Group Assignment Time Series

MBD S2.2 Group 8

2/2/2021

[Introduction](#Introduction) [Time series 1](#time_series_1) [Time series 2](#time_series_2) [Time series 3](#time_series_3) [Time series 4](#time_series_4) [Time series 5](#time_series_5) [Time series 6](#time_series_6) [Time series 7](#time_series_7)

Introduction

We are glad to present to you our first Group Assignment! We also wanted to practice R further, so we are delivering our findings in a markdown format.

# Summary of findings

For each series, we first recurred to both a graphical analysis and formal test (the ADF Test) to determine if the time series were mean and variance stationary. Our analysis determined that while all series exhibited variance stationarity , series 2,3,4 and 7 were deemed non-stationary around the mean (due to sustained upward and downward trends), a result later confirmed by the ADF Test, which suggested 1 difference transformations for all but series 4, for which 2 differences were required.

Secondly, we combined an eye-test analysis of both ACF and PACF functions with the Box - Ljung Test to assess every series autocorrelation and identify if they could represent a WN process (i.e a stationary stochastic process with 0 mean, constant variance and no autocorrelation). The results conveyed series 1 and 5 represented WN processes and the same conclusion could be made for series 2 and 7 after transformations were undertaken. Therefore, we can conclude that for these time series, a linear model is deemed ineffective when attempting to predict future values.

At this stage, note that the graphical analysis was sometimes misleading to determine time series autocorrelation. For, instance, in series 1 , while lag 12 is “out of bounds”, its small magnitude (approx. 0.11 ) and the lack of additional information about the dataset’s time frequency (e.g is it monthly or yearly data?) made us prefer to rely on the formal test instead.

On the other hand, for the remaining time series (2, 3, 3 transformed, 4, 4 transformed, 6 and 7) we defended that a linear model would be helpful. Ultimately, requiring (or not) a non-linear model to analyse this time series will remain unknown since we must first remove the linear part of the model before conducting further analysis.

Finally, we looked into the time series marginal distribution, in an effort to understand if they are normally distributed. For this purpose, we computed the time series histograms against a reference line following N ~ (0,1), as well as the Shapiro and Jarque-Bera Test. From this analysis, we concluded that time series 1 and 2 transformed , 3 transformed, and 4 transformed are normally distributed ( Please note that , in case of conflict between formal and eye tests, we often opted to trust the latter in detriment of the tests rigidity, a situation exemplified in series 1). From those , given that series 1 and 2 transformed are already White Noise, they can also be deemed Gaussian white noise and Strict White Noise since an uncorrelated variable whose multivariate joint distribution is Normal is forcefully independent and identically distributed(i.i.d). On the other hand, since time series 3 and 4 transformed are not White Noise, they cannot be Gaussian White Noise, and therefore neither Strict White Noise, despite being normally distributed.

Out of the non-normally distributed series (series 2 to 7 and 7 transformed) , Series 5 and 7 transformed required an additional analysis because they are WN processes that do not follow a normally distribution, meaning the question about the series being SWN is left unanswered ( as well as its non-linear modelling requirements). By taking the squares of our series and replicating the analysis conducted above to infer autocorrelation, we were able to determine that both series required a non linear model to capture the relationship between their pasts and presents values ( lags of the ACF & PACF out of bonds and results corroborated by formal tests) and therefore are not SWN processes.

**Note : All test were analysed considering a 95% confidence level (p.value = 0.05).**

## Original Data

kbl(summarytable\_ts\_or)

ORIGINAL DATA

Series 1

Series 2

Series 3

Series 4

Series 5

Series 6

Series 7

Stationary

YES

NO

NO

NO

YES

YES

NO

Marginal Normal distribution

YES

NO

NO

NO

NO

NO

NO

WN

YES

NO

NO

NO

YES

NO

NO

SWN

YES

NO

NO

NO

NO

NO

NO

GWN

YES

NO

NO

NO

NO

NO

NO

Linear model

NO

YES

YES

YES

NO

YES

YES

Nonlinear model

NO

??

??

??

YES

??

??

Transformations

NO

YES

YES

YES

NO

NO

YES

## Transformed Data

kbl(summarytable\_ts\_tr)

TRANSFORMED DATA

Series 1

Series 2

Series 3

Series 4

Series 5

Series 6

Series 7

Difference

0 diff

1 diff

1 diff

2 diff

NA

NA

1 diff

Stationary

NA

YES

YES

YES

NA

NA

YES

Marginal Normal distribution

NA

YES

YES

YES

NA

NA

NO

WN

NA

YES

NO

NO

NA

NA

YES

SWN

NA

YES

NO

NO

NA

NA

NO

GWN

NA

YES

NO

NO

NA

NA

NO

Linear model

NA

NO

YES

YES

NA

NA

NO

Nonlinear model

NA

NO

??

??

