

Lecture 9

ARIMA Models

Colin Rundel

02/15/2017

$MA(\infty)$

From last time,

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

Properties:

$$E(y_t) = \delta$$

$$\text{Var}(y_t) = (1 + \theta_1^2 + \theta_2^2 + \cdots + \theta_q^2) \sigma_w^2$$

$$\text{Cov}(y_t, y_{t+h}) = \begin{cases} \theta_h + \theta_1 \theta_{1+h} + \theta_2 \theta_{2+h} + \cdots + \theta_{q-h} \theta_q & \text{if } |h| \leq q \\ 0 & \text{if } |h| > q \end{cases}$$

and is stationary for any values of θ_i

If we let $q \rightarrow \infty$ then process will still be stationary if the moving average coefficients (θ 's) are square summable,

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

since this is necessary for $\text{Var}(y_t) < \infty$.

Sometimes, a slightly strong condition called absolute summability, $\sum_{i=1}^{\infty} |\theta_i| < \infty$, is necessary (e.g. for some CLT related asymptotic results).

Invertibility

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

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

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 is invertible.

Differencing

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

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

Δ can also be expressed in terms of the lag operator L ,

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

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 stationary as well. Is Δy_t stationary?

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. Is Δy_t stationary? What about $\Delta^2 y_t$, is it stationary?

ARIMA

Autoregressive integrated moving average are just an extension of an ARMA model to include differencing of degree d to y_t , which is most often used to address trend in the data.

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

Autoregressive integrated moving average are just an extension of an ARMA model to include differencing of degree d to y_t , which is most often used to address trend in the data.

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

Box-Jenkins approach:

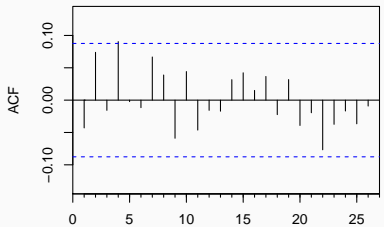
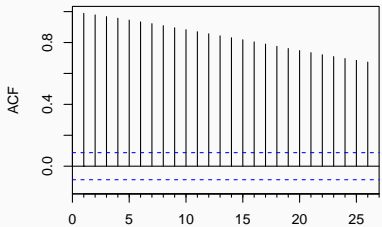
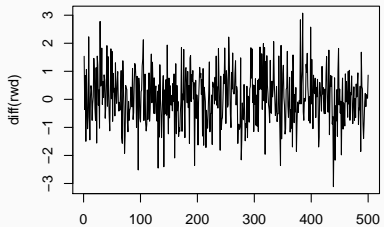
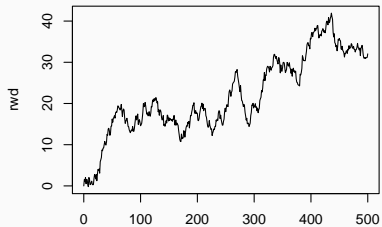
1. Transform data if necessary to stabilize variance
2. Choose order (p , d , and q) of ARIMA model
3. Estimate model parameters (ϕ s and θ s)
4. Diagnostics

Using forecast - random walk with drift

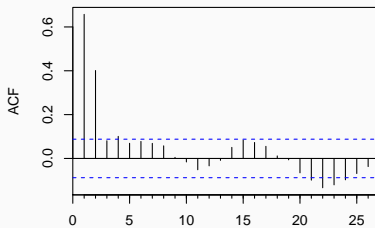
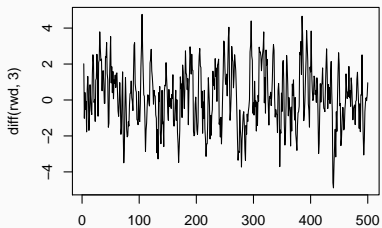
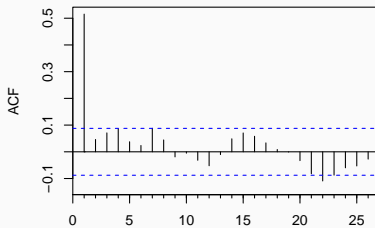
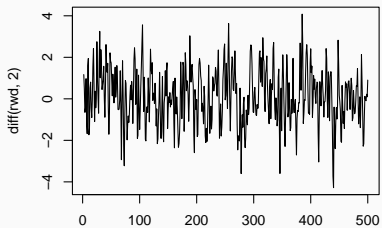
Some of R's base timeseries handling is a bit wonky, the `forecast` package offers some useful alternatives and additional functionality.

