Models for Non-Stationary & Seasonal Time Series

Kushal Kharel

10/23/2021

Models for NonStationary & Seasonal Time Series

Suppose we have a time series model of the form,

$$Y_t = \mu_t + X_t$$

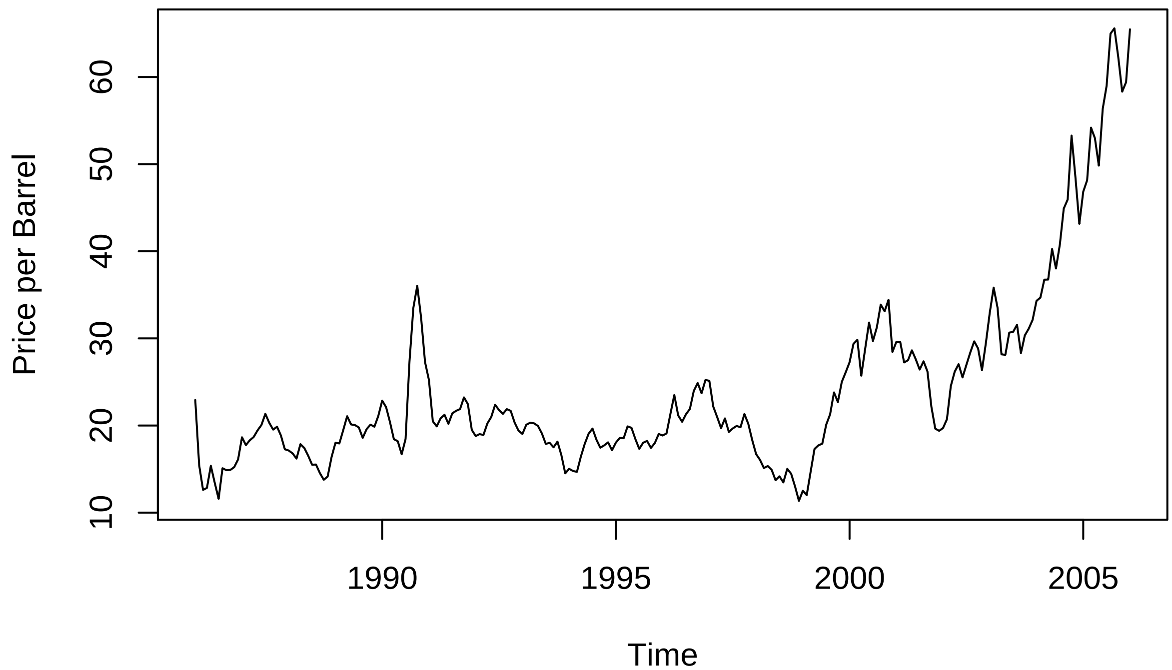
where μ_t is is not a constant mean and X_t is a zero-mean stationary series. We know that any time series without a constant mean is nonstationary. We can apply a technique of differencing to achieve stationarity. For later discussion of this topic, we will consider a dataset of Monthly Proce of Oil from January 1986 to January 2006 taken from TSA package, springer textbook. We will just see the plot here and we will come back to this later on this notes.

```
## Warning: package 'TSA' was built under R version 4.0.5

##
## Attaching package: 'TSA'

## The following objects are masked from 'package:stats':
##
##     acf, arima

## The following object is masked from 'package:utils':
##
##     tar
```



We can see from the plot above that this time series displays variation across time specially from 2001 and it is reasonable to think that we cannot fit the stationary model in this case.

Let us consider the AR(1) model,

$$Y_t = \phi Y_{t-1} + e_t$$

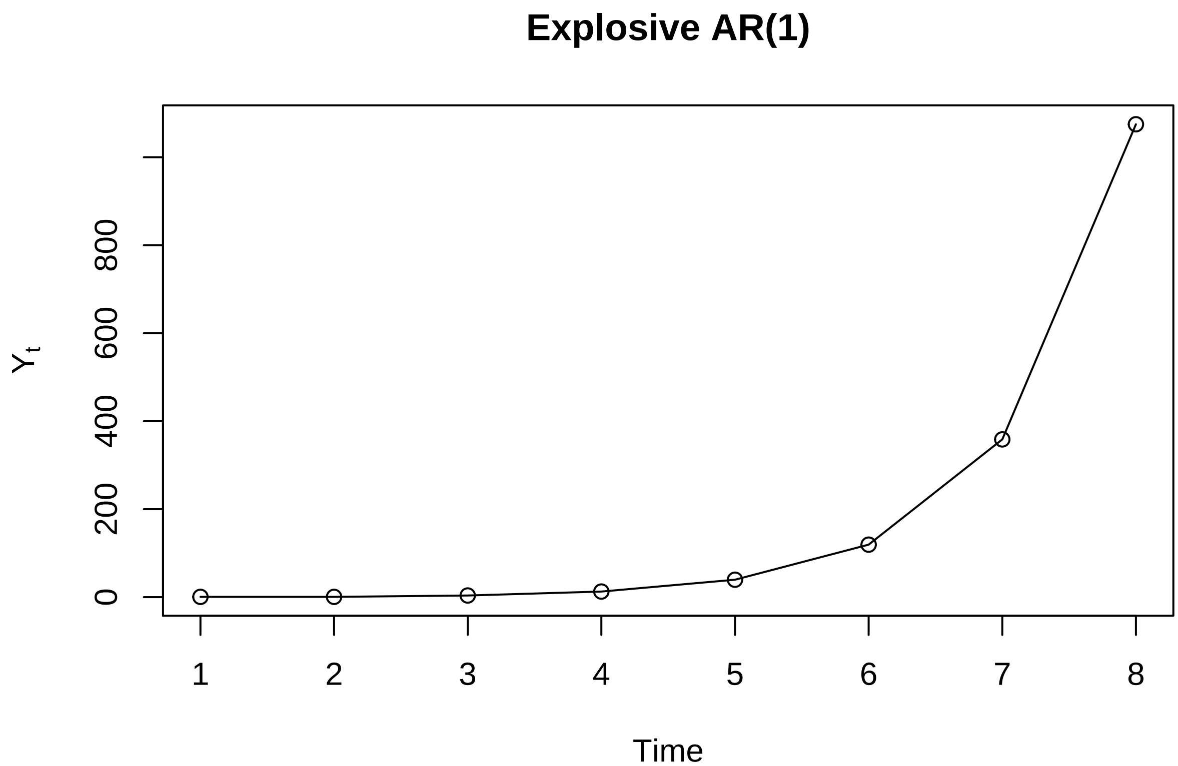
where, e_t is a true "innovation" that is e_t is uncorrelated with $Y_{t-1}, Y_{t-2}, \ldots$. So, we must have $|\phi| < 1$. Let us consider that ϕ in above equation is 3 then we have a equation of form,

$$Y_t = 3Y_{t-1} + e_t$$

If we iterate into the past we obtain,

$$Y_t = e_t + 3e_{t-1} + 3^2 e_{t-2} + ... + 3^{t-1} e_1 + 3^t Y_0$$

We can see that distant past values does indeed have an effect and the weights applied to Y_0 and e_t grows exponentially large. The process is explosive. Let us plot this model and see how it looks like:



This explosive behavior can be seen in model's variance and covariance functions which is given below;

$$Var(Y_t) = \frac{1}{8}(9^t - 1)\sigma_e^2$$
$$Cov(Y_t, Y_{t-k}) = \frac{3^k}{8}(9^{t-k} - 1)\sigma_e^2$$
$$Corr(Y_t, Y_{t-k}) = 3^k\sqrt{\frac{9^{t-k} - 1}{9^t - 1}} \approx 1$$

for large t and moderate k.