NA

NA

YES

Transformations

NA

NO

NO

NO

NA

NA

NO

## Tests

Additionally we wanted to list all the tests we carried out in one place in order for it to be more understandable

kbl(teststable)

…1

Test

Type

Series 1

Series 2

Series 3

Series 4

Series 5

Series 6

Series 7

y

Stationarity

ndiffs

0

1

1

2

0

0

1

NA

Normality

Shaphiro test

p-value = 0.01941

p-value = 1.133e-08

p-value = 1.019e-06

p-value = 2.779e-12

p-value = 3.456e-10

p-value < 2.2e-16

p-value < 2.2e-16

NA

Correlation

Box.test

p-value = 0.5346

p-value < 2.2e-16

p-value < 2.2e-16

p-value < 2.2e-16

p-value = 0.03248

p-value < 2.2e-16

p-value < 2.2e-16

y^2

Corrlation

Box.test

p-value = 0.8219

p-value < 2.2e-16

p-value < 2.2e-16

p-value < 2.2e-16

p-value < 2.2e-16

p-value < 2.2e-16

p-value < 2.2e-16

z

Stationary

ndiffs

NA

0

0

0

NA

NA

0

NA

Normality

Shaphiro test

NA

p-value = 0.69

p-value = 0.2178

p-value = 0.3416

NA

NA

p-value < 2.2e-16

NA

Correlation

Box.test

NA

p-value = 0.8996

p-value < 2.2e-16

p-value < 2.2e-16

NA

NA

p-value = 0.004978

z^2

Corrlation

Box.test

NA

p-value = 0.9718

p-value = 0.0004868

p-value = 6.562e-08

NA

NA

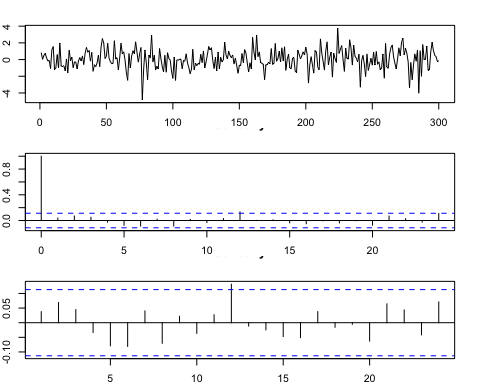
p-value < 2.2e-16

Now in case you want to see all the tests and thought process behind this table, please find them below.

## Time Series 1

### Graphical Representation of Time Series and respective ACF & PACF:

y<-series1  
par(mar=c(2,2,2,2)) # to adjust graphic size  
par(mfrow=c(3,1)) # plot the series, its acf and pacf together  
ts.plot(y)   
acf(y)  
pacf(y)



### Computation of basic descriptive statistics:

## [1] 0.03205

## [1] 1.193712

## [1] -0.2795834  
## attr(,"method")  
## [1] "moment"

## [1] 4.040454  
## attr(,"method")  
## [1] "moment"

The series plot suggest this dataset is stationary around both mean and variance. To verify our deductions, we conducted a formal unit root test: the Augmented Dickey Fuller (ADF) with the following hypothesis:

*Ho: the process is not stationary.(i.e. We need a unit root)* *H1: the process is stationary.(i.e. No differences needed)*

**Stationarity Formal Test:**

## [1] 0

The ADF test indicates 0 required differences, hence enabling us to reject Ho and validate our previous inference. Furthermore, we are also capable of claiming no data transformation is needed. When looking at both ACF & PACF function we observe that, at a 95% confidence level, past information appears unable to guide future forecasting efforts. Note that while lag 12 surpasses the confidence interval boundary, its small magnitude (approx. 0.11 ) and the lack of addition information about the dataset’s time frequency (e.g is it monthly or yearly data?) make it negligible. To formally assess the series autocorrelation, we opted to realise a Box-Ljung Test with the following hypotheses:

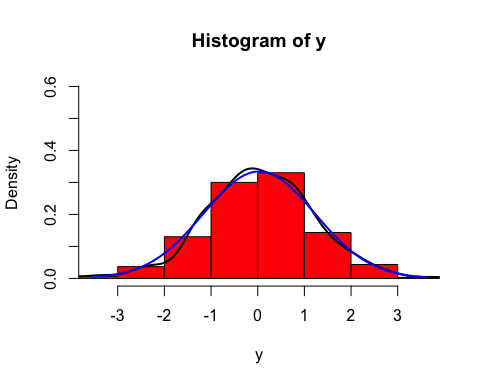
*Ho: Uncorrelated data* *H1: Correlated data*

**Testing for White-Noise(WN)/Autocorrelation:**

##   
## Box-Ljung test  
##   
## data: y  
## X-squared = 18.804, df = 20, p-value = 0.5346

The test p.value of 0.5 suggests that, at a 95% confindence level, we cannot reject the null hypothesis that the data is uncorrelated, a result that corroborates the results showcased by our graphical analysis. Consequently, given the series’ mean of 0.03205, we feel comfortable suggesting this series represents a white noise (WN) process. In addition, we can also conclude that no linear model would be helpful to model/predict future values of this time series (i.e. mean is the best point estimmate). Now, we will try and identify if the series’ marginal distribution is normally distributed (or not) by recurring to both graphical tools and nornality tests (i.e. Shapiro and Jarque-Bera Tests).

### Testing for Normality:

 According to our graphical analysis, time series 1 joint distribution seems normally distributed given the black (our series) and blue (normal distribution reference line N ~ (0,1)) overlap significantly with the exception of a small detour on the left side of the distribution around x = -2.