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

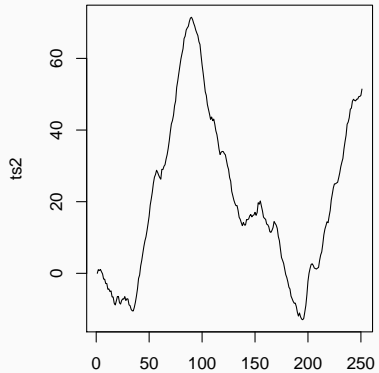
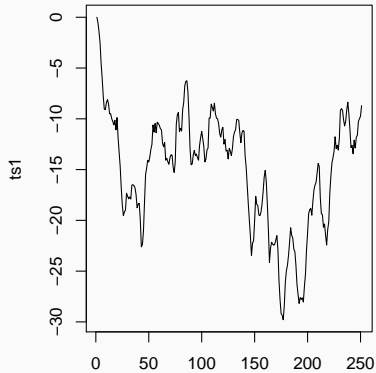
library(forecast)
Arima(rwd, order = c(0,1,0), include.constant = TRUE)
## Series: rwd
## ARIMA(0,1,0) with drift
##
## Coefficients:
##      drift
##      0.0641
## s.e.  0.0431
##
## sigma^2 estimated as 0.9323:  log likelihood=-691.44
## AIC=1386.88   AICc=1386.91   BIC=1395.31
```

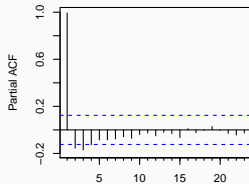
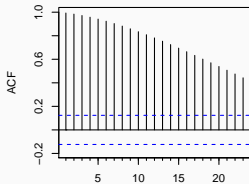
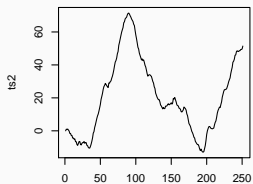
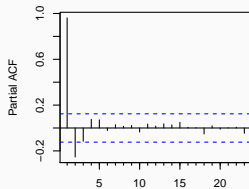
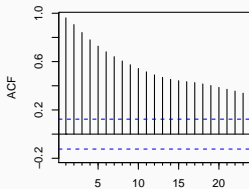
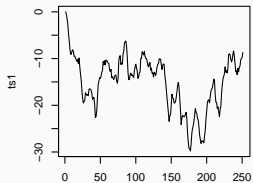


Over differencing

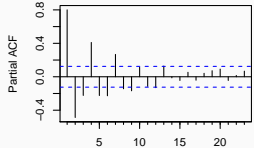
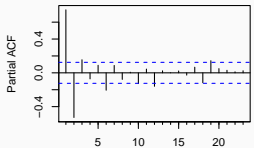
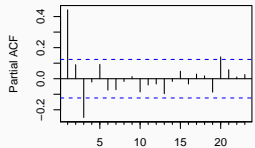
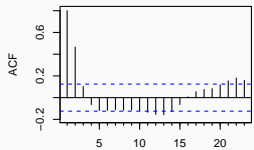
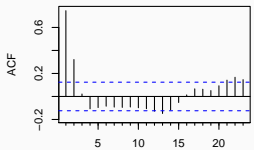
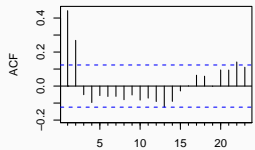
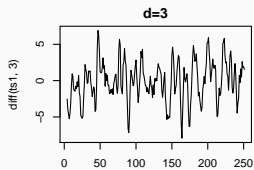
