1\, w_{t-1} + \theta_2\, w_{t-2} + \cdots + \theta_q\, w_{t-q}$$

Properties:

$$E(y_t) = \delta$$

$$\gamma(0) = Var(y_t) = (1 + \theta_1^2 + \theta_2 + \cdots + \theta_q^2)\, \sigma_w^2$$

$$\gamma(h) = \begin{cases} \theta_h + \theta_1\, \theta_{1+h} + \theta_2\, \theta_{2+h} + \cdots + \theta_{q-h}\, \theta_q & \text{if } h \in \{1, \dots, q\} \\ 0 & \text{otherwise} \end{cases}$$

and is stationary for any values of $(\theta_1, \dots, \theta_q)$

$MA(\infty)$

If we let $q \to \infty$ then process will be stationary if and only if the moving average coefficients (θ 's) are square summable, i.e.

$$\sum_{i=1}^{\infty} \theta_i^2 < \infty$$

which is necessary such that $Var(y_t) < \infty$ so that the weak stationarity are met.

Sometimes, a slightly stronger condition known as absolute summability,

$$\sum_{i=1}^{\infty} |\theta_i| < \infty$$

, is necessary (e.g. for some CLT related asymptotic results).

Invertibility

If an $MA(q)$ process, $y_t = \delta + \theta_q(L)w_t$, can be rewritten as a stationary AR process then the process is said to be invertible.

$MA(1)$ w/ $\delta = 0$ example:

$$y_t = w_t + \theta w_{t-1}$$

$$
\begin{aligned}
w_t &= y_t - \theta w_{t-1} \\
&= y_t - \theta\left(y_{t-1} - \theta w_{t-2}\right) \\
&= y_t - \theta y_{t-1} + \theta^2 w_{t-2} \\
&= y_t - \theta y_{t-1} + \theta^2 y_{t-2} - \theta^3 w_{t-3} \\
&\;\;\vdots \\
&= y_t + \sum_{i=1}^{\infty} (-\theta)^i y_{t-i} = y_t + \sum_{i=1}^{p} (-\theta)^i y_{t-i} + \cancel{(-\theta)^{p+1} w_{t-p-1}} \to 0
\end{aligned}
$$

stationary iff $|\theta| < 1$

Invertibility vs Stationarity

A $MA(q)$ process is *invertible* if $y_t = \delta + \theta_q(L)\, w_t$ can be rewritten as an exclusively AR process (of possibly infinite order), i.e. $\phi(L)\, y_t = \alpha + w_t$.

Invertibility vs Stationarity

A $MA(q)$ process is *invertible* if $y_t = \delta + \theta_q(L)\, w_t$ can be rewritten as an exclusively AR process (of possibly infinite order), i.e. $\phi(L)\, y_t = \alpha + w_t$.

Conversely, an $AR(p)$ process is *stationary* if $\phi_p(L)\, y_t = \delta + w_t$ can be rewritten as an exclusively MA process (of possibly infinite order), i.e. $y_t = \delta + \theta(L)\, w_t$.

Invertibility vs Stationarity

A $MA(q)$ process is *invertible* if $y_t = \delta + \theta_q(L)\, w_t$ can be rewritten as an exclusively AR process (of possibly infinite order), i.e. $\phi(L)\, y_t = \alpha + w_t$.

Conversely, an $AR(p)$ process is *stationary* if $\phi_p(L)\, y_t = \delta + w_t$ can be rewritten as an exclusively MA process (of possibly infinite order), i.e. $y_t = \delta + \theta(L)\, w_t$.

So using our results w.r.t. $\phi(L)$ it follows that if all of the roots of $\theta_q(L)$ are outside the complex unit circle then the moving average process is invertible.

AR(1): $y_t = \phi y_{t-1} + w_t$

$$y_1 = v_1$$

$$y_2 = \phi y_{t-1} + w_2 = \phi v_1 + w_2$$

$$y_3 = \phi y_2 + w_3 = \phi^2 v_1 + \phi w_2 + w_3$$

$$\vdots$$

$$y_t = v_t + \sum_{i=1}^{t-1} \phi^i v_i$$

Differencing

Difference operator

We will need to define one more notational tool for indicating differencing

$$\Delta y_t = y_t - y_{t-1}$$

Difference operator

We will need to define one more notational tool for indicating differencing

$$\Delta y_t = y_t - y_{t-1}$$

Just like the lag operator we will indicate repeated applications of this operator using exponents

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

Difference operator

We will need to define one more notational tool for indicating differencing

$$\Delta y_t = y_t - y_{t-1}$$

Just like the lag operator we will indicate repeated applications of this operator using exponents

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

Note that Δ can even be expressed in terms of the lag operator L,

$$\Delta^d = (1 - L)^d$$

Differencing and Stocastic Trend

Using the two component time series model

$$y_t = \mu_t + x_t$$

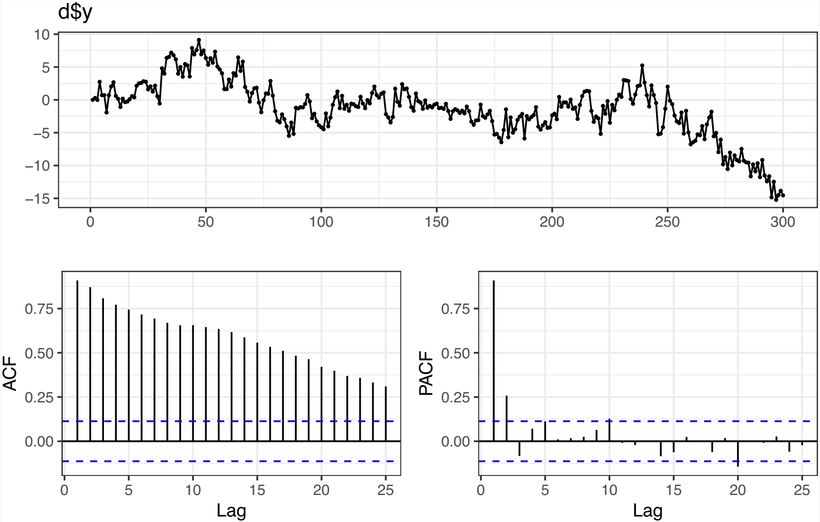
where μ_t is a non-stationary trend component and x_t is a mean zero stationary component.

