

# Series4

*Group E*

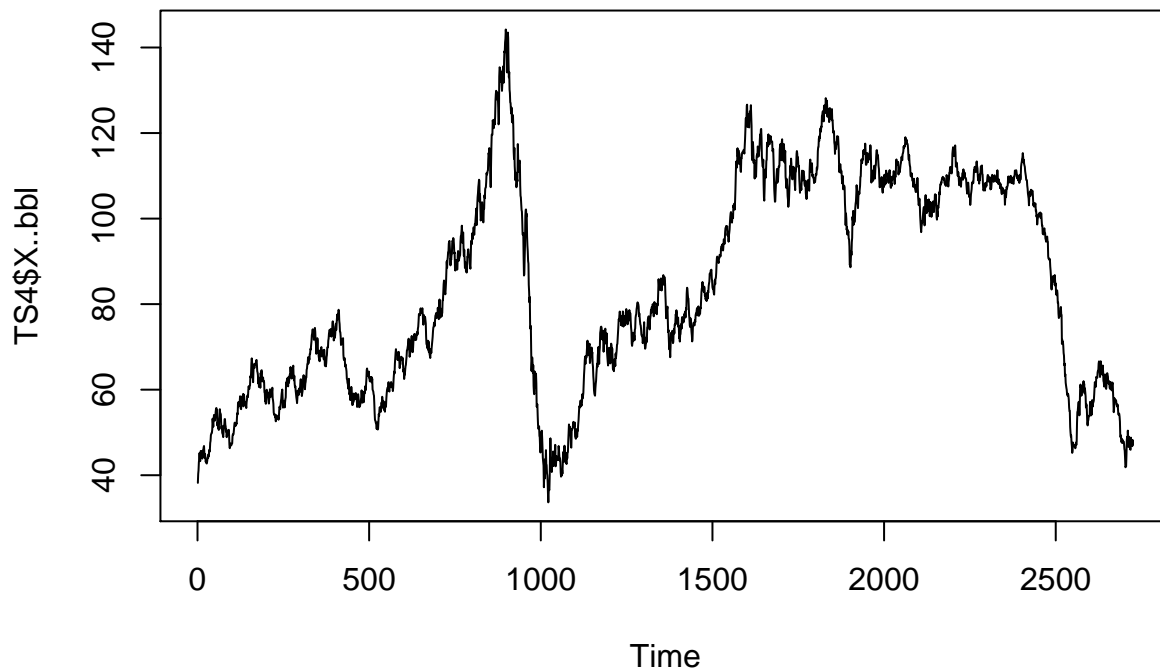
*February 17, 2016*

Box - Jenkins Analysis refers to a systematic method of identifying, fitting, checking, and using integrated autoregressive, moving average (ARIMA) time series models. The method is appropriate for time series of medium to long length (at least 50 observations).

If future values can be described only by their probability distribution, the series is said to be a statistical or stochastic process.

A special class of stochastic processes is a stationary stochastic process. A statistical process is stationary if the probability distribution is the same for all starting values of  $t$ . This implies that the mean and variance are constant for all values of  $t$ . A series that exhibits a simple trend is not stationary because the values of the series depend on  $t$ . A stationary stochastic process is completely defined by its mean, variance, and autocorrelation function. One of the steps in the Box - Jenkins method is to transform a non-stationary series into a stationary one.

```
ts.plot(TS4$X..bbl)
```



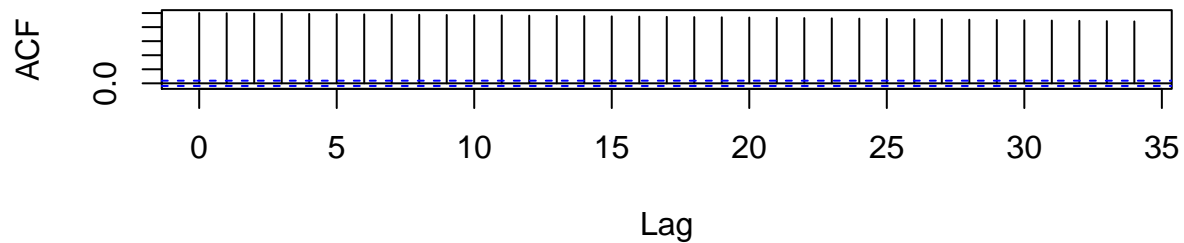
```
#skewness(TS4$X..bbl,na.rm = TRUE)
#kurtosis(TS4$X..bbl,na.rm = TRUE)
```

The stationary assumption allows us to make simple statements about the correlation between two successive values,  $X_t$  and  $X_{t+k}$ . This correlation is called the autocorrelation of lag  $k$  of the series. The autocorrelation function displays the autocorrelation on the vertical axis for successive values of  $k$  on the horizontal axis. The following figure shows the autocorrelation and partial auto-correlation function of the brent prices data.