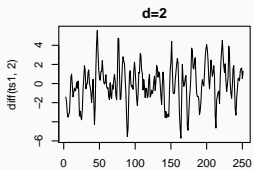
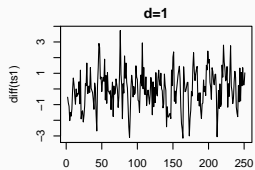


AR or MA?

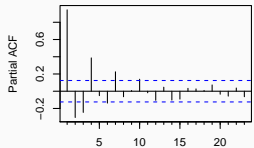
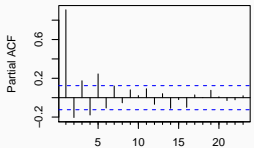
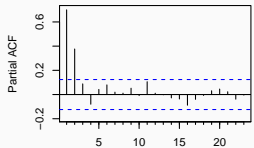
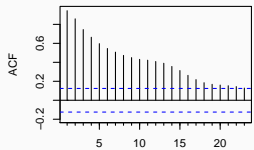
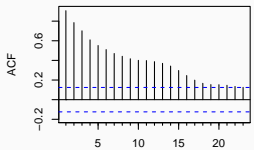
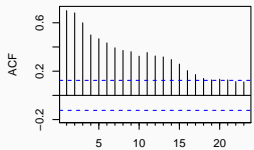
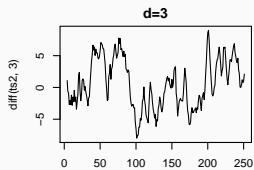
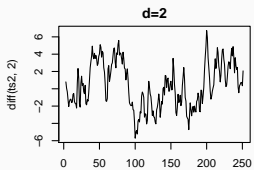
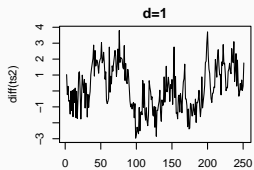




ts1 - Finding d



ts2 - Finding d



p	d	q	AIC	BIC
0	1	2	729.43	740.00
1	1	2	731.23	745.31
2	1	2	731.57	749.18
2	1	1	744.29	758.38
2	1	0	747.55	758.12
1	1	0	747.61	754.65
1	1	1	748.65	759.21
0	1	1	764.98	772.02
0	1	0	800.43	803.95

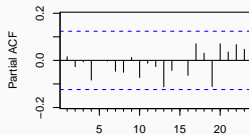
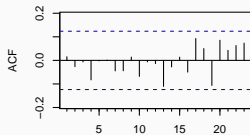
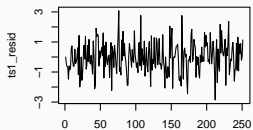
p	d	q	AIC	BIC
2	1	0	683.12	693.68
1	1	2	683.25	697.34
2	1	1	683.83	697.92
2	1	2	685.06	702.67
1	1	1	686.38	696.95
1	1	0	719.16	726.20
0	1	2	754.66	765.22
0	1	1	804.44	811.48
0	1	0	890.32	893.85