##   
## Shapiro-Wilk normality test  
##   
## data: y  
## W = 0.98868, p-value = 0.01941

##   
## Jarque-Bera test for normality  
##   
## data: y  
## JB = 18.193, p-value = 0.001

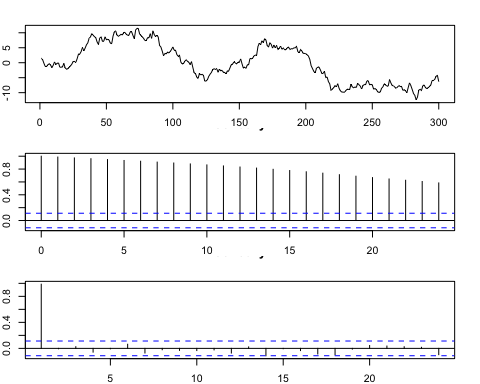
Both formal tests follow the same hypothesis structure:

*Ho: the data is normally distributed* *H1: the data is not normally distributed*

While both formal tests preconised that we should reject Ho at a 95% confidence level (p.values of roughly 0.02 and 0.003 respectively), we deemed the results excessively rigid for this particular case and decided to stick with the eye test. Therefore, by infering this series joint distribution is marginally distributed, we concluded it is not only a WN but also Gaussian White Noise (GWN) and subsquently Strict White Noise (SWN) process since an uncorrelated variable whose multivariate joint distribution is Normal is forcefully indepedent and identically distributed(i.i.d). Given our latest findings, we can also infer that no non-linear model would be of help in future forecasting endeavours.

## Time Series 2

### Graphical Representation of Time Series and respective ACF & PACF:



### Computation of basic descriptive statistics:

## [1] -0.22853

## [1] 6.31268

## [1] 0.04883792  
## attr(,"method")  
## [1] "moment"

## [1] 1.765163  
## attr(,"method")  
## [1] "moment"

The series plot suggests this dataset is not stationary around the mean (it exhbits clear sustained positive and negative fluctuations), a result hinted by the series -0.22 mean. Regarding the variance, the graph shows no clear and sustainable variance trend acrross time, hence we deemed series 2 variance stationary. To verify our deductions, we conducted a formal unit root test: the Augmented Dickey Fuller (ADF) with the following hypothesis:

*Ho: the process is not stationary.(i.e. We need a unit root)* *H1: the process is stationary.(i.e. No differences needed)*

\*\*Stationarity Formal Test:

## [1] 1

The ADF test indicates 1 difference is required for the data to become stationary in the mean, hence enabling us to accept Ho and validate our previous inference. When observing both ACF & PACF function we infer that, at a 95% confidence level, past information (especially lag 1) could prove to be extremely helpful to guide future forecasting efforts. To formally assess the series autocorrelation, we opted to realise a Box-Ljung Test with the following hypotheses:

*Ho: Uncorrelated data* *H1: Correlated data*

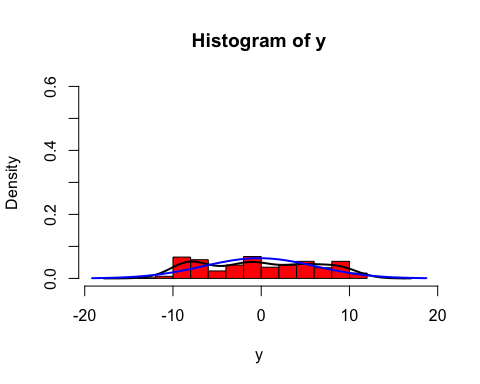
**Testing for White-Noise(WN)/Autocorrelation:**

##   
## Box-Ljung test  
##   
## data: y  
## X-squared = 4509.5, df = 20, p-value < 2.2e-16

The extremely small p-value reported confirmed our graphical analysis and we are now able to infer our time series is not a WN process of any kind. Consequently, we can also conclude that a linear model is needed model/predict future values of this time series. Ultimately, requiring (or not) a non-linear model to analyse this time series will remain unknown since we must first isolate the linear part of the model to make such inference.

Now, we will try and identify if the series’ marginal distribution is normally distributed by recurring to both graphical tools and normality tests (i.e. Shapiro and Jarque-Bera Tests).

### Testing for Normality:



According to our graphical analysis, the time series 3 joint distribution cannot be normally distributed. We decided to verify our intuition with formal tests.

##   
## Shapiro-Wilk normality test  
##   
## data: y  
## W = 0.94921, p-value = 1.133e-08

##   
## Jarque-Bera test for normality  
##   
## data: y  
## JB = 18.817, p-value = 0.0015

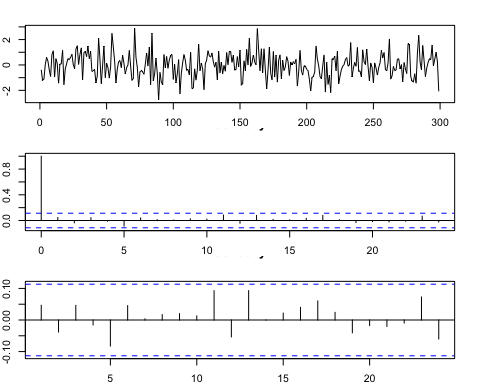
Both formal tests follow the same hypothesis structure:

*Ho: the data is normally distributed* *H1: the data is not normally distributed*

Both Jarque-Bera and Shapiro tests reported very small p-values, indicating we should reject the null hypothesis at a 95% confidence level. Therefore, we were able to validate the conclusions based on the team’s eye test.

