UNIVERSITY of LIMERICK
OLLSCOIL LUIMNIGH

Time Series Analysis

MS 4218

joseph.lynch@ul.ie

Outline

- Seasonal Arima models

Seasonal Arima model

In TSLecture3 we encountered the seasonal means model and the cosine model.

These models included a deterministic time trend and the coefficients of this time trend were determined using the lm function.

In many cases, the assumption of deterministic trend is suspect and consequently the residuals from these models are often highly auto-correlated at several lags.

Seasonal Arima models are Arima models with a stochastic time trend.

Seasonal Arima form

The seasonal and non-seasonal parts of an ARIMA model have the same type of structure:

either may have an AR factor, an MA factor, and/or an order of differencing.

The model is classified as an ARIMA$(p, d, q)x(P, D, Q)$ model, where $P =$number of seasonal autoregressive (SAR) terms, $D =$number of seasonal differences, $Q =$number of seasonal moving average (SMA) terms.

In the seasonal part of the model, all of these factors operate across multiples of lag s (the number of periods in a season).

Seasonal Arima model cont.

In identifying a seasonal model, the first step is to determine whether or not a seasonal difference is needed, in addition to or perhaps instead of a non-seasonal difference.

Look at time series plots and acf and pacf plots, if necessary, for all possible combinations of 0 or 1 non-seasonal difference and 0 or 1 seasonal difference.

Never use more than 1 seasonal difference, nor more than 2 total differences (seasonal and non-seasonal combined).

If the seasonal pattern is both strong and stable over time (e.g., high in the summer and low in winter, or vice versa), then use a seasonal difference regardless of whether a non-seasonal difference is used.

This will prevent the seasonal pattern from "dying out" in the long-term forecasts.

Seasonal Arima model cont.

The signature of pure SAR or pure SMA behavior is similar to the signature of pure AR or pure MA behavior, except that the pattern appears across multiples of lag s in the acf and pacf,

e.g., a pure SAR(1) process has spikes in the acf at lags s, $2s$, $3s$, etc., while the pacf cuts off after lag s.

A pure SMA(1) process has spikes in the pacf at lags s, $2s$, $3s$, etc., while the acf cuts off after lag s.

An SAR signature usually occurs when the autocorrelation at the seasonal period is positive, whereas an SMA signature usually occurs when the seasonal autocorrelation is negative.

Do not mix SAR and SMA terms in the same model, and avoid using more than one of either kind.

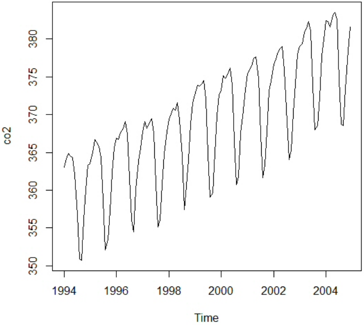
Seasonal Arima model cont.

Although a seasonal ARIMA model has only a few parameters, forecasting requires the estimation of one or two seasons' worth of implicit parameters to initialize it, and usually 4 or 5 seasons of data are required to fit a seasonal ARIMA model.

The most commonly used seasonal Arima model is the $(0, 1, 1) \times (0, 1, 1)$ model, i.e., an MA(1)$\times$SMA(1) model with both a seasonal and a non-seasonal difference.

When seasonal ARIMA models are fitted to logged data, they are capable of tracking a multiplicative seasonal pattern.

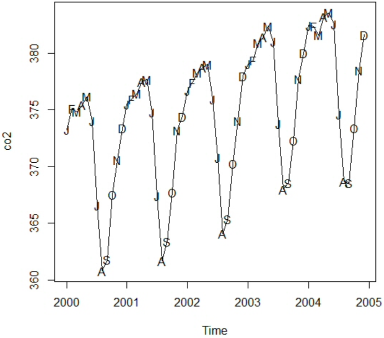
Carbon dioxide levels in Northern Canada (Jan1994;Dec2004)




```
data(co2)
plot(co2,ylab="co2")
```

Upward trend ⇒ consider a non-stationary model.

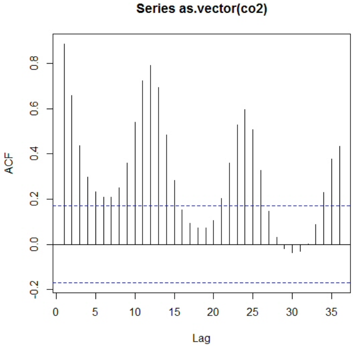
Seasonal component




```
plot(window(co2,start=c(2000,1)),ylab="co2")
Month=c("J","F","M","A","M","J","J","A","S","O","N","D")
points(window(co2,start=c(2000,1)),pch=Month)
```

Co2 levels are higher in winter.