```
Arima(ts1, order = c(0,1,2))  
## Series: ts1  
## ARIMA(0,1,2)  
##  
## Coefficients:  
##          ma1      ma2  
##      0.4138  0.4319  
## s.e.  0.0547  0.0622  
##  
## sigma^2 estimated as 1.064: log likelihood=-361.72  
## AIC=729.43  AICc=729.53  BIC=740
```

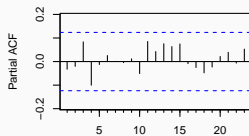
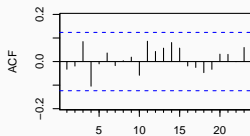
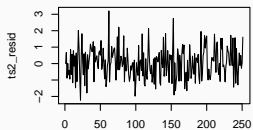
```
Arima(ts2, order = c(2,1,0))  
## Series: ts2  
## ARIMA(2,1,0)  
##  
## Coefficients:  
##          ar1      ar2  
##      0.4392  0.3770  
## s.e.  0.0587  0.0587  
##  
## sigma^2 estimated as 0.8822:  log likelihood=-338.56  
## AIC=683.12  AICc=683.22  BIC=693.68
```

Residuals

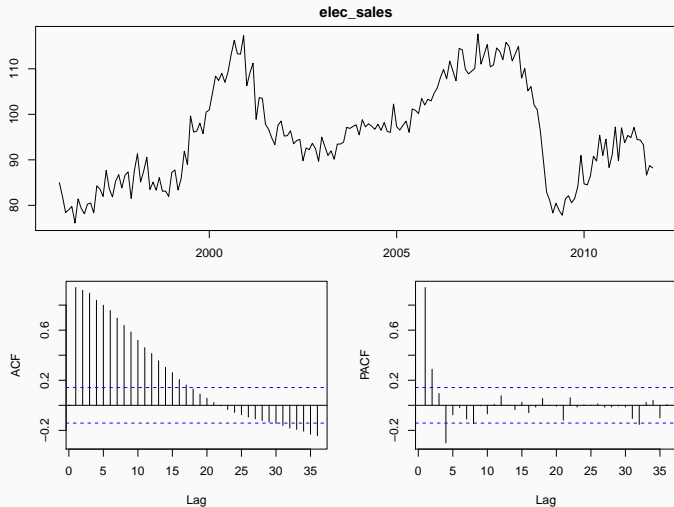
ts1 Residuals



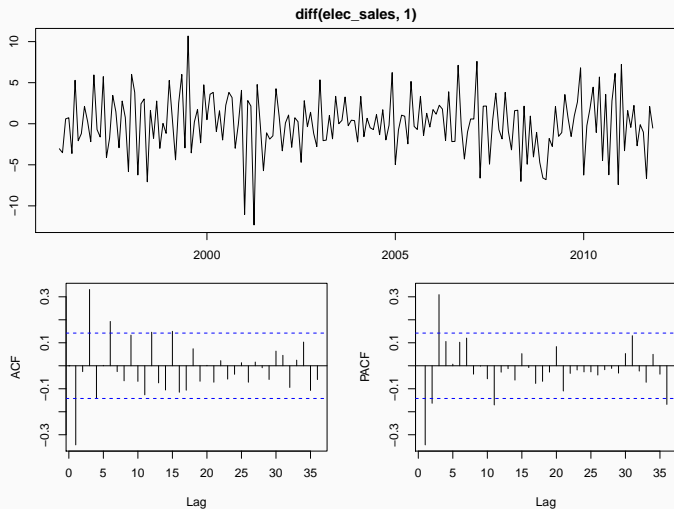
ts2 Residuals



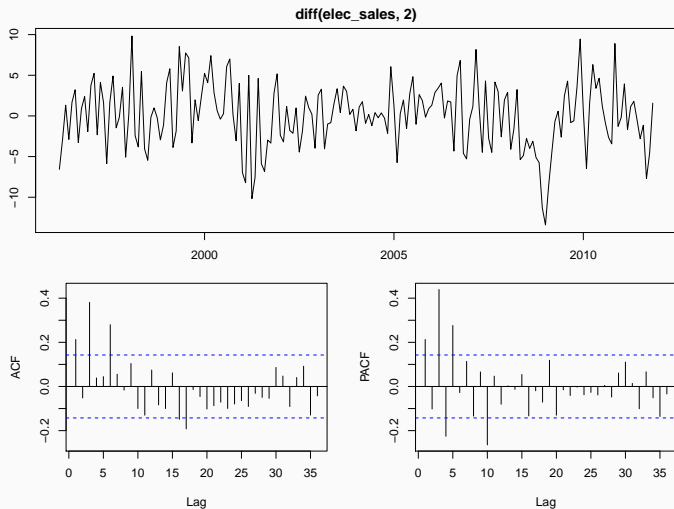
Electrical Equipment Sales



1st order differencing



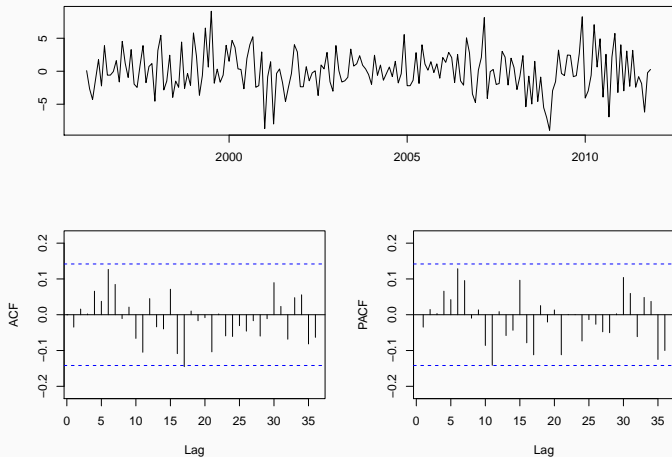
2nd order differencing