Now, consider the above model when $|\phi| = 1$, then the equation becomes,

$$Y_t = Y_{t-1} + e^t$$

We can rewrite it as follows:

$$\nabla Y_t = e^t$$

where, $\nabla Y_t = Y_t - Y_{t-1}$ is the first difference of Y_t

Suppose,

$$Y_t = M_t + X_t$$

where M_t is a series that is changing only slowly over time and it can be either deterministic or stochastic. If we assume that M_t is approximately constant over every two consecutive time points, we can predict M_t at t by choosing β_0 so that,

$$\sum_{j=0}^{1}(Y_{t-j} - \beta_{0,t})^2$$

is minimized which leads to

$$\hat{M}_t = \frac{1}{2}(Y_t + Y_{t-1})$$

and the "de-trended series at time t is then,

$$Y_t - \hat{M}_t = Y_t - \frac{1}{2}(Y_t + Y_{t-1}) = \frac{1}{2}(Y_t - Y_{t-1}) = \frac{1}{2}\nabla Y_t$$

Now, suppose that M_t is stochastic and changes slowly over time which is governed by random walk model.

$$Y_t = M_t + e_t$$

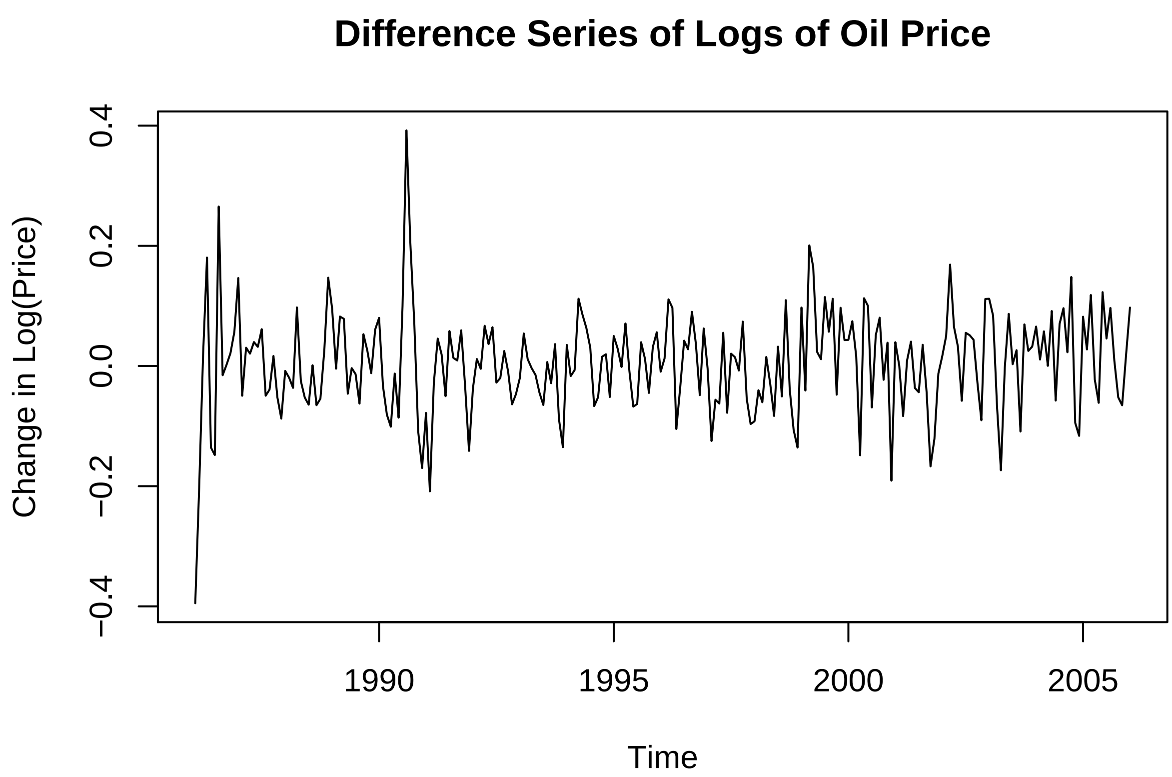
with $M_t = M_{t-1} + \epsilon_t$ where e_t and ϵ_t are independent white noise series. Then,

$$\nabla Y_t = \nabla M_t + \nabla e_t = \epsilon_t + e_t - e_{t-1}$$

which will have the acf of MA(1) series with

$$\rho_1 = -\{\frac{1}{[2 + (\frac{\sigma_\epsilon^2}{\sigma_e^2})]}\}$$

Now, let us apply the differencing to our oil prices series and see the plot.



We can see from the plot above that the series looks much more stationary than the original one. Note that there are outliers in this case which needs to be cleaned up and we will come back to this later in the discussion.

We can also consider second difference models that leads to stationarity. Now, assume that M_t is linear over three consecutive time points. We can predict M_t at the middle point t by choosing $\beta_{0,t}$ and $\beta_{1,t}$ to minimize

$$\sum_{j=-1}^{1}(Y_{t-j} - (\beta_{0,t} + j\beta_{1,t}))^2$$

which yields,

$$\hat{M}_t = \frac{1}{3}(Y_{t+1} + Y_t + Y_{t-1})$$

and the detrended series is

$$\begin{aligned}
Y_t - \hat{M}_t &= Y_t - \frac{Y_{t+1} + Y_t + Y_{t-1}}{3} \\
&= -\frac{1}{3}(Y_{t+1} - 2Y_t + Y_{t-1}) \\
&= -\frac{1}{3}\nabla(\nabla Y_{t+1}) \\
&= -\frac{1}{3}\nabla^2(Y_{t+1})
\end{aligned}$$

a constant multiple of the centered second difference of Y_t. We have differenced twice, but both differences are at lag 1.

Alternatively, we might assume that

$$Y_t = M_t + e_t$$

where $M_t = M_{t-1} + W_t$ and $W_t = W_{t-1} + \epsilon_t$ with e_t and ϵ_t independent white noise time series. The stochastic trend M_t os such that its rate of change ∇M_t is changing slowly over time. Then,

$$\begin{aligned}
\nabla Y_t &= \nabla M_t + \nabla e_t = W_t + \nabla e_t \\
&and \\
\nabla^2 Y_t &= \nabla W_t + \nabla^2 e_t \\
&= \epsilon_t + (e_t - e_{t-1}) - (e_{t-1} - e_{t-2}) \\
&= \epsilon_t + e_t - 2e_{t-1} + e_{t-2}
\end{aligned}$$

which has acf of MA(2) process. The second difference of the nonstationary process Y_t is stationary which leads us to the definition of integrated autoregressive moving average time series models.

ARIMA Models

A time series Y_t is said to follow ARIMA model if the d^{th} difference $W_t = \nabla^d Y_t$ is a stationary ARMA process. For practical purposes, we can usually take $d = 1$ or at most $d = 2$.

Suppose we have $ARIMA(p, 1, q)$ process with $W_t = Y_t - Y_{t-1}$ then we have,

$$W_t = \phi_1 W_{t-1} + \phi_2 W_{t-2} + ... + \phi_p W_{t-p} + e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2} - ... - \theta_q e_{t-q}$$

In terms of observed series,

$$Y_t - Y_{t-1} = \phi_1(Y_{t-1} - Y_{t-2}) + \phi_2(Y_{t-2} - Y_{t-3}) + ... + \phi_P(Y_{t-p} - Y_{t-p-1}) + e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2} - ... - \theta_q e_{t-q}$$

which can be rewritten as a difference equation form,

$$Y_t = (1+\phi_1)Y_{t-1} + (\phi_2-\phi_1)Y_{t-2} + (\phi_3-\phi_2)Y_{t-3} + ... + (\phi_p-\phi_{p-1})Y_{t-p} - \phi_p Y_{t-p-1} + e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2} - ... - \theta_q e_{t-q}$$

Note that it appears to be $ARMA(p+1,q)$ process and the characteristic polynomial satisfies

$$1 - (1+\phi_1)x - (\phi_2-\phi_1)x^2 - (\phi_3-\phi_2)x^3 - ... - (\phi_p - \phi_{p-1})x^p + \phi_p x^{p+1} = (1 - \phi_1 x - \phi_2 x^2 - ... - \phi_p x^p)(1-x)$$

We can see from the factorization that the root is at x = 1 which implies nonstationarity. The remaining roots however are the roots of the characteristic polynomial o the stationary process ∇Y_t

Since nonstationary processes are not in statistical equilibrium, we cannot assume that they go infinitely into the past or that they start at $t = -\infty$. However, we can assume that they start at some time point $t = -m$ where $-m$ is earlier than time $t = 1$ at which the first series is observed. We take $Y_t = 0$ for $t < -m$. Then, the difference equation $Y_t - Y_{t-1} = W_t$ can be solved by summing both sides from $t = -m$ tp $t = t$,

$$Y_t = \sum_{j=-m}^{t} W_j$$

for $ARIMA(p,1,q)$ process. THe $ARIMA(p,2,q)$ can de dealt with in a similar way by summing twice,

$$Y_t = \sum_{j=-m}^{t} \sum_{i=-m}^{j} W_i = \sum_{j=0}^{t+m} (j+1) W_{t-j}$$

We can use these representations to investigate the covariance properties of ARIMA models.