We have already shown that differencing can address deterministic trend (e.g. $\mu_t = \beta_0 + \beta_1\, t$). In fact, if μ_t is any k-th order polynomial of t then $\Delta^k y_t$ is stationary.

Differencing can also address stochastic trend such as in the case where μ_t follows a random walk.

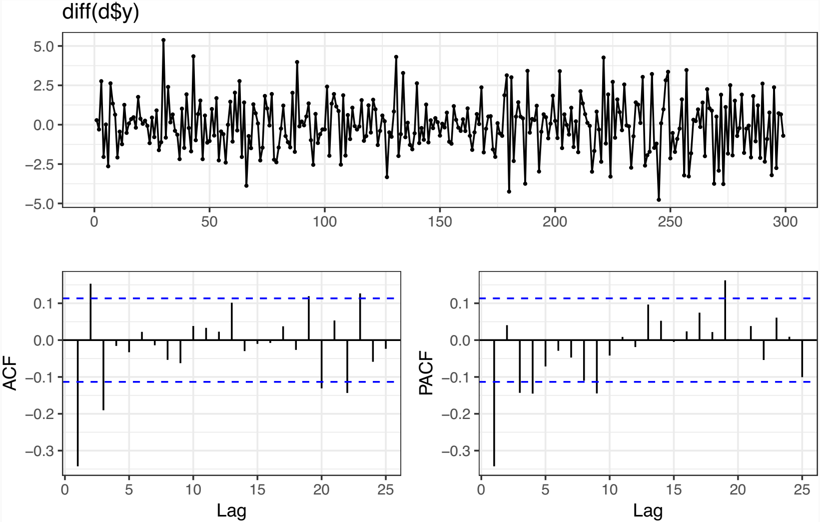
Stochastic trend - Example 1

Let $y_t = \mu_t + w_t$ where w_t is white noise and $\mu_t = \mu_{t-1} + v_t$ with v_t being a stationary process with mean 0.



Differenced stochastic trend

```
forecast::ggtsdisplay(diff(d$y))
```



Stationary?

Is y_t stationary?

$$y_t = \mu_t + w_t \qquad \mu_t = \mu_{t-1} + v_t$$

$$\mu_0 = 0$$
$$\mu_1 = v_1$$
$$\mu_2 = v_1 + v_2$$
$$\mu_3 = v_1 + v_2 + v_3$$
$$\vdots$$
$$\mu_t = \sum_{i=1}^{t} v_i$$

$$\boxed{y_t = \sum_{i=1}^{t} v_i + w_t}$$

① $E(y_t) = 0$

② $$\begin{aligned} Var(y_t) &= Var\left(\sum v_i\right) + Var(w_t) \\ &= \sum_{i=1}^{t} \sigma_v^2 + \sigma_w^2 \\ &= t\,\sigma_v^2 + \sigma_w^2 \end{aligned}$$

$$\lim_{t \to \infty} Var(y_t) = \infty$$

✗

Difference Stationary?

Is Δy_t stationary?

$$\boxed{y_t = \sum_{i=1}^{t} v_i + w_t}$$

$$\Delta y_t = y_t - y_{t-1}$$

$$= \left(\cancel{\sum_{i=1}^{t} v_i} + w_t \right) - \left(\cancel{\sum_{i=1}^{t-1} v_i} + w_{t-1} \right)$$

$$= v_t + v_t - v_{t-1}$$

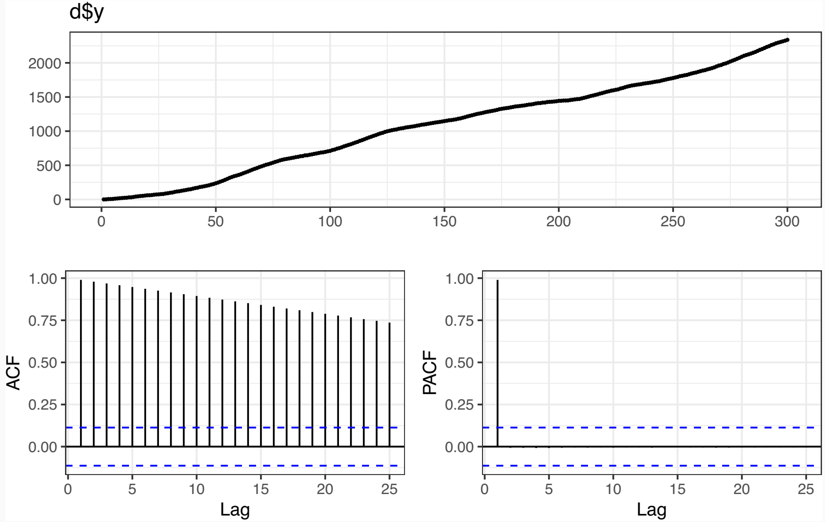
① $E(\Delta y_t) = 0$

② $\mathrm{Var}(\Delta y_t) = \sigma_v^2 + 2\sigma_w^2$

③ $$\gamma(h) = \begin{cases} \sigma_v^2 + 2\sigma_w^2 & \text{if } h = 0 \\ \sigma_w^2 & \text{if } |h| = 1 \\ 0 & \text{if } |h| > 1 \end{cases}$$