## Time Series 2 (1 Difference Transformation):

### Graphical Representation of Time Series and respective ACF & PACF:



### Computation of basic descriptive statistics:

## [1] -0.02598662

## [1] 0.9620915

## [1] 0.1359436  
## attr(,"method")  
## [1] "moment"

## [1] 3.096639  
## attr(,"method")  
## [1] "moment"

The series plot suggest this dataset is stationary around both mean and variance. To verify our deductions, we conducted a formal unit root test: the Augmented Dickey Fuller (ADF) with the following hypothesis:

*Ho: the process is not stationary.(i.e. We need a unit root)* *H1: the process is stationary.(i.e. No differences needed)*

\*\*Stationarity Formal Test:

## [1] 0

The ADF test indicates 0 required differences, hence enabling us to reject Ho and validate our previous inference. Furthermore, we are also capable of claiming no further data transformation will be needed. When analysing both ACF & PACF function we observe that, at a 95% confidence level, past information appears unable to contribute to futuristic predictions (i.e. all lags are within the confidence interval limit boundaries). To formally assess the series autocorrelation, we opted to realise a Box-Ljung Test with the following hypotheses:

*Ho: Uncorrelated data* *H1: Correlated data*

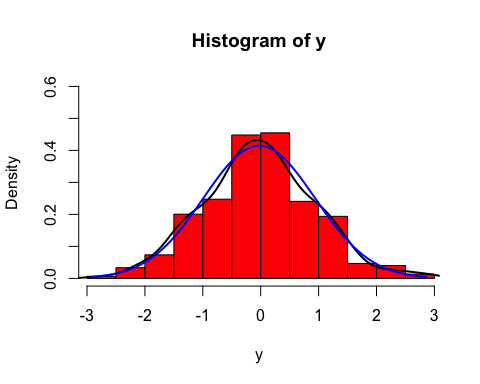
**Testing for White-Noise(WN)/Autocorrelation:**

##   
## Box-Ljung test  
##   
## data: y  
## X-squared = 12.452, df = 20, p-value = 0.8996

The test p.value of 0.9 suggests that, at a 95% confindence level, we should accept null hypothesis that the data is uncorrelated and that the past is usefuless for forecasting purposes. Therefore, given the series’ mean of -0.22, we feel comfortable suggesting this series represents a white noise (WN) process. In addition, we can also conclude that a linear model would no longer be helpful to model/predict future values of this time series (i.e. mean is the best point estimmate).

Now, we will try and identify if the series’ marginal distribution is normally distributed (and hence try and determine if such WN process is also GWN) by recurring to both graphical tools and nornality tests (i.e. Shapiro and Jarque-Bera Tests).

### Testing for Normality:



According to our graphical analysis, the time series 2 (trasnformed) joint distribution seems normally distributed given the black (our series) and blue (normal distribution reference line N ~ (0,1)) overlap significantly despite being the narrower Bell curve exhibited by the time series studied between (-1;1).

##   
## Shapiro-Wilk normality test  
##   
## data: y  
## W = 0.99619, p-value = 0.69

##   
## Jarque-Bera test for normality  
##   
## data: y  
## JB = 1.1021, p-value = 0.5235

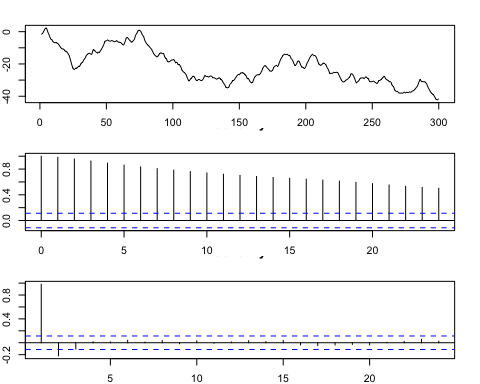
Both formal tests follow the same hypothesis structure:

*Ho: the data is normally distributed* *H1: the data is not normally distributed*

Both Jarque-Bera (p-value 0.55) and Shapiro tests (p-value 0.69) indicate we should accept the null hypothesis at a 95% confidence level. Therefore, we were able to validate the conclusions based on the team’s eye test. As per consequence, by infering this series joint distribution is marginally distributed, we concluded it is not only a WN but also Gaussian White Noise (GWN) and subsquently Strict White Noise (SWN) process since an uncorrelated variable whose multivariate joint distribution is Normal is forcefully i.i.d. Given our latest findings, we can also infer that no non-linear model would be of help in future forecasting endeavours.

## Time Series 3

### Graphical Representation of Time Series and respective ACF & PACF:



### Computation of basic descriptive statistics:

## [1] -21.65212

## [1] 10.42905

## [1] 0.3495993  
## attr(,"method")  
## [1] "moment"

## [1] 2.197106  
## attr(,"method")  
## [1] "moment"

The series plot suggests this dataset is not stationary around the mean (it exhbits a clear sustained downwards trend), a result hinted by the series -21.65 mean. Regarding the variance, the graph shows no clear and sustainable variance trend acrross time, hence we deemed series 3 variance stationary. To verify our deductions, we conducted a formal unit root test: the Augmented Dickey Fuller (ADF) with the following hypothesis:

*Ho: the process is not stationary.(i.e. We need a unit root)* *H1: the process is stationary.(i.e. No differences needed)*

**Stationarity Formal Test:**

## [1] 1

The ADF test indicates 1 difference is required for the data to become stationary in the mean, hence enabling us to accept Ho and validate our previous inference. When observing both ACF & PACF function we infer that, at a 95% confidence level, past information (particularly lags 2 and 3) could prove to be extremely helpful to guide future forecasting efforts. To formally assess the series autocorrelation, we opted to realise a Box-Ljung Test with the following hypotheses:

*Ho: Uncorrelated data* *H1: Correlated data*

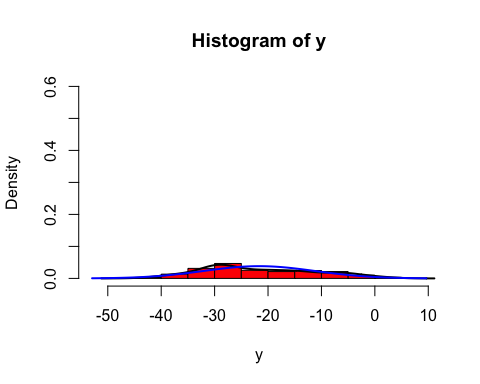
**Testing for White-Noise(WN)/Autocorrelation:**

##   
## Box-Ljung test  
##   
## data: y  
## X-squared = 3598.9, df = 20, p-value < 2.2e-16

The extremely small p-value reported confirmed our graphical analysis and we are now able to infer our time series is not a WN process of any kind. Consequently, we can also conclude that a linear model is needed model/predict future values of this time series. Ultimately, requiring (or not) a non-linear model to analyse this time series will remain unknown since we must first isolate the linear part of the model to make such inference.

Now, we will try and identify if the series’ marginal distribution is normally distributed by recurring to both graphical tools and nornality tests (i.e. Shapiro and Jarque-Bera Tests).

### Testing for Normality:



According to our graphical analysis, the time series 3 joint distribution cannot be normally distributed. We decided to verify our intuition with formal tests.

##   
## Shapiro-Wilk normality test  
##   
## data: y  
## W = 0.9645, p-value = 1.019e-06

##   
## Jarque-Bera test for normality  
##   
## data: y  
## JB = 13.938, p-value = 0.007

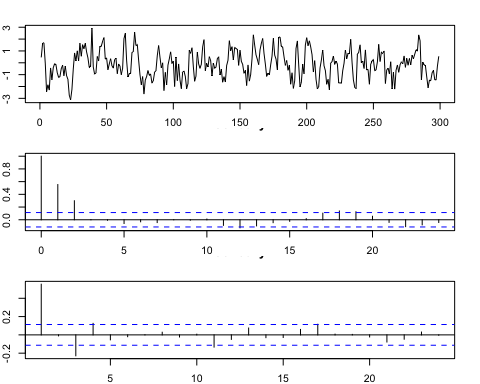
Both formal tests follow the same hypothesis structure:

*Ho: the data is normally distributed* *H1: the data is not normally distributed*

Both Jarque-Bera and Shapiro tests reported very small p-values, indicating we should reject the null hypothesis at a 95% confidence level. Therefore, we were able to validate the conclusions based on the team’s eye test.

## Time Series 3 (1 Difference Transformation):

### Graphical Representation of Time Series and respective ACF & PACF:



### Computation of basic descriptive statistics:

## [1] -0.1338495

## [1] 1.172368

## [1] 0.06834054  
## attr(,"method")  
## [1] "moment"

## [1] 2.523204  
## attr(,"method")  
## [1] "moment"

The series plot suggest this dataset is stationary around both mean and variance. To verify our deductions, we conducted a formal unit root test: the Augmented Dickey Fuller (ADF) with the following hypothesis:

*Ho: the process is not stationary.(i.e. We need a unit root)* *H1: the process is stationary.(i.e. No differences needed)*

**Stationarity Formal Test:**

## [1] 0

The ADF test indicates 0 required differences, hence enabling us to reject Ho and validate our previous inference. Furthermore, we are also capable of claiming no further data transformation will be required. Nevertheless, when analysing both ACF & PACF function we observe that, at a 95% confidence level, past information appears capable of contributing to future predictions (i.e. lags 1 and 3 are out of bounds and with significant magnitude). To formally assess the series autocorrelation, we opted to realise a Box-Ljung Test with the following hypotheses:

*Ho: Uncorrelated data* *H1: Correlated data*

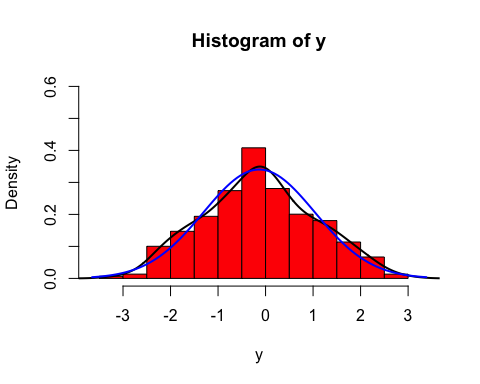
**Testing for White-Noise(WN)/Autocorrelation:**

##   
## Box-Ljung test  
##   
## data: y  
## X-squared = 148.71, df = 20, p-value < 2.2e-16

The extremely small p-value reported verifies our graphical analysis based on the ACF & PACF and conveys the idea that this time series is correlated with its past values. Consequently, we can infer our time series is not a WN process (nor SWN and GWN as result). Consequently, we can also conclude that a linear model is needed model/predict future values of this time series. Ultimately, requiring (or not) a non-linear model to analyse this time series will remain unknown since we must first isolate the linear part of the model to make such inference.

Now, we will try and identify if the series’ marginal distribution is normally distributed by recurring to both graphical tools and nornality tests (i.e. Shapiro and Jarque-Bera Tests).