If the process has no AR terms then it is $IMA(d,1)$ and if it has no MA terms then it is $ARI(p,q)$. Let us discuss more in details of these types of processes below.

IMA(1,1) Model

This model represents most of the processes arising in business and economics. Consider the model in difference equation form,

$$Y_t = Y_{t-1} + e^t - \theta e_{t-1}$$

By writing Y_t explicitly as a function of present and past noise values using $Y_t = \sum_{j=-m}^{t} W_j$ and the fact that $W_t = e_t - \theta e_{t-1}$,

$$Y_t = e_t + (1-\theta)e_{t-1} + (1-\theta)e_{t-2} + ... + (1-\theta)e_{-m} - \theta e_{-m-1}$$

We are assuming that $-m < 1$ and $t > 0$. So, Y_t can be seen as equally weighted accumulation of a larger number of white noise values.

Let us derive the variances and correlations. From above equation, we have,

$$Var(Y_t) = [1+\theta^2+(1-\theta)^2(t+m)]\sigma_e^2 and Corr(Y_t, Y_{t-k}) = \frac{1-\theta+\theta^2+(1-\theta)^2(t+m-k)}{[Var(Y_t)Var(Y_{t-k})]^{0.5}} \approx \sqrt{\frac{t+m-k}{t+m}} \approx 1$$

for large m and moderate k. We can see that as t increases, $Var(Y_t)$ increases and can be large and the correlation between Y_t and Y_{t-k} will be strongly positive for many lags $k = 1, 2, ...$

The ARI(1,1) Model

The $ARI(1,1)$ process satisfies,

$$Y_t - Y_{t-1} = \phi(Y_{t-1} - Y_{t-2}) + e_t$$

or,

$$Y_t = (1+\phi)Y_{t-1} - \phi Y_{t-2} + e_t$$

where, $|\phi| < 1$. In order to fins the Φ-weights, we need to use the technique that will generalize to arbitrary ARIMA models. We can obtain weights by equating like powers of x in the identity:

$$1 - \phi_1 x - \phi_2 x^2 - ... - \phi_p x^p (1-x)^d (1 + \Psi_1 x + \Psi_2 x^2 + \Psi_3 x^3 + ...) = (1 - \theta_1 x - \theta_2 x^2 - \theta_3 x^3 - ... - \theta_q x^q$$

which reduces to;

$$(1-\phi x)(1-x)(1 + \Psi_1 x + \Psi_2 x^2 + \Psi_3 x^3 + ...) = 1$$

or,

$$[1 - (1+\phi)x + \phi x^2][1 + \Psi_1 x + \Psi_2 x^2 + \Psi_3 x^3 + ...] = 1$$

equating like powers of x on both sides, we obtain

$$-(1+\phi) + \Psi_1 = 0\phi - (1+\phi)\Psi_1 + \Psi_2 = 0$$

In general,

$$\Psi_k = (1+\phi)\Psi_{k-1} - \phi\Psi_{k-2}$$

for $k \geq 2$ with $\Psi_0 = 1$ and $\Psi_1 = 1 + \phi$. This recursion with starting values allows us to compute as many weights as necessary. An explicit solution to the recursion in this case can be found by

$$\Psi_k = \frac{1 - \phi^{k+1}}{1 - \phi}$$

for $k \geq 1$

Constant Terms in ARIMA Models

A nonzero constant mean, μ in a stationary ARMA model W_t can be accommodated in either of two ways. We can assume that

$$W_t - \mu = \phi_1(W_{t-1} - \mu) + \phi_2(W_{t-2} - \mu) + ... + \phi_p(W_{t-p} - \mu) + e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2} - ... - \theta_q e_{t-q}$$

Alternatively, we can introduce the constant mean term θ_0 into the model as follows:

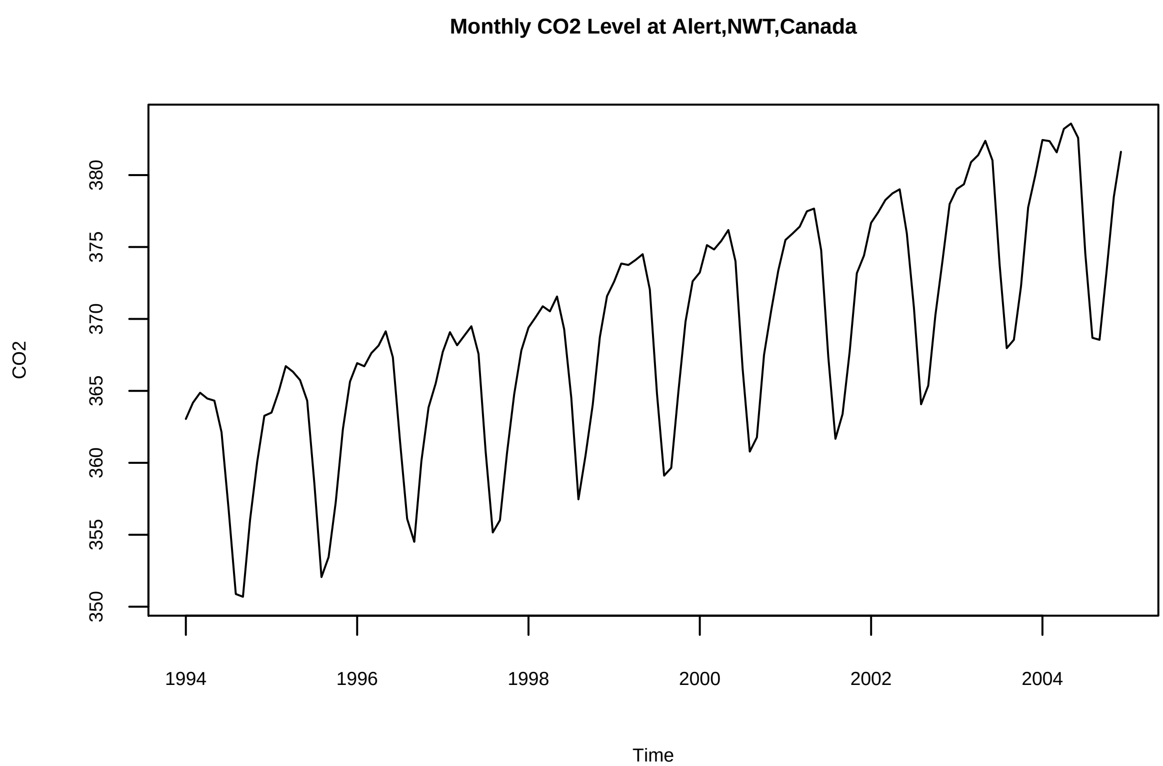
$$W_t = \theta_0 + \phi_1 W_{t-1} + \phi_2 W_{t-2} + ... + \phi_p W_{t-p} + e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2} - ... - \theta_q e_{t-q}$$