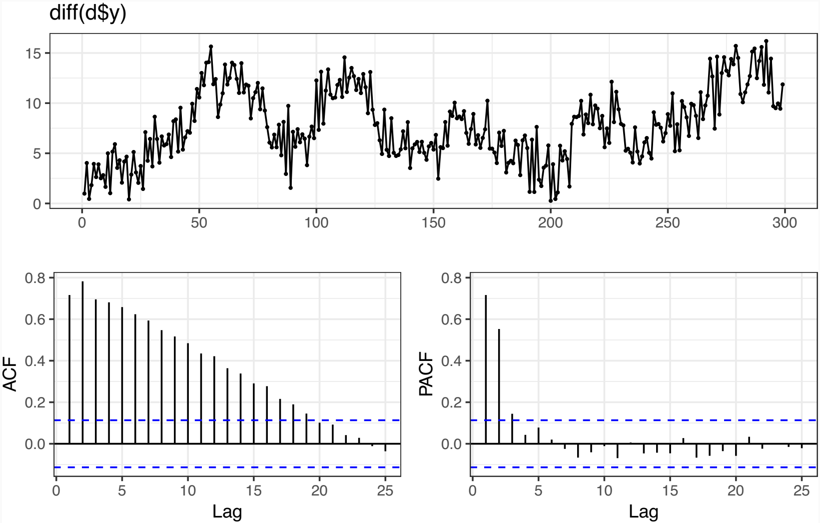
Stochastic trend - Example 2

Let $y_t = \mu_t + w_t$ where w_t is white noise and $\mu_t = \mu_{t-1} + v_t$ but now $v_t = v_{t-1} + e_t$ with e_t being stationary.



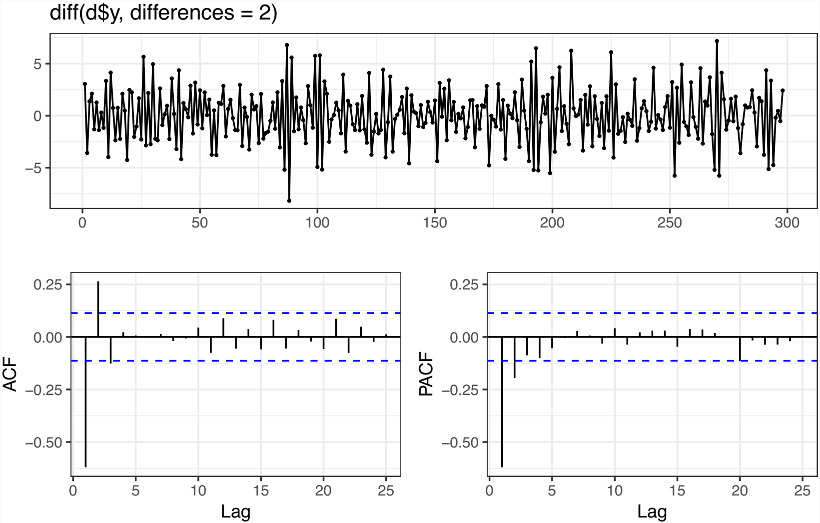
Differenced stochastic trend

```
forecast::ggtsdisplay(diff(d$y))
```



Twice differenced stochastic trend

```
forecast::ggtsdisplay(diff(d$y,differences = 2))
```



Difference stationary?

Is Δy_t stationary?

$$\boxed{V_t = V_{t-1} + e_t}$$

$$V_0 = 0$$

$$V_1 = e_1$$

$$V_2 = e_1 + e_2$$

$$V_3 = e_1 + e_2 + e$$

$$\vdots$$

$$V_t = \sum_{i=1}^{t} e_i$$

$$\boxed{\begin{aligned}\mu_t &= \mu_{t-1} + V_t \\ &= \mu_{t-1} + \sum_{i=1}^{t} e_i\end{aligned}}$$

$$\mu_0 = 0$$

$$\mu_1 = \mu_0 + e_1 = e_1$$

$$\mu_2 = \mu_1 + e_2 + e_1 = 2e_1 + e_2$$

$$\mu_3 = 3e_1 + 2e_2 + e_3$$

$$\vdots$$

$$\mu_t = \sum_{i=1}^{t} (t+1-i)\, e_i$$

$$Y_t = \sum_{i=1}^{t} (t+1-i)\, e_i + w_t$$

(1) $E(Y_t) = 0$

$\cancel{(2)}$ $Var(Y_t) < \infty$

Difference stationary?

Is Δy_t stationary?

$$y_t = \sum_{i=1}^{t} (t+1-i)\, e_i + w_t$$

(1) $E(\Delta y_t) = 0$

(2) $\mathrm{Var}(\Delta y_t) = t\sigma_e^2 + 2\sigma_v^2$

$$\Delta y_t = y_t - y_{t-1}$$

$$= (\mu_t + w_t) - (\mu_{t-1} + w_{t-1})$$

$$= (\mu_t - \mu_{t-1}) + (v_t - v_{t-1})$$

$$= (\cancel{\mu_{t-1}} + v_t - \cancel{\mu_{t-1}}) + (w_t - w_{t-1})$$

$$= \sum_{i=1}^{t} e_i + (v_t - v_{t-1})$$

2nd order difference stationary?

What about $\Delta^2 y_t$, is it stationary? $\Delta y_t = v_t + w_t - w_{t-1}$

$$\Delta^2 y_t = (v_t + w_t - w_{t-1}) - (v_{t-1} + w_{t-1} - v_{t-2})$$

$$= (v_t - v_{t-1}) + (w_t - 2w_{t-1} + w_{t-2})$$

$$= (\cancel{v_{t-1}} + e_t - \cancel{v_{t-1}}) + (w_t - 2w_{t-1} + w_{t-2})$$

$$= e_t + (w_t - 2w_{t-1} + w_{t-2})$$

(1) $E(\Delta^2 y_t) = 0$

(2) $Var(\Delta^2 y_t) = \sigma_e^2 + \sigma_w^2 + 4\sigma_w^2 + \sigma_w^2 = \sigma_e^2 + 6\sigma_w^2$

(3) $Cov(\Delta^2 y_t, \Delta^2 y_{t-h}) = ?$

differencing

$ARIMA$

$ARIMA$ Models

Autoregressive integrated moving average are just an extension of an $ARMA$ model to include differencing of degree d to y_t before including the autoregressive and moving average components.