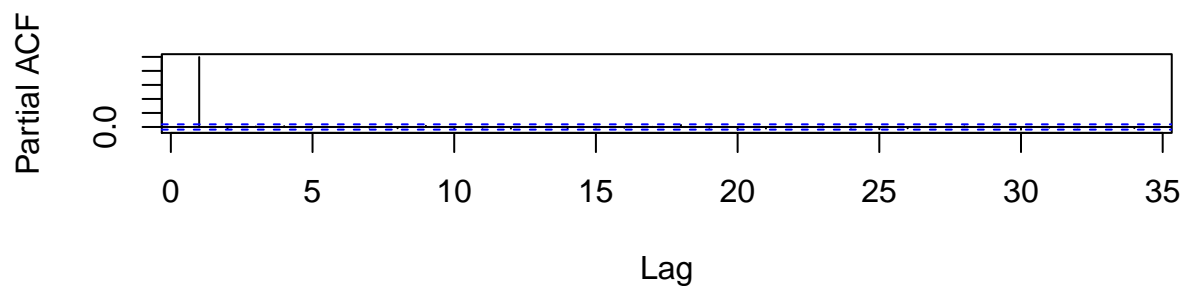
The autocorrelation diagram helps decide on the order of the MA process for the ARIMA model. The partial autocorrelation plot helps decide on the order of the AR process for the ARIMA model.

```
par(mfrow=c(2,1))
acf(TS4$X..bbl)
pacf(TS4$X..bbl)
```

### Series TS4\$X..bbl



### Series TS4\$X..bbl

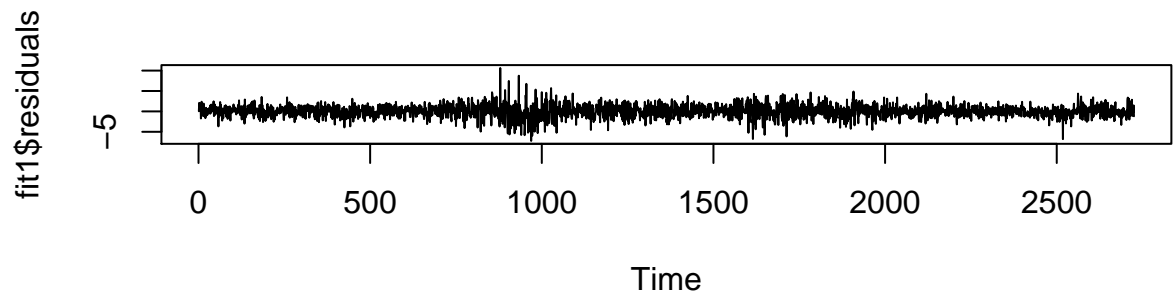


```
fit1 = arima(TS4$X..bbl,order = c(0,1,1))
fit1
```

```
##
## Call:
## arima(x = TS4$X..bbl, order = c(0, 1, 1))
##
## Coefficients:
##          ma1
##         0.0550
## s.e.  0.0194
##
## sigma^2 estimated as 2.34:  log likelihood = -5023.33,  aic = 10050.66
```