Auto-correlation function

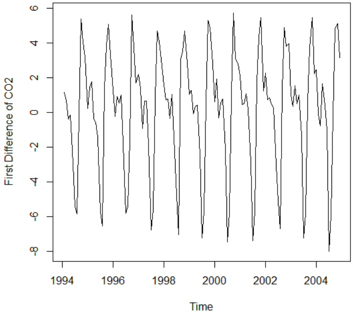



```
acf(as.vector(co2),lag.max=36)
```

Strong correlation at lags 12, 24, 36 etc.

Other correlations also present.

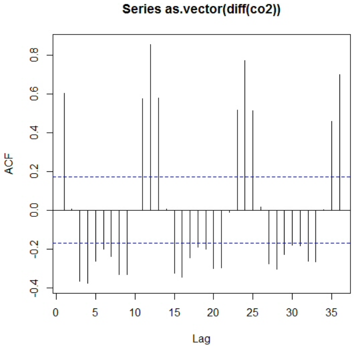
Time plot of first difference




```
plot(diff(co2), xlab="Time",
ylab="First Difference of CO2" )
```

Trend has gone.

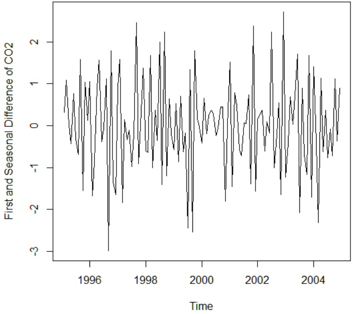
Acf of first difference




```
acf(as.vector(diff(co2)),lag.max=36)
```

Seasonality still present, so try seasonal differencing.

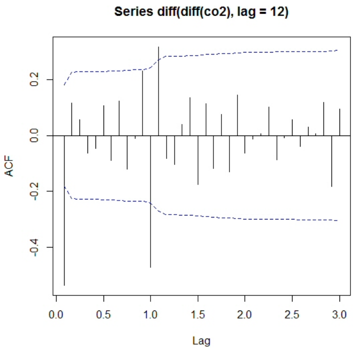
Time plot of seasonal difference




```
plot(diff(diff(co2),lag=12),
ylab="First and Seasonal Difference of CO2",
xlab="Time")
```

Seasonality has now disappeared.

Acf of seasonal difference




```
acf(diff(diff(co2),lag=12),lag.max=36,
ci.type="ma")
```

Little auto-correlation left.

? a simple model which incorporates lags 1 and 12 auto-correlations.

Arima (0,1,1)$\times(0,1,1)_{12}$

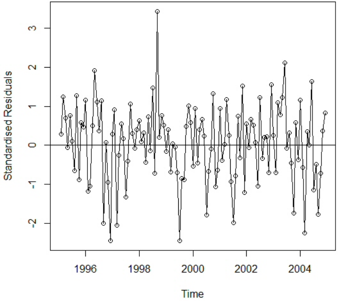
```
m1.co2=arima(co2,order=c(0,1,1),
                 seasonal=list(order=c(0,1,1),
                                   period=12))
m1.co2

#Coefficients:
#          ma1     sma1
#      -0.5792  -0.8206
#s.e.   0.0791   0.1137

#sigma^2 estimated as 0.5446:
#log likelihood = -139.54,  aic = 283.08
```

Both coefficients are highly significant.

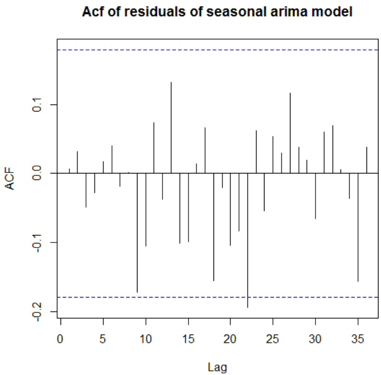
Diagnostic checking $(0,1,1) \times (0,1,1)_{12}$ model




```
plot(window(rstandard(m1.co2),
start=c(1995,2)),
ylab="Standardised Residuals",
type="o");abline(h=0)
```

A little central irregularity, but nothing major.

Acf of residuals of $(0,1,1) \times (0,1,1)_{12}$ model




```
acf(as.vector(window(rstandard(m1.co2),
start=c(1995,2))),
lag.max=36)
```

Only one significant auto-correlation, at lag 22.