$$ARIMA(p, d, q): \qquad \phi_p(L)\, \Delta^d\, y_t = \delta + \theta_q(L) w_t$$

$ARIMA$ Models

Autoregressive integrated moving average are just an extension of an $ARMA$ model to include differencing of degree d to y_t before including the autoregressive and moving average components.

$$ARIMA(p, d, q): \qquad \phi_p(L)\, \Delta^d\, y_t = \delta + \theta_q(L) w_t$$

Box-Jenkins approach:

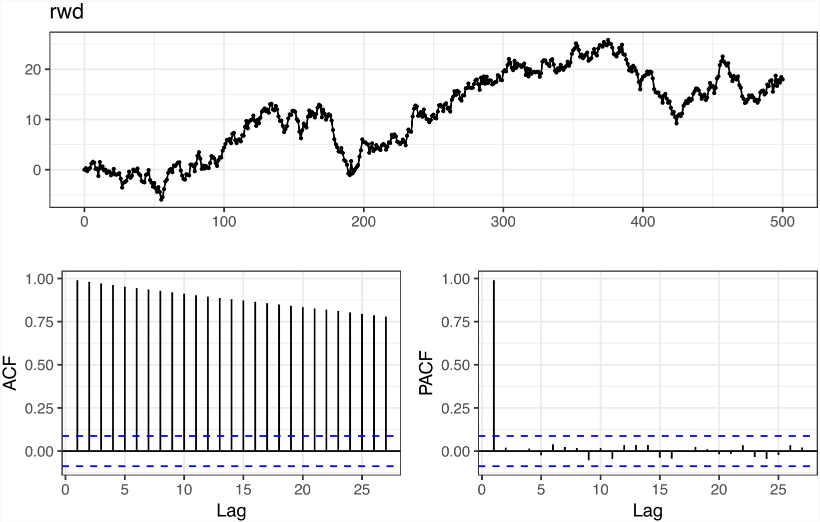
1. Transform data if necessary to stabilize variance
2. Choose order (p, d, q) of ARIMA model
3. Estimate model parameters (δ, ϕs, and θs) + σ_w^2
4. Diagnostics

Using `forecast` - random walk with drift

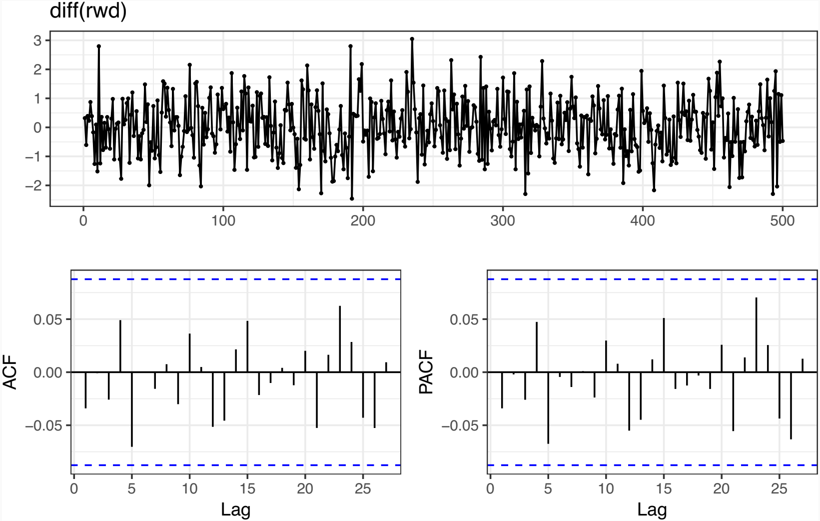
Some of R's base timeseries handling is a bit wonky, the `forecast` package offers some useful alternatives and additional functionality.

```
rwd = arima.sim(n=500, model=list(order=c(0,1,0)), mean=0.1)

forecast::Arima(rwd, order = c(0,1,0), include.constant = TRUE)
## Series: rwd
## ARIMA(0,1,0) with drift
##
## Coefficients:
##        drift
##       0.0359
## s.e.  0.0435
##
## sigma^2 estimated as 0.95:  log likelihood=-696.15
## AIC=1396.31   AICc=1396.33   BIC=1404.74
```

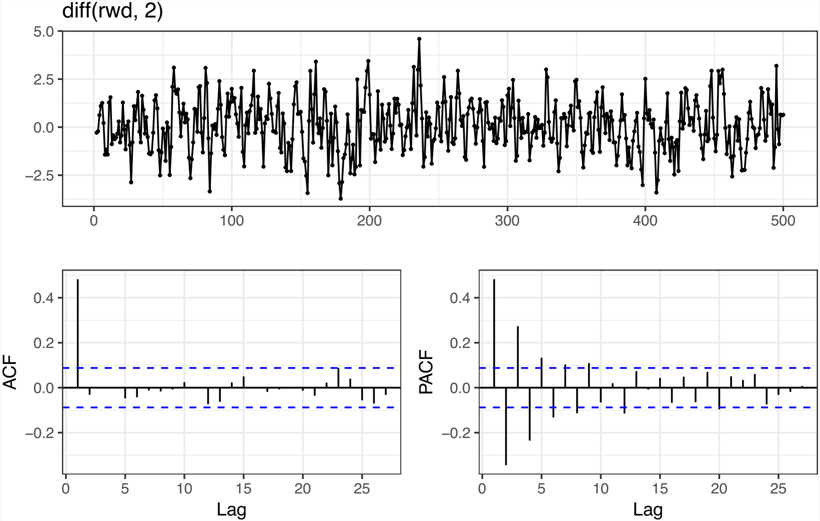
EDA


rwd
20
10
0
0
100
200
300
400
500
1.00
0.75
0.50
0.25
0.00
ACF
0
5
10
15
20
25
Lag
1.00
0.75
0.50
0.25
0.00
PACF
0
5
10
15
20
25
Lag