### Testing for Normality:



According to our graphical analysis, the time series 3 (trasnformed) joint distribution could be normally distributed given the black (our series) and blue (normal distribution reference line N ~ (0,1)) have a similar shapes. We decided to verify our intuition with formal tests.

##   
## Shapiro-Wilk normality test  
##   
## data: y  
## W = 0.99344, p-value = 0.2178

##   
## Jarque-Bera test for normality  
##   
## data: y  
## JB = 2.8694, p-value = 0.198

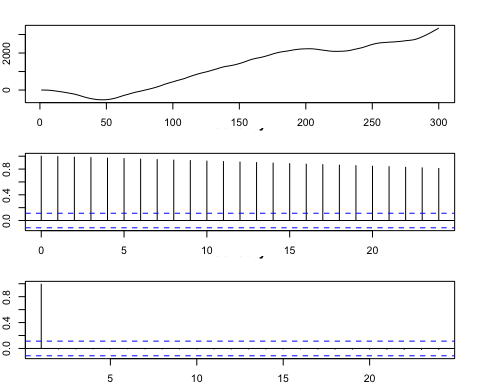
Both formal tests follow the same hypothesis structure:

*Ho: the data is normally distributed* *H1: the data is not normally distributed*

Both Jarque-Bera (p-value 0.20) and Shapiro tests (p-value 0.22) indicate we should accept the null hypothesis at a 95% confidence level. Therefore, we were able to validate the conclusions based on the team’s eye test. As per consequence, this series joint distribution is normally distributed.

## Time Series 4

### Graphical Representation of Time Series and respective ACF & PACF:



### Computation of basic descriptive statistics:

## [1] 1256.843

## [1] 1154.857

## [1] -0.1488262  
## attr(,"method")  
## [1] "moment"

## [1] 1.559977  
## attr(,"method")  
## [1] "moment"

The series plot suggests this dataset is not stationary around the mean (it exhibits a clear sustained upward trend), a result hinted by the series 1256.8 mean. Regarding the variance, the graph shows no clear and sustainable variance trend across time, hence we deemed series 2 variance stationary. To verify our deductions, we conducted a formal unit root test: the Augmented Dickey Fuller (ADF) with the following hypothesis:

*Ho: the process is not stationary.(i.e. We need a unit root)* *H1: the process is stationary.(i.e. No differences needed)*

**Stationarity Formal Test:**

## [1] 2

The ADF test indicates 2 differences are required for the data to become stationary in the mean, hence enabling us to accept Ho and validate our previous inference. When observing both ACF & PACF function we infer that, at a 95% confidence level, past information (especially lag 1) could prove to be extremely helpful to guide future forecasting efforts. To formally assess the series autocorrelation, we opted to realise a Box-Ljung Test with the following hypotheses:

*Ho: Uncorrelated data* *H1: Correlated data*

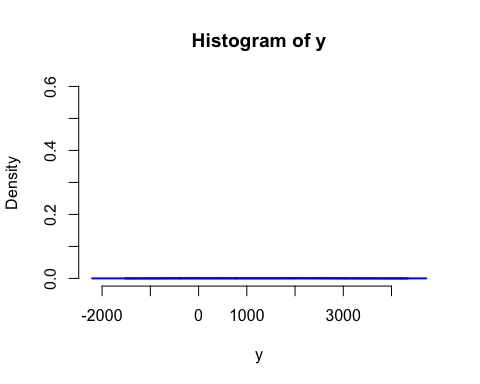
**Testing for White-Noise(WN)/Autocorrelation:**

##   
## Box-Ljung test  
##   
## data: y  
## X-squared = 5290.9, df = 20, p-value < 2.2e-16

The extremely small p-value reported confirmed our graphical analysis and we are now able to infer our time series is not a WN process of any kind. Consequently, we can also conclude that a linear model is needed model/predict future values of this time series. Ultimately, requiring (or not) a non-linear model to analyse this time series will remain unknown since we must first isolate the linear part of the model to make such inference.

Now, we will try and identify if the series’ marginal distribution is normally distributed by recurring to both graphical tools and nornality tests (i.e. Shapiro and Jarque-Bera Tests).

### Testing for Normality:



According to our graphical analysis, the time series 4 joint distribution cannot be normally distributed. We decided to verify our intuition with formal tests.

##   
## Shapiro-Wilk normality test  
##   
## data: y  
## W = 0.91161, p-value = 2.779e-12

##   
## Jarque-Bera test for normality  
##   
## data: y  
## JB = 26.665, p-value = 0.001

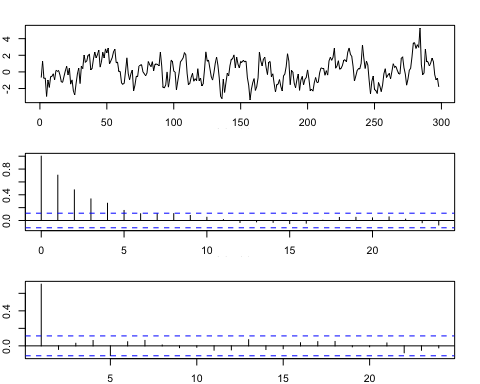
Both formal tests follow the same hypothesis structure:

*Ho: the data is normally distributed* *H1: the data is not normally distributed*

Both Jarque-Bera and Shapiro tests reported extremely small p-values, indicating we should reject the null hypothesis at a 95% confidence level. Therefore, we were able to validate the conclusions based on the team’s eye test.