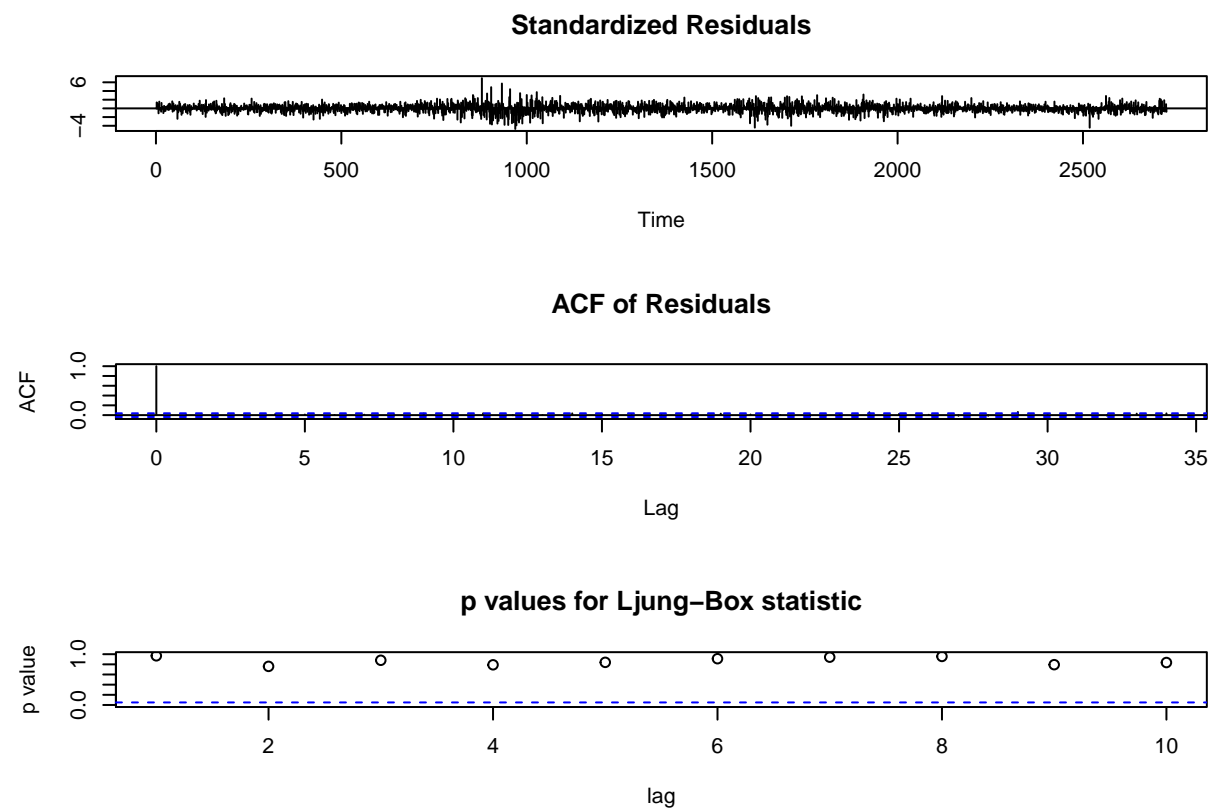
```
ts.plot(fit1$residuals)
```

*#We checked aic of several other models and all of them seemed to be higher than this value.*



The box Ljung test is a test for evaluating the fit of residuals vs lag. If for even higher lags the Box- Ljung test returns high p-values then the residuals can be said to be close to white noise.

```
tsdiag(fit1)
```



```
Box.test(fit1$residuals,lag = 13,type = 'Ljung')
```