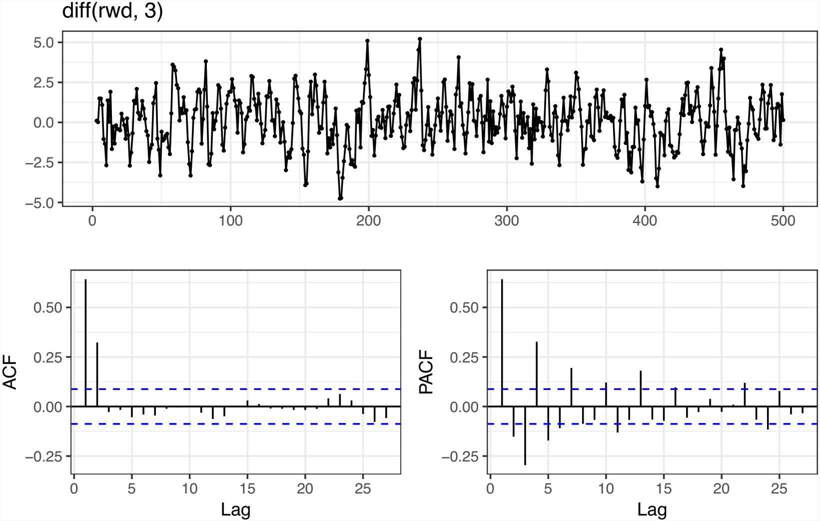
Differencing - Order 1



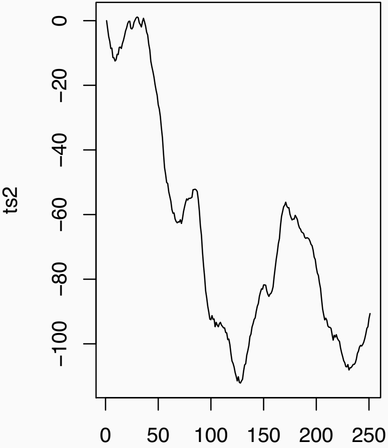
Differencing - Order 2



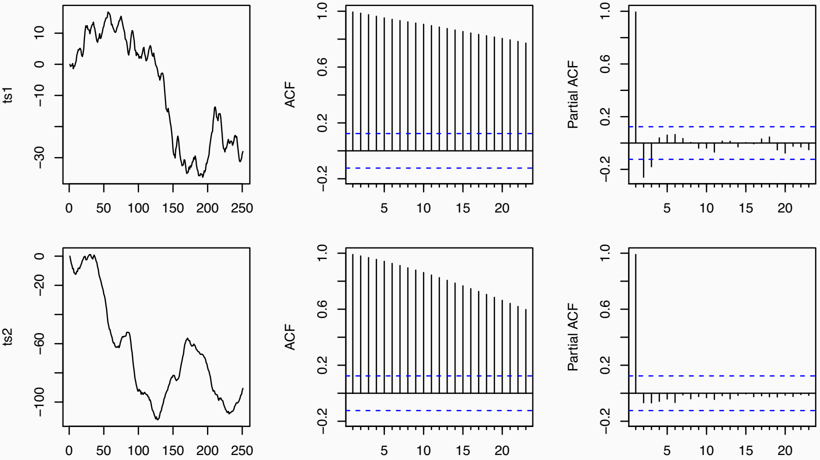
Differencing - Order 3



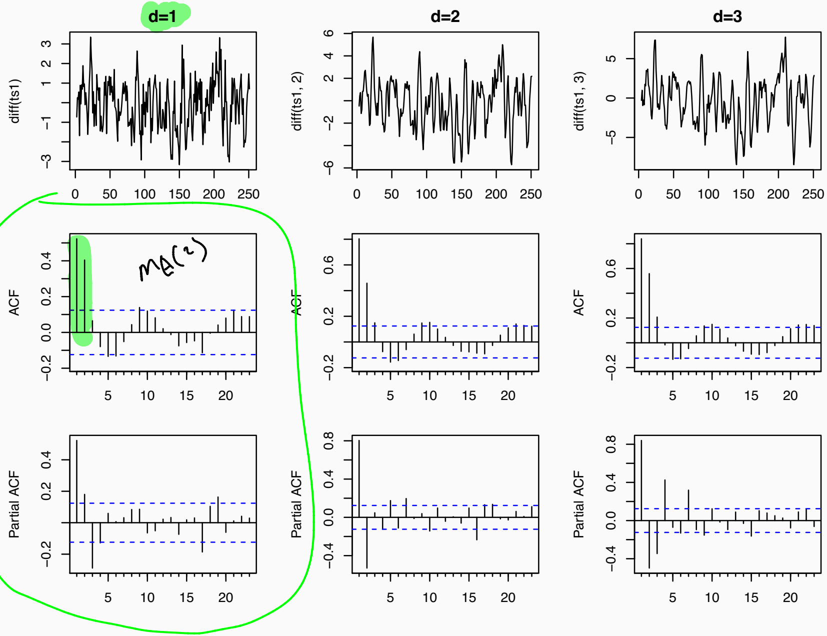
AR or MA - Time Series 1?