Taking expectation on both sides, we get,

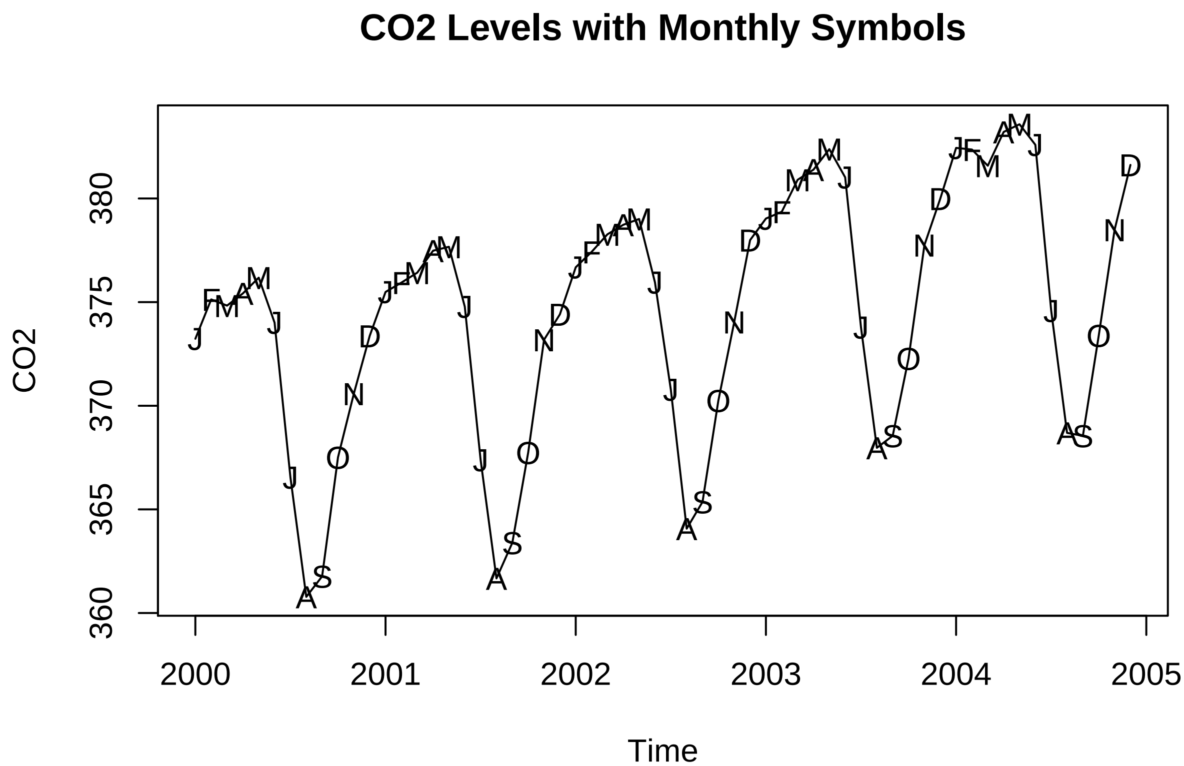
$$\mu = \theta_0 + (\phi_1 + \phi_2 + ... + \phi_p)\mu\mu = \frac{\theta_0}{1 - \phi_1 - \phi_2 - ... - \phi_p}\theta_0 = \mu(1 - \phi_1 - \phi_2 - ... - \phi_p)$$

Seasonal Models

In business and economics, the assumption of deterministic trend is not good since there seems to be cyclical tendencies in the series. Here, we are going to look at the monthly CO_2 levels from January 1994 through December 2004 at one of the site in Alert,North-west Territories,Canada near the Artic Circle.



We can see from the plot above that there is a strong upward trend and also a seasonality component. Let us get a clear picture using monthly plotting symbols.



We can see from the plot above that the carbon dioxide levels are higher during winter months and much lower in summer months. We can consider the deterministic seasonal models but we know that such models do not explain the behavior of this series. It can be shown that for this model, the residuals from a seasonal means plus linear time trend model are highly autocorrelated at many lags. So, we consider the stochastic seasonal model which works very well for this type of series.

Seasonal ARIMA Models

Let s denote the known seasonal period; for monthly series $s = 12$ and for quarterly series $s = 4$ and so on. Consider the time series model given below,

$$Y_t = e_t - \Theta e_{t-12}$$

Notice,

$$Cov(Y_t, Y_{t-1}) = Cov(e_t - \Theta e_{t-12}, e_{t-1} - \Theta e_{t-13}) = 0$$

But,

$$Cov(Y_t, Y_{t-12}) = Cov(e_t - \Theta_{t-12}, e_{t-12} - \Theta e_{t-24}) = -\Theta \sigma_e^2$$

Such a series is stationary and has non-zero autocorrelations only at lag 12.

Let us define a seasonal $MA(Q$ model of order Q with seasonal period s by,

$$Y_t = e_t - \Theta_1 e_{t-s} - \Theta_2 e_{t-2s} - ... - \Theta_Q e_{t-Qs}$$

with seasonal MA characteristic polynomial,

$$\Theta(x) = 1 - \Theta_1 x^s - \Theta_2 x^{2s} - ... - \Theta_Q x^{Qs}$$

Such a series is always stationary and that the acf will be non-zero only at the seasonal lags of $s, 2s, 3s, ..., Qs$. In particular,

$$\rho_{ks} = \frac{-\Theta_k + \Theta_1\Theta_{k+1} + \Theta_2\Theta_{k+2} + ... + \Theta_{Q-k}\Theta_Q}{1 + \Theta_1^2 + \Theta_2^2 + ... + \Theta_Q^2}$$

for $k = 1, 2, ..., Q$. Note that for the model to be invertible, the roots of $\Theta(x) = 0$ must all exceed 1 in absolute value. The seasonal $MA(Q)$ model can also be viewed as special case of nonseasonal MA model of order $q = Qs$ but with all θ-values zero except at the seasonal lags $s, 2s, 3s, ..., Qs$. Seasonal autoregressive models can also be define. Consider the model below,

$$Y_t = \Phi Y_{t-12} + e_t$$

where $|\Phi| < 1$ and e_t is independent of $Y_{t-1}, Y_{t-2}, ...$ We can show that $|\Phi| < 1$ ensures stationarity. Hence, we can argue that $E(Y_t) = 0$. Multiplying above equation by Y_{t-k}, taking expectation and dividing by γ_0 yields,

$$\rho_k = \Phi\rho_{k-12}$$

for $k \geq 1$

Clearly,

$$\rho_{12} = \Phi\rho_0 = \Phi andrho_{24} = \Phi\rho_{12} = \Phi^2$$

In general terms,

$$\rho_{12k} = \Phi^k$$

for $k = 1, 2, ...$

Setting $k = 1$ and $k = 11$ and using $\rho_k = \rho_{-k}$, we get,

$$\rho_1 = \Phi\rho_{11} and \rho_{11} = \Phi\rho_1$$

which implies that $\rho_1 = \rho_{11} = 0$. Similarly, we can show that $\rho_k = 0$ except at seasonal lags $12, 24, 36, ...$ The acf decays exponentially like an $AR(1)$ model.

Now, let us define a seasonal $AR(p)$ model of order P and seasonal period s by,

$$Y_t = \Phi_1 Y_{t-s} + \Phi_2 Y_{t-2s} + ... + \Phi_P Y_{t-Ps} + e_t$$

with seasonal characteristic polynomial,

$$\Phi(x) = 1 - \Phi_1 x^s - \Phi_2 x^{2s} - ... - \Phi_P x^{Ps}$$

The acf is nonzero only at lags $s, 2s, 3s, ...$ where it behaves like a combination of decaying exponentials and damped sine functions. In particular, above equations can be generalized to general seasonal $AR(1)$ model to give,

$$\rho_{ks} = \Phi^k$$

for $k = 1, 2, ...$ with zero correlation at other lags.

Nonstationary Seasonal ARIMA Models

The seasonal difference of period s for the series Y_t is denoted as $\nabla_s Y_t$ and is defined as,

$$\nabla_s Y_t = Y_t - Y_{t-s}$$

For example, for monthly series we consider the changes from January to January, February to February and so on for successive years. For a series of length n, the seasonal difference series will be of length $n-2$ that is s data values are lost due to seasonal differencing.

To get the idea of where the seasonal differencing might be appropriate, consider a process given below:

$$Y_t = S_t + e_t$$

with $S_t = S_{t-s} = \epsilon_t$ where e_t and ϵ_t are independent white noise. Here, S_t is a seasonal random walk and if $\sigma_\epsilon << \sigma_e$, S_t would model a slow changing seasonal component.