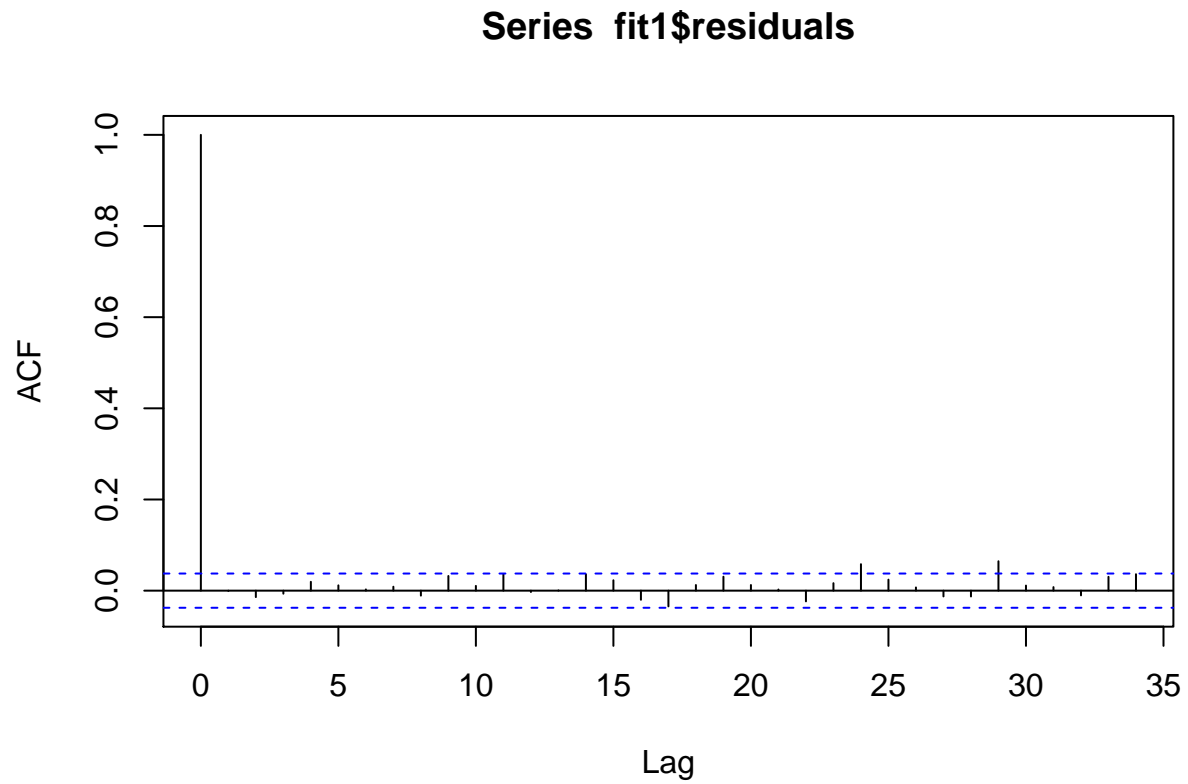
```
##
## Box-Ljung test
##
## data: fit1$residuals
## X-squared = 9.3667, df = 13, p-value = 0.7447
```

```
##(Ljung-Box test: we accept the null of rho's=0 if p-value is bigger than 0.05)
shapiro.test(fit1$residuals)
```

```
##
```

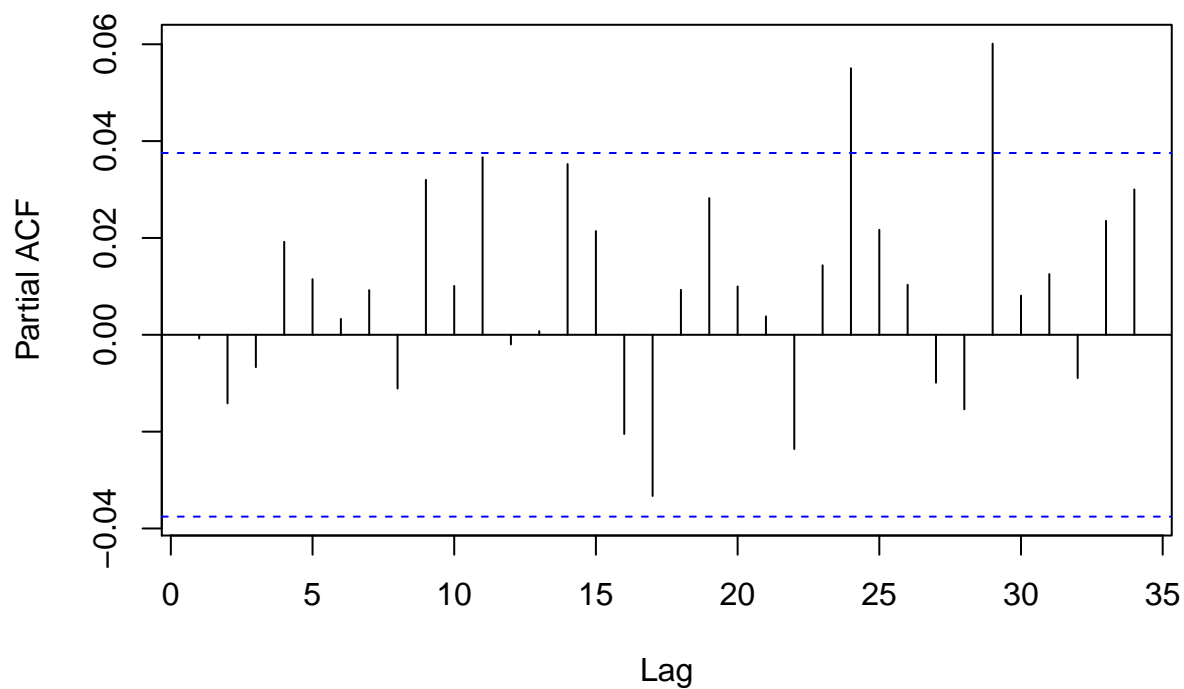
```
## Shapiro-Wilk normality test
##
## data: fit1$residuals
## W = 0.97714, p-value < 2.2e-16
```

```
acf(fit1$residuals)
```



```
pacf(fit1$residuals)
```