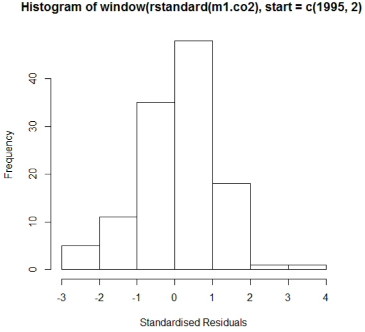
Ljung-Box test of residual independence

H_0: the residuals are un-correlated.

```
LB.test(m1.co2)
#Box-Ljung test
#data:  residuals from  m1.co2
#X-squared = 7.051, df = 10,
#p-value = 0.7206
```

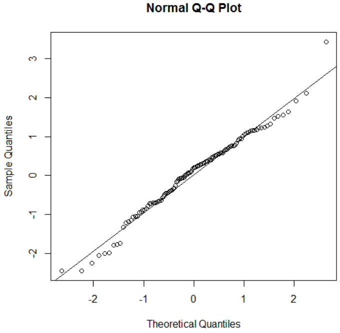
High p value, $\Rightarrow$ do not reject H_0, i.e., model has captured dependence.

Test of Normality of residuals




```
hist(window(rstandard(m1.co2),
start=c(1995,2)),
xlab="Standardised Residuals")
```

Bell-shaped, but not perfectly so.




```
qqnorm(window(rstandard(m1.co2),
start=c(1995,2))
qqline(window(rstandard(m1.co2),
start=c(1995,2))
```

Outlier in upper tail.

Shapiro-Wilk test of Normality

```
Shapiro.test(window(rstandard(m1.co2),
start=c(1995,2)))

#data:  window(rstandard(m1.co2),
#start = c(1995, 2))
#W = 0.982, p-value = 0.1134
```

High $p-$ value means we cannot reject H_0, thus the residuals are Normally distributed.

Over-fitted model

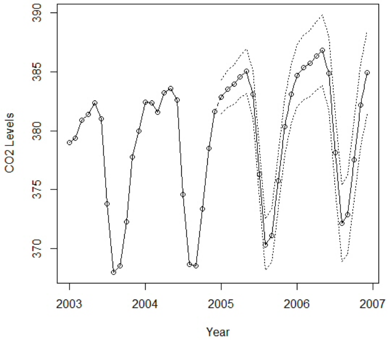
```
m2.co2=arima(c02, order=c(0,1,2),
seasonal=list(order=c(0,1,1),period=12))
m2.co2

#Coefficients:
#           ma1      ma2     sma1
#       -0.5714  -0.0165  -0.8274
#s.e.    0.0897   0.0948   0.1224

#sigma^2 estimated as 0.5427:
#log likelihood = -139.52,  aic = 285.05
```

Little change in ma1 and sma1 estimates. Se similar.
ma2 not significantly different from 0.
Aic has increased.

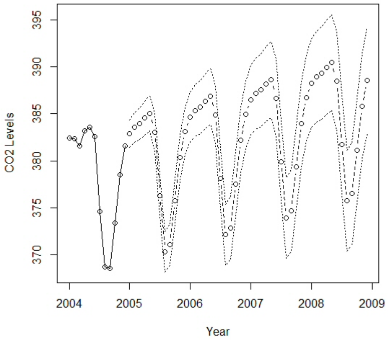
2-year forecast and limits for CO_2 model




```
plot(m1.co2,n1=c(2003,1),n.ahead=24,
xlab="Year", type="o", ylab="CO2 Levels")
```

Forecasts mimic stochastic periodicity. Limits are narrow.

Long term forecast for CO_2 model




```
plot(m1.co2,n1=c(2004,1),n.ahead=48,
xlab="Year", type="b", ylab="CO2 Levels")
```

Limits get wider as time goes on.

Generic Seasonal Arima Formula

```
get.best.series<-function(tsdata, maxord=c(1,1,1,1,1,1))
{ best.aic<-1e8
  for(p in maxord[1]) for(d in maxord[2]) for(q in maxord[3])
  for(P in maxord[4]) for(D in maxord[5]) for(Q in maxord[6])
    {
     fit<-arima(tsdata, order=c(p,d,q),
                        seas=list(order=c(P,D,Q),
                        frequency(tsdata))
                )
     fit.aic<--2*fit$loglik+(log(n)+1)*length(fit$coef)
        if(fit.aic<best.aic)
           {best.aic<-fit.aic
            best.fit<-fit
            best.model<-c(p,d,q,P,D,Q)
           }
     }
  list(best.aic,best.fit,best.model)
}
```

Next

Time Series in the Frequency domain