Due to nonstationarity of S_t, Y_t is also nonstationary. However, if we seasonally difference Y_t then we find,

$$\nabla_s Y_t = S_t - S_{t-s} + e_t - e_{t-s} = \epsilon_t + e_t - e_{t-s}$$

We can see that $\nabla_s Y_t$ is now stationary and has acf of MA(1) model.

We can also generalize the above model to account for nonseasonal slwoly changing stochastic trend. Consider,

$$Y_t = M_t + S_t + e_t$$

with

$$S_t = S_{t-s} + \epsilon_t$$

and

$$M_t = M_{t-1} + \zeta_t$$

where e_t, ϵ_t, ζ_t are mutually independent white noise series. Here, we take both a seasonal difference and an ordinary nonseasonal difference to obtain

$$\nabla\nabla_s Y_t = \nabla(M_t - M_{t-s} + \epsilon_t + e_t - e_{t-s}) = (\zeta_t + \epsilon_t + e_t) - (\epsilon_{t-s} + e_{t-s}) + e_{t-s-1}$$

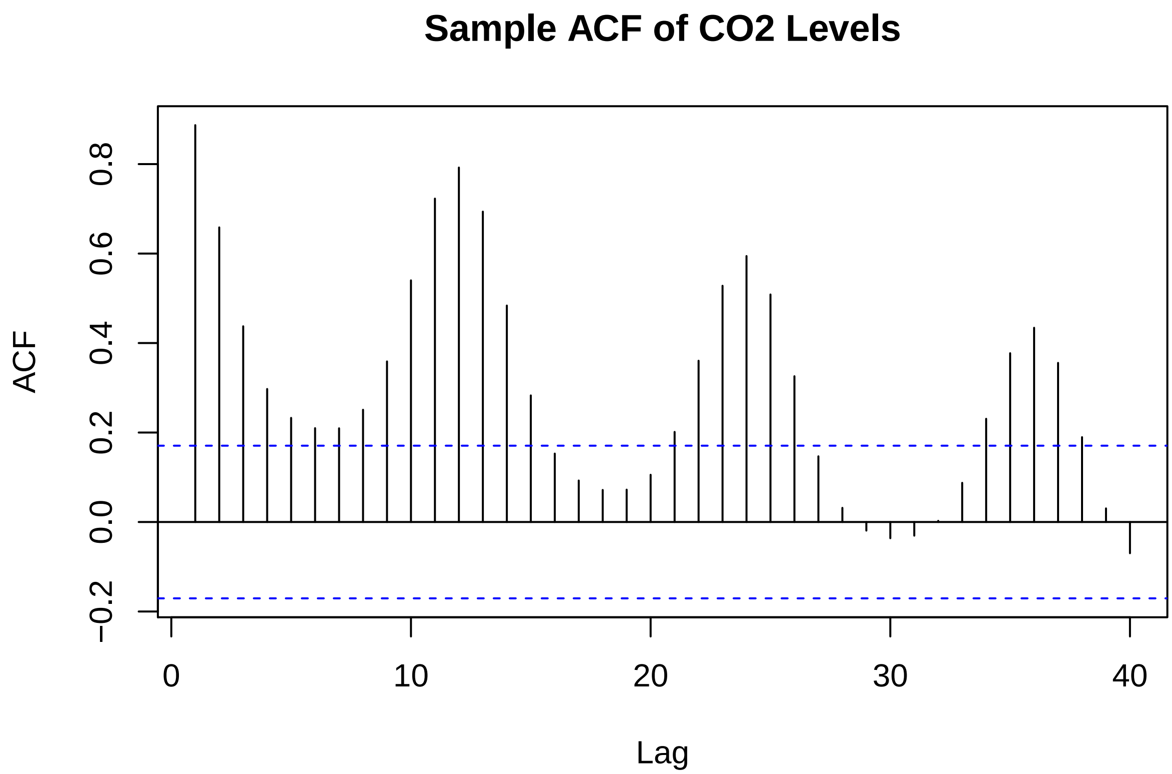
The process here is stationary an has nonzero autocorrelation only at lags $1, s-1, s, s+1$ which agress with the autocoreelation structure of the multiplicative seasonal model $ARMA(0,1) * (0,1)$ with seasonal period s

A process Y_t is said to be multiplicative seasonal $ARIMA$ model with nonseasonal(regular) orders p, d, q, seasonal orders P, D, Q and seasonal period s if the differenced series

$$W_t = \nabla^d \nabla_s^D Y_t$$

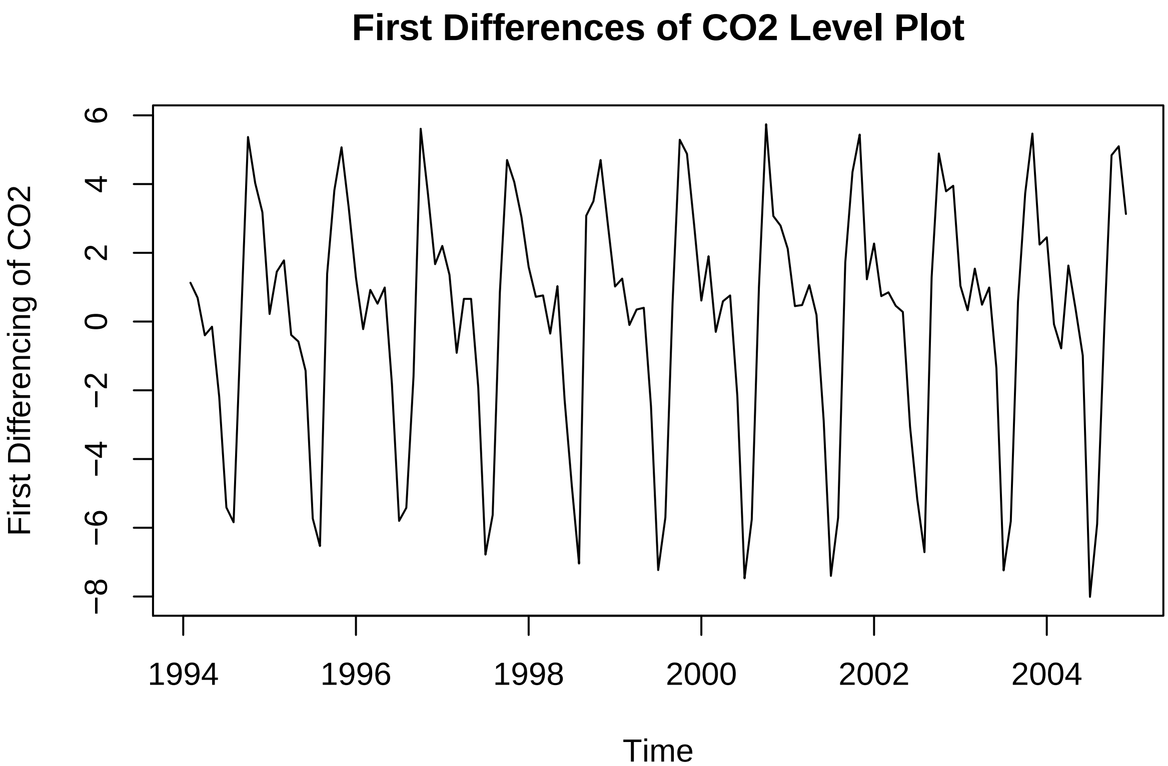
satisfies an $ARMA(p,q) * (P,Q)_s$ model with seasonal period s. We say that Y_t is an $ARIMA(p,d,q) * (P,D,Q)_s$ model with seasonal period s

Let us go back to our CO2 example from above. The upward trend alone would lead us to specify the nonstationary model. Let us see the sample acf for this series.