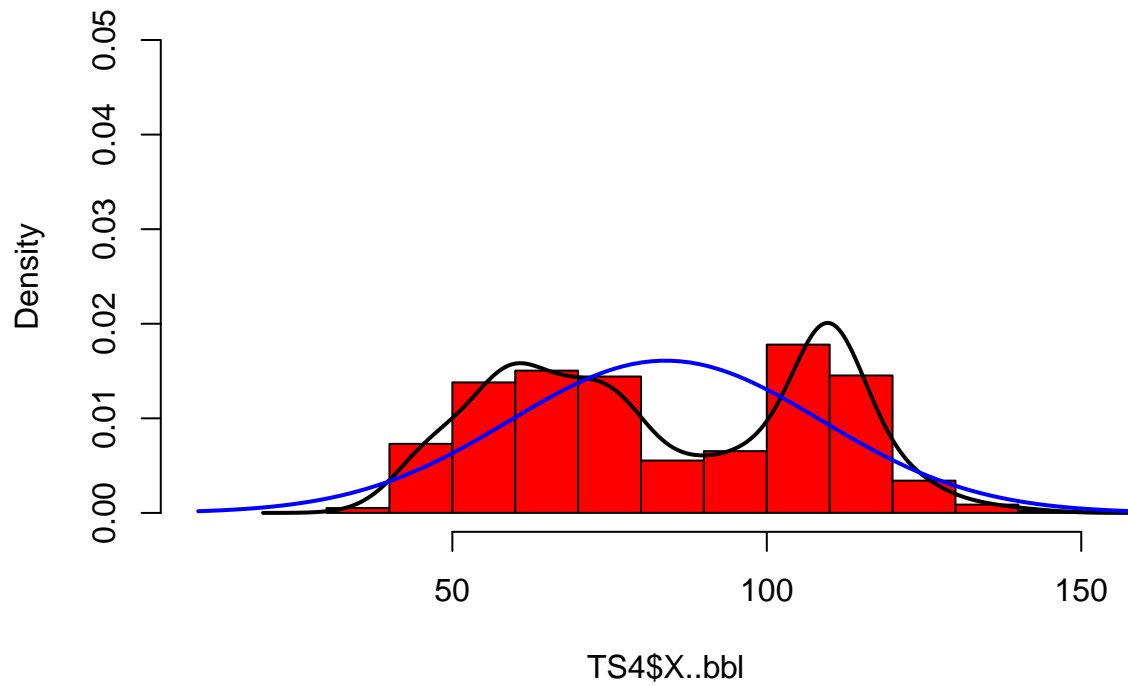
## Series fit1\$residuals



Now it seems pretty clear that the residuals do not have any obvious pattern.

```
hist(TS4$X..bbl,prob=T,ylim=c(0,0.05),
     xlim=c(mean(TS4$X..bbl)-3*sd(TS4$X..bbl),
             mean(TS4$X..bbl)+3*sd(TS4$X..bbl)),col="red")
lines(density(TS4$X..bbl),lwd=2)
mu=mean(TS4$X..bbl)
sigma=sd(TS4$X..bbl)
x = seq(mu-3*sigma,mu+3*sigma,length=100)
yy<-dnorm(x,mu,sigma)
lines(x,yy,lwd=2,col="blue")
```

## Histogram of TS4\$X..bbl



We do band forecasting instead of point forecasting.

```
y.pred<-predict(fit1,n.ahead=6)
ts.plot(TS4$X..bbl, xlim = c(2500,2800),ylim = c(0,100))
lines(y.pred$pred,col="red")
lines(y.pred$pred+1.96*y.pred$se,col="red",lty=3)
lines(y.pred$pred-1.96*y.pred$se,col="red",lty=3)
```