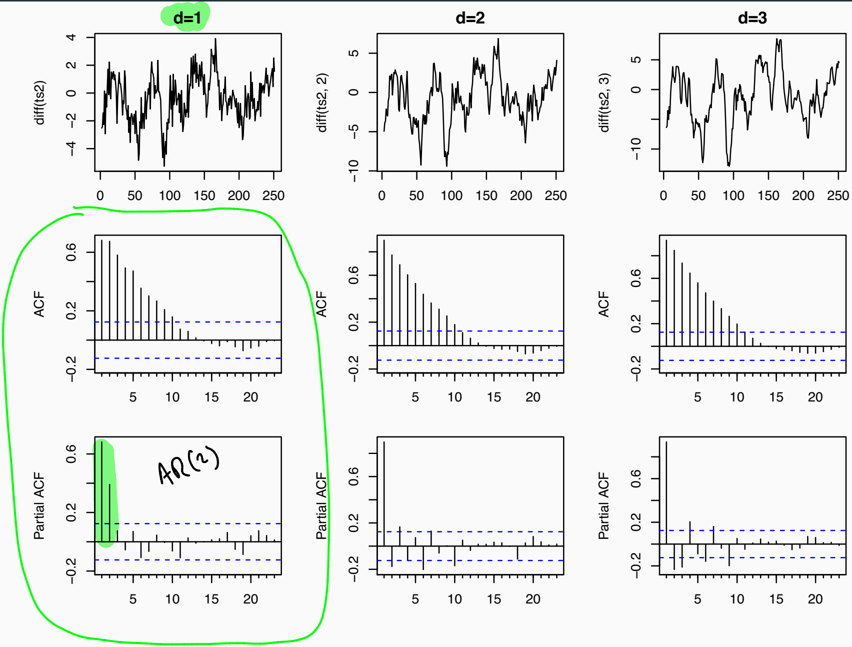
EDA



ts1 - Finding d



ts2 - Finding d



`ts1` - Models

p	d	q	aic	aicc	bic
0	1	0	830.71	830.73	834.23
1	1	0	751.57	751.62	758.61
2	1	0	745.00	745.10	755.56
0	1	1	784.70	784.75	791.74
1	1	1	749.63	749.73	760.20
2	1	1	739.39	739.55	753.47
0	1	2	726.90	727.00	737.47
1	1	2	723.19	723.35	737.27
2	1	2	725.14	725.39	742.75

p	d	q	aic	aicc	bic
0	1	0	967.07	967.09	970.60
1	1	0	800.69	800.74	807.74
2	1	0	756.04	756.14	766.61
0	1	1	885.78	885.83	892.82
1	1	1	762.34	762.43	772.90
2	1	1	756.77	756.93	770.85
0	1	2	822.62	822.72	833.19
1	1	2	756.88	757.05	770.97
2	1	2	758.50	758.74	776.10

ts1 - final model

Fitted:

```
forecast::Arima(ts1, order = c(0,1,2))
## Series: ts1
## ARIMA(0,1,2)
##
## Coefficients:
##          ma1     ma2
##       0.3823  0.5217
## s.e.  0.0522  0.0613
##
## sigma^2 estimated as 1.052:  log likelihood=-360.45
## AIC=726.9   AICc=727   BIC=737.47
```

Truth:

```
ts1 = arima.sim(n=250, model=list(order=c(0,1,2), ma=c(0.4,0.5)))
```

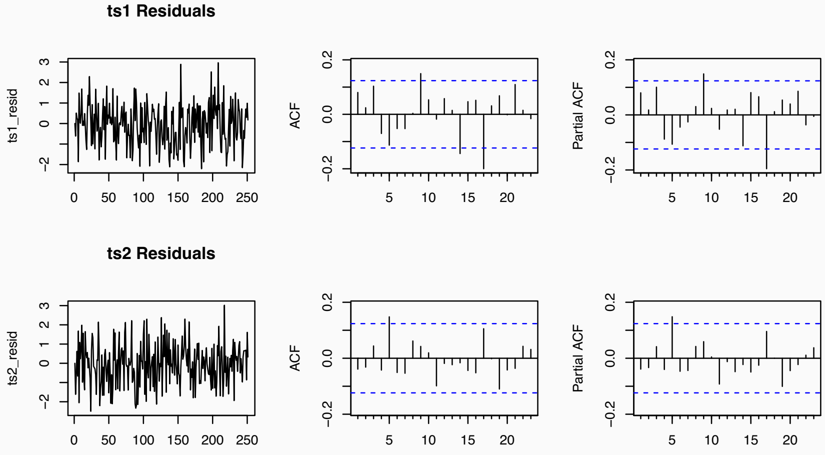
Fitted:

```
forecast::Arima(ts2, order = c(2,1,0))
## Series: ts2
## ARIMA(2,1,0)
##
## Coefficients:
##          ar1     ar2
##       0.4148  0.4148
## s.e.  0.0573  0.0578
##
## sigma^2 estimated as 1.181:  log likelihood=-375.02
## AIC=756.04   AICc=756.14   BIC=766.61
```

Truth:

```
ts2 = arima.sim(n=250, model=list(order=c(2,1,0), ar=c(0.4,0.5)))
```

Residuals


ts1 Residuals
ts1_resid
ACF
Partial ACF
ts2 Residuals
ts2_resid
ACF
Partial ACF

Automatic model selection

`ts1`:

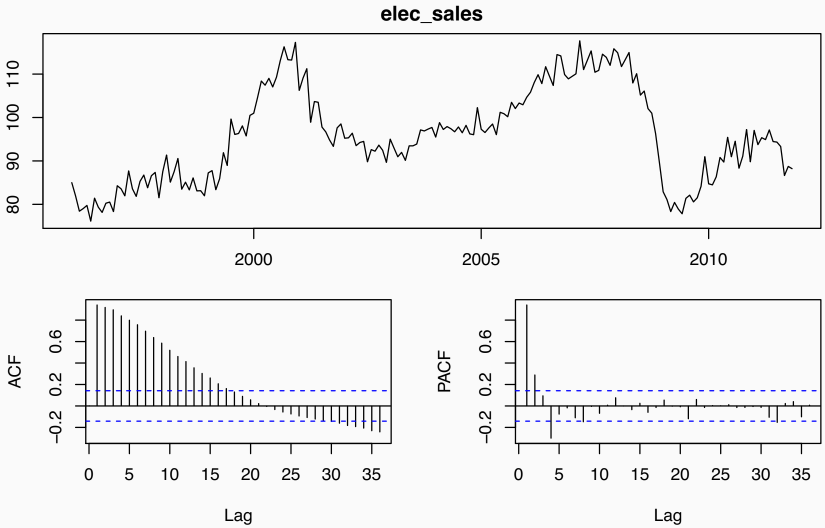
```
forecast::auto.arima(ts1)
## Series: ts1
## ARIMA(1,1,2)
##
## Coefficients:
##          ar1     ma1     ma2
##       0.2804  0.1838  0.4512
## s.e.  0.1199  0.1092  0.0736
##
## sigma^2 estimated as 1.033:  log likelihood=-357.59
## AIC=723.19   AICc=723.35   BIC=737.27
```