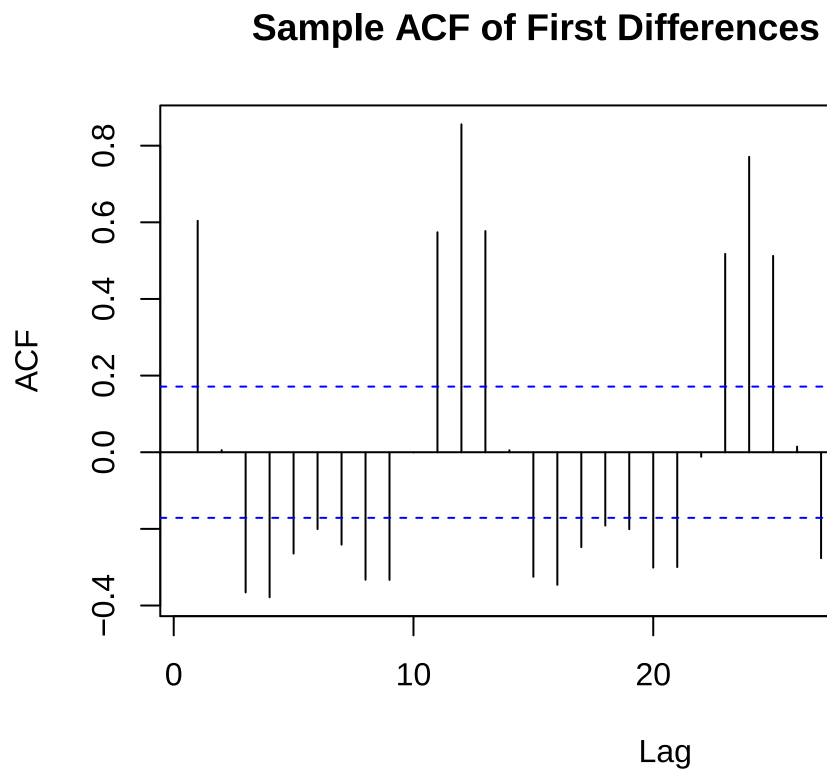


We can see from the plot that there exists a strong correlation at lags 12,24,36 and so on. There is also substantial other correlation that needs to be modeled.

Now, let us take the first difference of the time series.

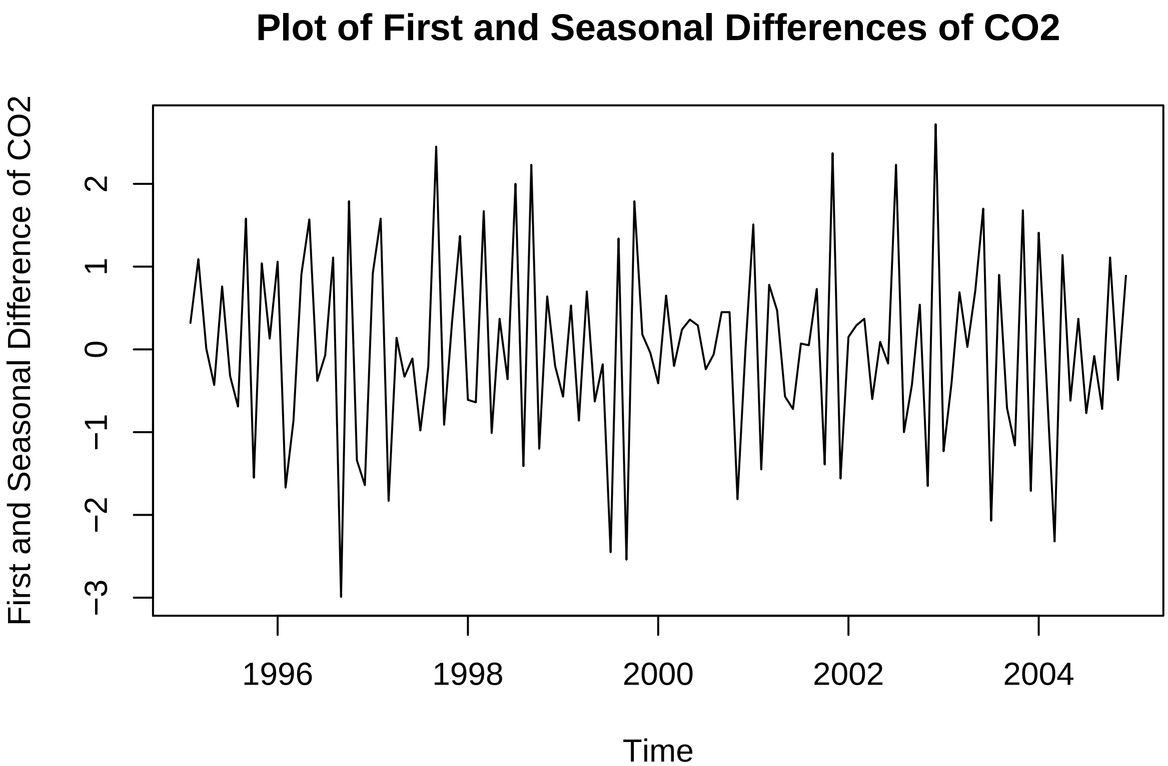


We can see from the plot above, we have omitted the general upward trend. but the strong seasonality is still present.



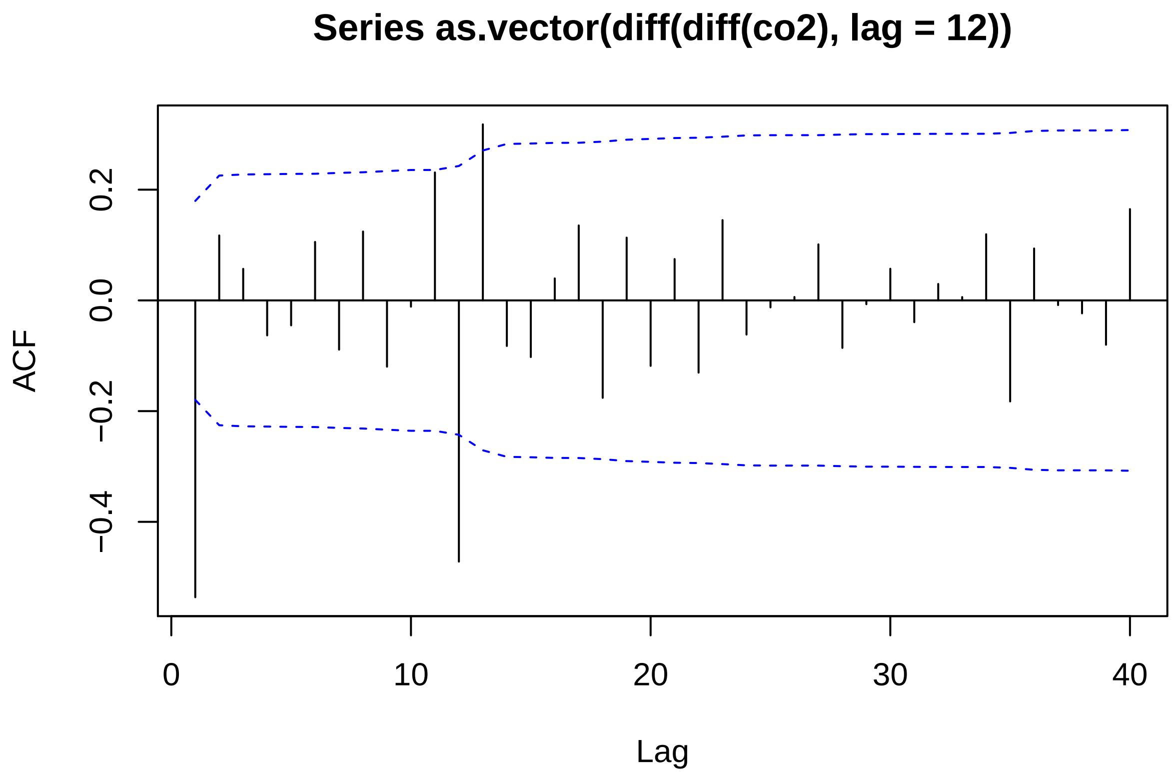
Now, let us inspect the sample ACF of first differences
Perhaps seasonal differencing will bring us to a series that may be modeled parsimoniously.

Now let us plot the time series after taking both a first difference and a seasonal difference.



From the plot above, it appears the most of the seasonality is gone now.

Now, let us inspect the sample ACF of first and Seasonal Differences.