```
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

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



```
Arima(elec_sales, order = c(3,1,0))$aic  
## [1] 979.3314
```

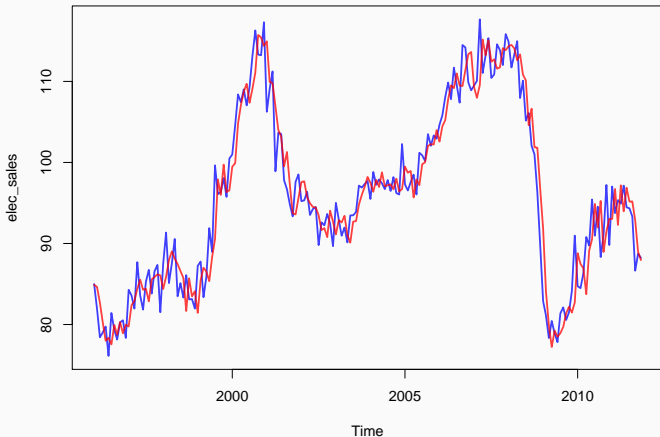
```
Arima(elec_sales, order = c(3,1,1))$aic  
## [1] 978.1664
```

```
Arima(elec_sales, order = c(4,1,0))$aic  
## [1] 978.9048
```

```
Arima(elec_sales, order = c(2,1,0))$aic  
## [1] 996.6795
```

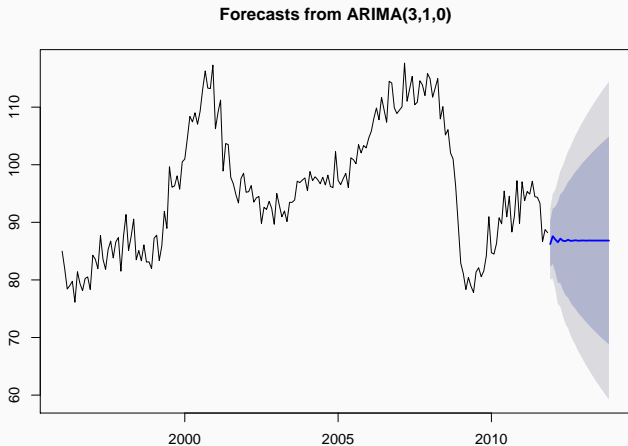
Model fit

```
plot(elec_sales, lwd=2, col=adjustcolor("blue", alpha.f=0.75))  
Arima(elec_sales, order = c(3,1,0)) %>% fitted() %>% lines(col=adjustcolor("red", alpha.f=0.75))
```



Model forecast

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



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 <https://people.duke.edu/~rnau/411arim3.htm>