`ts2`:

```
forecast::auto.arima(ts2)
## Series: ts2
## ARIMA(1,2,1)
##
## Coefficients:
##           ar1      ma1
##       -0.3052  -0.2661
## s.e.   0.1184   0.1217
##
## sigma^2 estimated as 1.227:  log likelihood=-377.92
## AIC=761.84   AICc=761.94   BIC=772.4
```

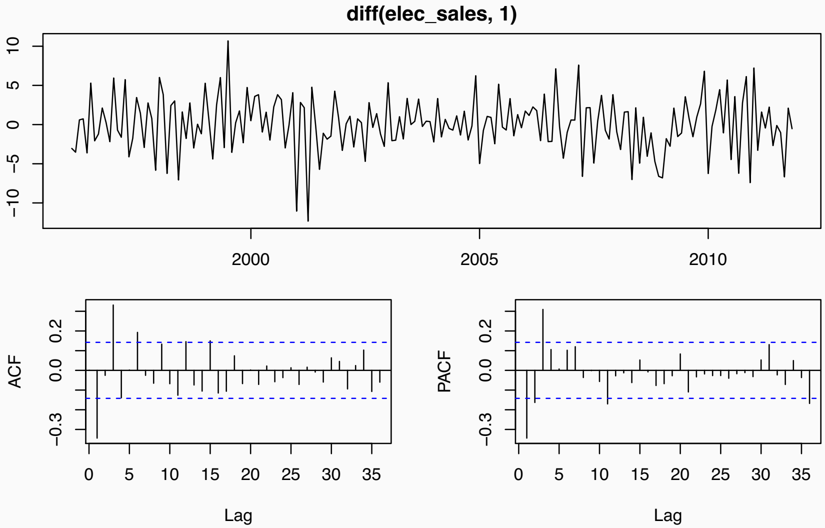
Electrical Equipment Sales

Data

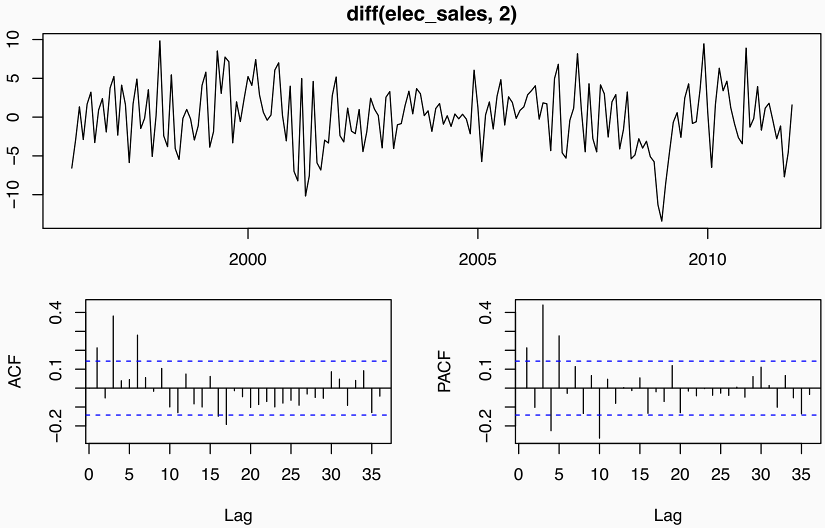


elec_sales
80
90
100
110
2000
2005
2010
ACF
-0.2
0.2
0.6
0
5
10
15
20
25
30
35
Lag
PACF
-0.2
0.2
0.6
0
5
10
15
20
25
30
35
Lag

1st order differencing



2nd order differencing

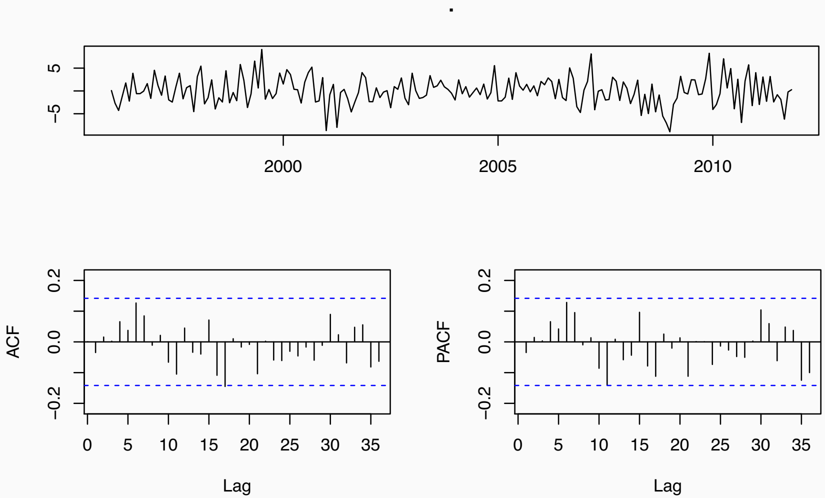


Model

```
forecast::Arima(elec_sales, order = c(3,1,0))
## Series: elec_sales
## ARIMA(3,1,0)
##
## Coefficients:
##           ar1      ar2     ar3
##       -0.3488  -0.0386  0.3139
## s.e.   0.0690   0.0736  0.0694
##
## sigma^2 estimated as 9.853:  log likelihood=-485.67
## AIC=979.33   AICc=979.55   BIC=992.32
```

Residuals

```r
forecast::Arima(elec_sales, order = c(3,1,0)) %>% residuals() %>%
  forecast::tsdisplay(points=FALSE)
```