From the plot above, we can see that very little autocorrelation remains in the series after two differences have been taken into account. This plot also suggests that a simple model which incorporates lag 1 and lag 12 autocorrelations might be adequate.

we will consider specifying the multiplicative, seasonal $ARIMA(0,1,1) * (0,1,1)_{12}$ model,

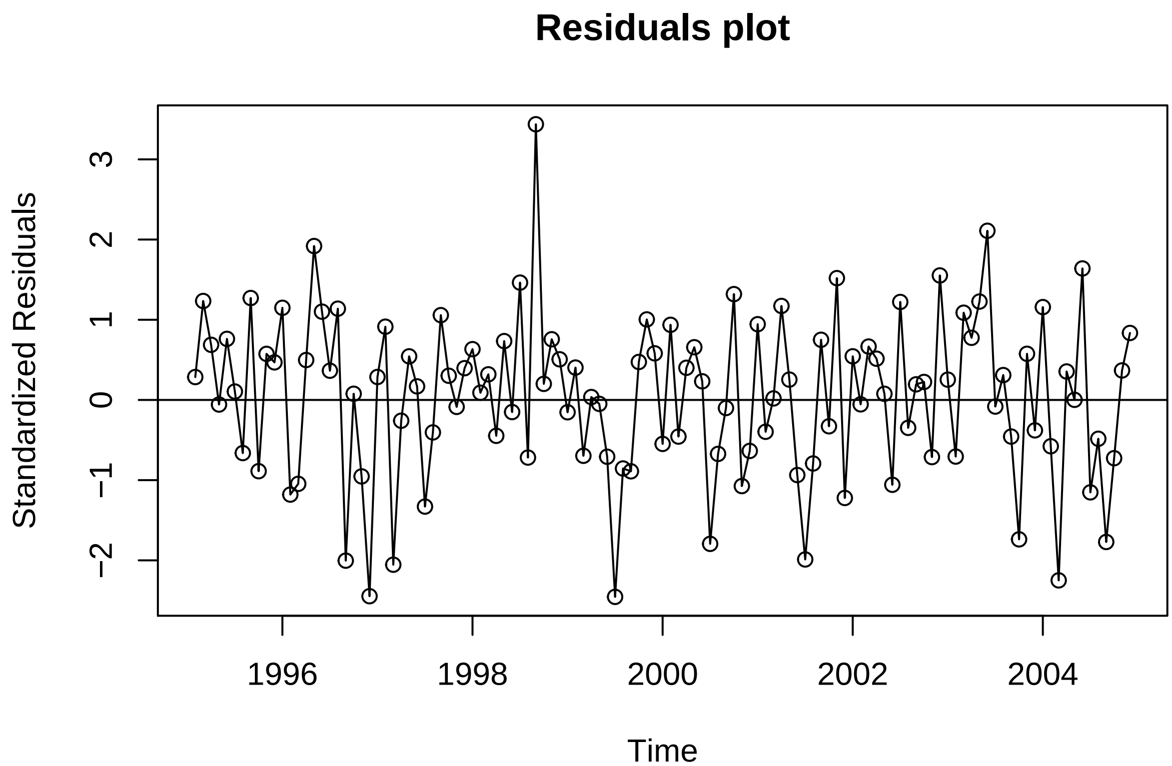
$$\nabla_{12}\nabla Y_t = e_t - \theta e_{t-1} - \Theta e_{t-12} + \theta\Theta e_{t-13}$$

which incorporates many of these requirements.

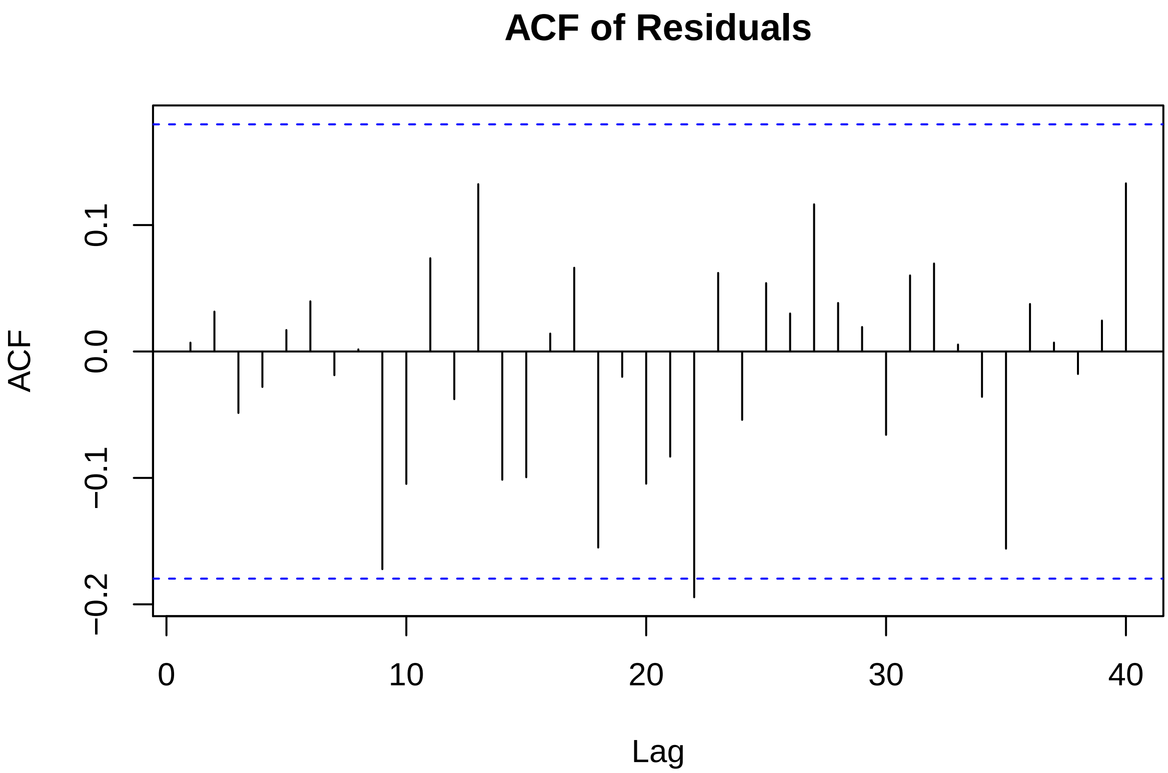
Now, let us look at the maximum likelihood estimates and their standard errors for the $ARIMA(0,1,1) * (0,1,1)_{12}$ model for CO2 levels.

```
##
## Call:
## arima(x = co2, order = c(0, 1, 1), seasonal = list(order = c(0, 1, 1), period = 12))
##
## 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
```

We can see that coefficients are all highly significant and we can proceed to check further on this model. To check the estimations of $ARIMA(0,1,1) * (0,1,1)_{12}$ model, we first look at the time series plot of the residuals. Now, let us plot the standardized residuals below:

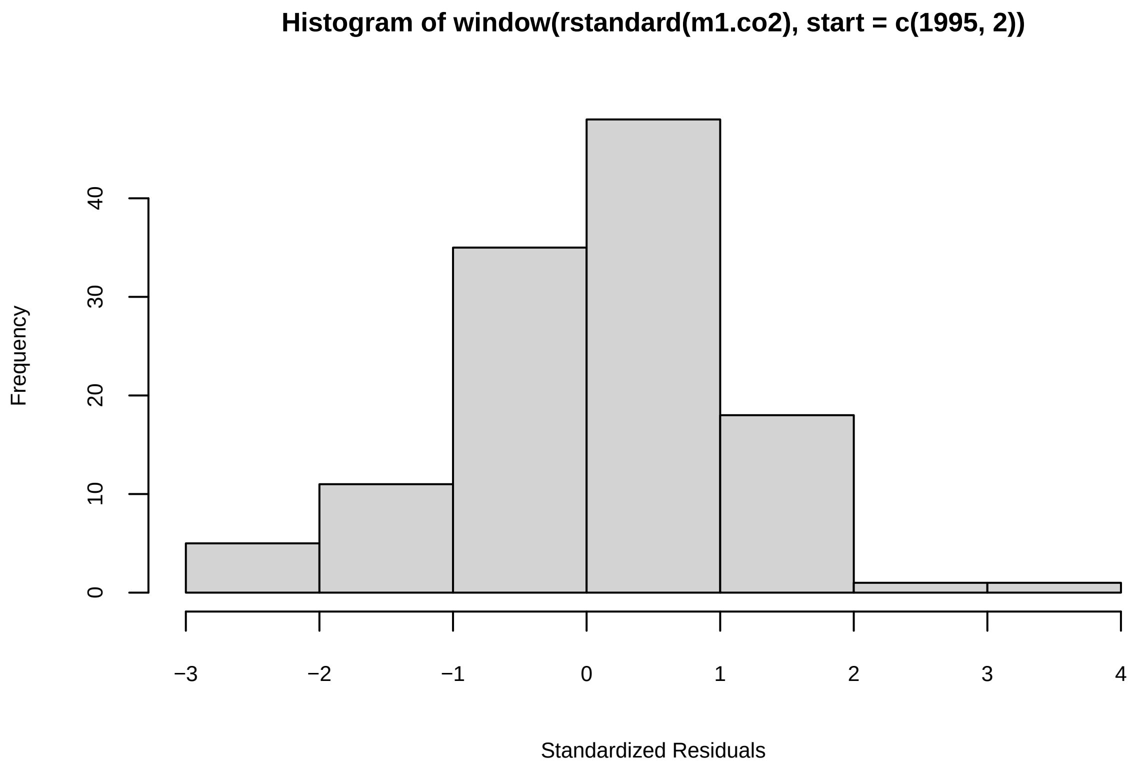


From the plot above, we can see that there are some strange behavior in the middle but it does not suggest any irregularities in the model. Let us look at the sample ACF of the residuals.

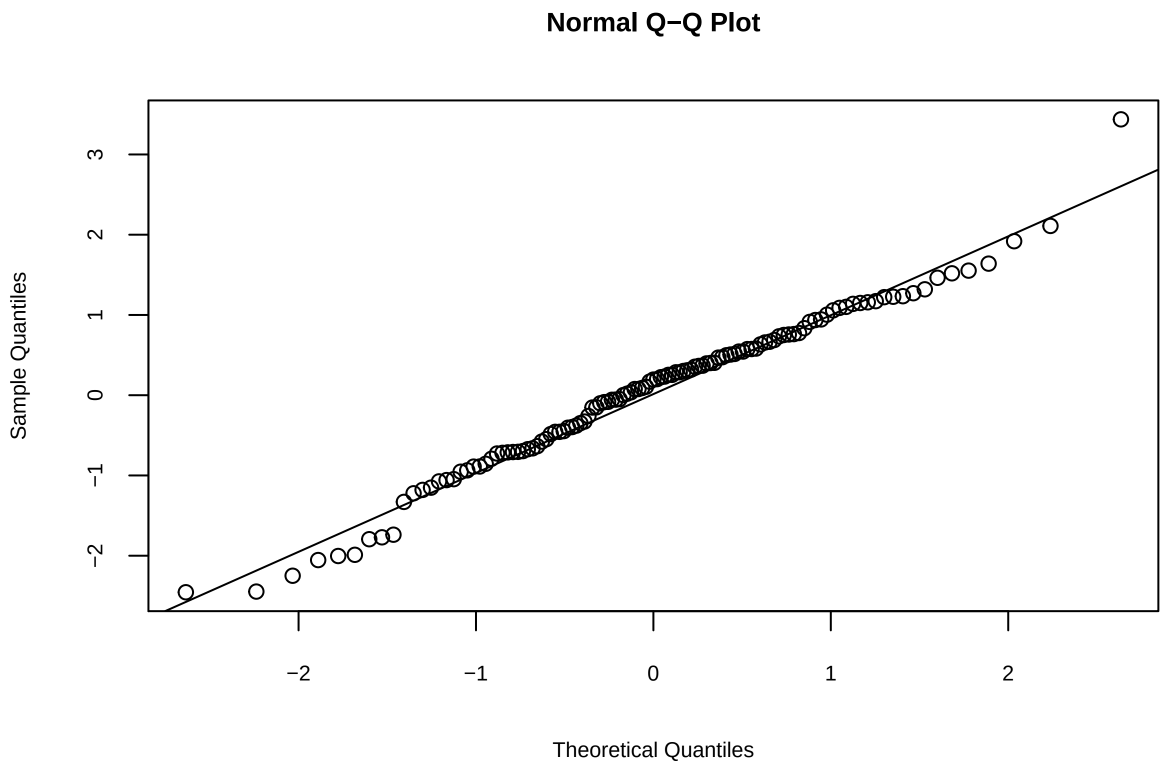


From the plot above, we can see that the only statistically significant correlation is at lag 22 and this correlation has a value of -0.17, a very small correlation. Except for marginal significance at lag 22, the model seems to have captured the essence of the dependence in the series.

Now, let us investigate the question of normality of the error terms through the residuals. Let us see the histogram of the residuals.



We can see from the plot above that the shape is somewhat bell-shaped but not ideal. Let us visualize the quantile-quantile plot which gives us much better insight than above histogram.



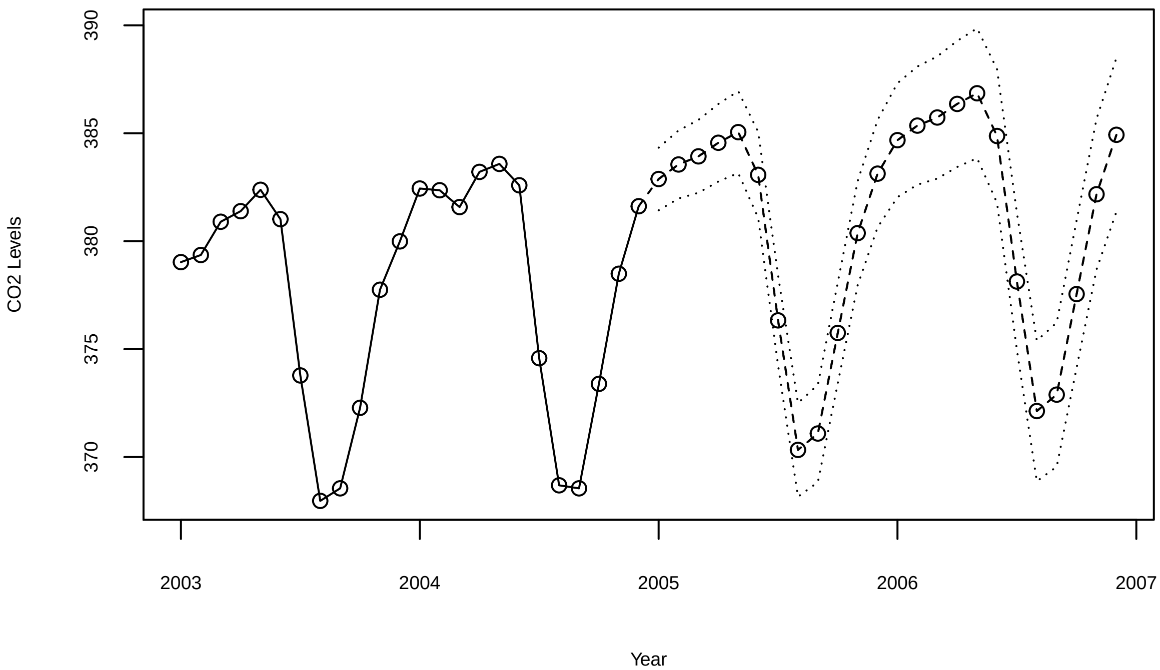
From the plot above, we can see that there is one outlier in the upper tail.

```
##
##  Shapiro-Wilk normality test
##
## data:  co2
## W = 0.98523, p-value = 0.1652
```

We can further see from Shapiro-Wilk test that normality is not rejected at any of the usual significance levels.

Forecasting

Let us forecast and have forecast limits for the CO2 model.



The above plot shows the forecasts and 95% forecast limits for two years ahead time for the $ARIMA(0,1,1) * (0,1,1)_{12}$ model that we fit. The forecast mimics the stochastic periodicity in the data quire well and forecast limits give a good band for the precision for the forecasts.

All this example and definition are taken from springer textbook. Feel free to explore springer textbook for additional mathematical details and explanation.