Model Comparison

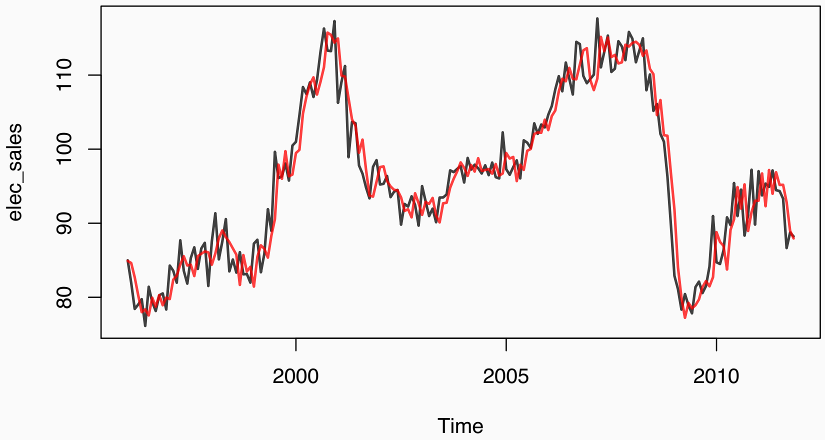
Model choices:

```
forecast::Arima(elec_sales, order = c(3,1,0))$aicc
## [1] 979.5477

forecast::Arima(elec_sales, order = c(3,1,1))$aicc
## [1] 978.4925

forecast::Arima(elec_sales, order = c(4,1,0))$aicc
## [1] 979.2309

forecast::Arima(elec_sales, order = c(2,1,0))$aicc
## [1] 996.8085
```

Automatic selection:

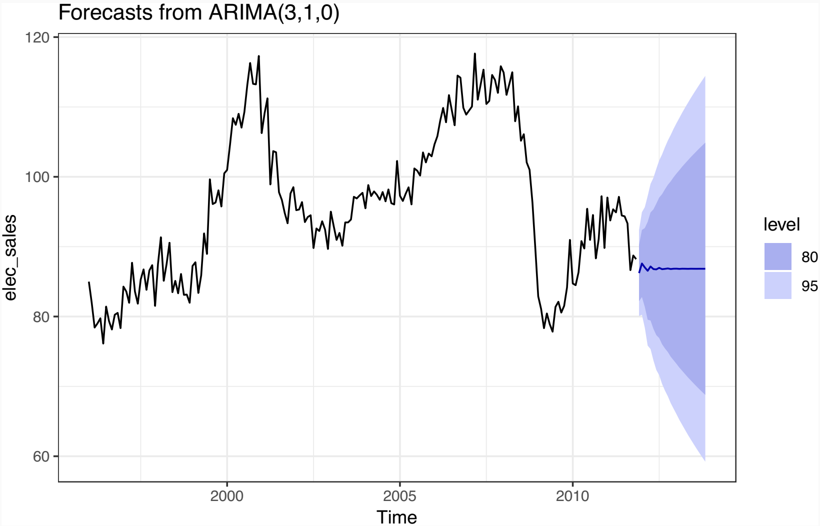
```
forecast::auto.arima(elec_sales)
## Series: elec_sales
## ARIMA(3,1,1)
##
## Coefficients:
##          ar1     ar2     ar3      ma1
##       0.0519  0.1191  0.3730  -0.4542
## s.e.  0.1840  0.0888  0.0679   0.1993
##
## sigma^2 estimated as 9.737:  log likelihood=-484.08
## AIC=978.17   AICc=978.49   BIC=994.4
```

Model fit



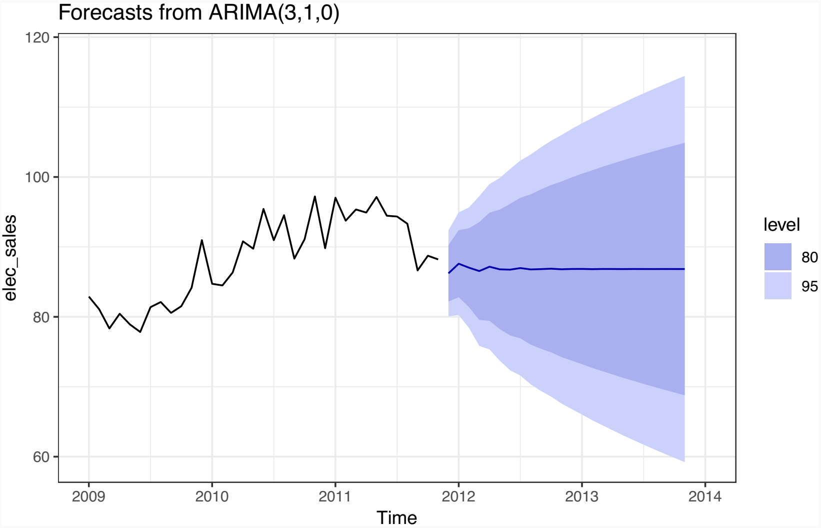
Model forecast

```
forecast::Arima(elec_sales, order = c(3,1,0)) %>%
  forecast::forecast() %>% autoplot()
```



Model forecast - Zoom

```
forecast::Arima(elec_sales, order = c(3,1,0)) %>%
  forecast::forecast() %>% autoplot() + xlim(2009,2014)
```



General Guidance

1. Positive autocorrelations out to a large number of lags usually indicates a need for differencing
2. Slightly too much or slightly too little differencing can be corrected by adding AR or MA terms respectively.
3. A model with no differencing usually includes a constant term, a model with two or more orders (rare) differencing usually does not include a constant term.
4. After differencing, if the PACF has a sharp cutoff then consider adding AR terms to the model.
5. After differencing, if the ACF has a sharp cutoff then consider adding an MA term to the model.
6. It is possible for an AR term and an MA term to cancel each other's effects, so try models with one fewer AR term and one fewer MA term.

Based on rules from https://people.duke.edu/~rnau/411arim2.htm and