## Time Series 4 (2 Difference Transformation):

### Graphical Representation of Time Series and respective ACF & PACF:



### Computation of basic descriptive statistics:

## [1] 0.1287819

## [1] 1.445643

## [1] 0.1289298  
## attr(,"method")  
## [1] "moment"

## [1] 2.761813  
## attr(,"method")  
## [1] "moment"

The series plot suggest this dataset is stationary around both mean and variance. To verify our deductions, we conducted a formal unit root test: the Augmented Dickey Fuller (ADF) with the following hypothesis:

*Ho: the process is not stationary.(i.e. We need a unit root)* *H1: the process is stationary.(i.e. No differences needed)*

**Stationarity Formal Test:**

## [1] 0

The ADF test indicates 0 required differences, hence enabling us to reject Ho and validate our previous inference. Furthermore, we are also capable of claiming no further data transformation will be needed. When analysing both ACF & PACF function we observe that, at a 95% confidence level, past information appears to contribute to futuristic predictions when it comes to the mean (i.e. all lags 1-2-3-4 are not within the confidence interval limit boundaries). To formally assess the series autocorrelation, we opted to realise a Box-Ljung Test with the following hypotheses:

*Ho: Uncorrelated data* *H1: Correlated data*

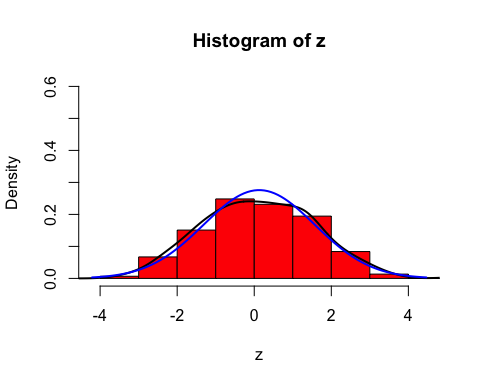
**Testing for White-Noise(WN)/Autocorrelation:**

##   
## Box-Ljung test  
##   
## data: z  
## X-squared = 297.76, df = 20, p-value < 2.2e-16

The test p.value of 2.2e-16 suggests that, at a 95% confindence level, we should reject the null hypothesis that the data is correlated and that the past is useful for forecasting purposes. Therefore, we feel comfortable suggesting this series has no white noise (WN) process. Because of that, the series will also not have strict white noise nor gaussian white noise. In addition, we can also conclude that a linear model is helpful to model/predict future values of this time series.Ultimately, requiring (or not) a non-linear model to analyse this time series will remain unknown since we must first isolate the linear part of the model to make such inference.

Now, we will try and identify if the series’ marginal distribution is normally distributed by recurring to both graphical tools and normality tests (i.e. Shapiro and Jarque-Bera Tests).

### Testing for Normality:



According to our graphical analysis, the time series 2 (trasnformed) joint distribution seems normally distributed given the black (our series) and blue (normal distribution reference line N ~ (0,1)) overlap significantly despite the Bell curve being higher than what exhibited by the time series 4.

##   
## Shapiro-Wilk normality test  
##   
## data: z  
## W = 0.99438, p-value = 0.3416

##   
## Jarque-Bera test for normality  
##   
## data: z  
## JB = 1.4325, p-value = 0.46

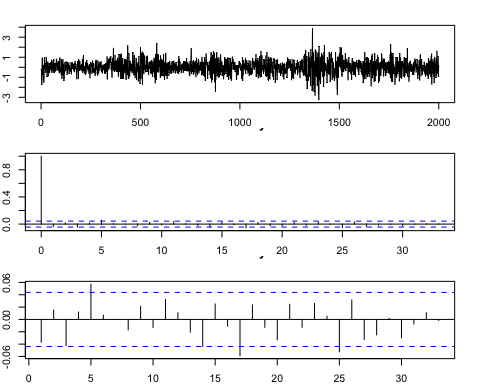
Both formal tests follow the same hypothesis structure:

*Ho: the data is normally distributed* *H1: the data is not normally distributed*

Both Jarque-Bera (p-value 0.4775) and Shapiro tests (p-value 0.3416) indicate we should accept the null hypothesis at a 95% confidence level. Therefore, we were able to validate the conclusions based on the team’s eye test. As per consequence,this series is marginally distributed.

## Time Series 5

### Graphical Representation of Time Series and respective ACF & PACF:



### Computation of basic descriptive statistics:

## [1] 0.0071755

## [1] 0.6797896

## [1] -0.08744834  
## attr(,"method")  
## [1] "moment"

## [1] 4.36924  
## attr(,"method")  
## [1] "moment"

The series plot suggests this dataset is stationary around the mean, a result hinted by the series mean of 0.0071. Regarding the variance, the graph shows no clear and sustainable variance trend across time (the few lags out of bound do not impact the R^2 in a significant manner), hence we deemed series 5 variance stationary. To verify our deductions, we conducted a formal unit root test: the Augmented Dickey Fuller (ADF) with the following hypothesis:

*Ho: the process is not stationary.(i.e. We need a unit root)* *H1: the process is stationary.(i.e. No differences needed)*

**Stationarity Formal Test:**

## [1] 0

The ADF test indicates 0 differences are required for the data to become stationary in the mean, hence enabling us to reject Ho and validate our previous inference. From this, we can also say that the data will not need transformation. When observing both ACF & PACF function we infer that, at a 95% confidence level, past information could not prove to be helpful to guide future forecasting efforts. To formally assess the series autocorrelation, we opted to realise a Box-Ljung Test with the following hypotheses:

*Ho: Uncorrelated data* *H1: Correlated data*

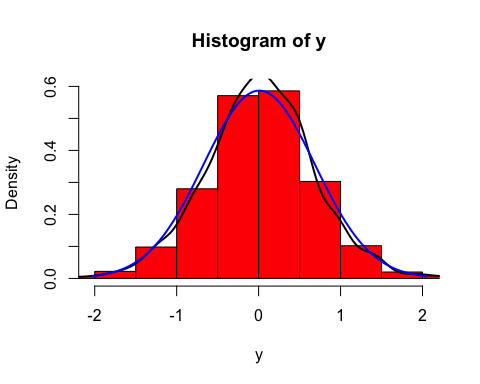
**Testing for White-Noise(WN)/Autocorrelation:**

##   
## Box-Ljung test  
##   
## data: y  
## X-squared = 33.151, df = 20, p-value = 0.03248

The p-value of 0.032 reported confirmed our graphical analysis and we are now able to infer our time series is a WN process. Because of this reason, we can also infer that a linear model is not needed to model/predict future values of this time series. Consequently, we can also conclude that a non linear model is needed to model/predict future values of time series 5.

Now, we will try and identify if the series’ marginal distribution is normally distributed by recurring to both graphical tools and normality tests (i.e. Shapiro and Jarque-Bera Tests).

### Testing for Normality:



According to our graphical analysis, the time series 5 joint distribution cannot be normally distributed. We decided to verify our intuition with formal tests.

##   
## Shapiro-Wilk normality test  
##   
## data: y  
## W = 0.99047, p-value = 3.456e-10

##   
## Jarque-Bera test for normality  
##   
## data: y  
## JB = 159.79, p-value < 2.2e-16

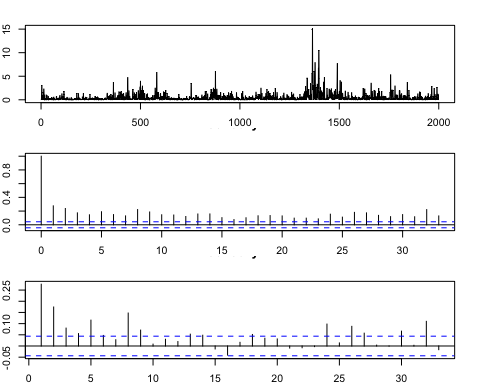
Both formal tests follow the same hypothesis structure:

*Ho: the data is normally distributed* *H1: the data is not normally distributed*

Both Jarque-Bera and Shapiro tests reported extremely small p-values (Jarque-Bera p-value < 2.2e-16, Shapiro p-value= 3.456e-10), indicating we should reject the null hypothesis at a 95% confidence level. Therefore, we were able to validate the conclusions based on the team’s eye test. Because of this, we can also infer that the series will not have Gaussian White noise.

### Testing for Strict White Noise:

# Graphically  
  
par(mar=c(2,2,2,2)) # to adjust graphic size  
  
par(mfrow=c(3,1)) # analysis of the squared data  
ts.plot(y^2)   
acf(y^2)  
pacf(y^2)



#Formal Tests:  
  
Box.test(y^2,lag=20, type="Ljung")

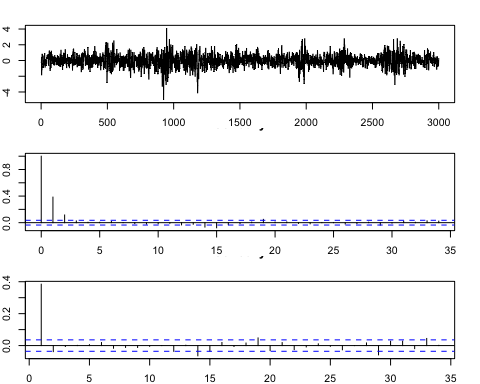
##   
## Box-Ljung test  
##   
## data: y^2  
## X-squared = 1061.5, df = 20, p-value < 2.2e-16

*Ho: uncorrelated data* *H1: correlated data*

When testing for strict white noise, the series plot suggests this dataset does not have strict white noise as there are many lags outside of bound in both the ACF and PACF. Our intuition is confirmed by the formal Box-Ljung test for the y^2. In fact, given a p-value < 2.2e-16, we reject the null hypothesis and find our data to be correlated. As there is not SWN, we can also exclude gaussian white noise from this series. Consequently, we can also conclude that a non linear model is needed to model/predict future values of this time series.

## Time Series 6

### Graphical Representation of Time Series and respective ACF & PACF:



### Computation of basic descriptive statistics:

## [1] 0.007541

## [1] 0.7605913

## [1] -0.2023789  
## attr(,"method")  
## [1] "moment"

## [1] 5.113352  
## attr(,"method")  
## [1] "moment"

The series plot suggests this dataset is stationary around the mean (it exhibits a constant trend), a result hinted by the series 0.0075 mean. Regarding the variance, the graph shows clear and sustainable variance trend across time, hence we deemed series 6 variance stationary. To verify our deductions, we conducted a formal unit root test: the Augmented Dickey Fuller (ADF) with the following hypothesis:

*Ho: the process is not stationary.(i.e. We need a unit root)* *H1: the process is stationary.(i.e. No differences needed)*

\*\*Stationarity Formal Test:

## [1] 0

The ADF test indicates 0 differences are required for the data to become stationary in the mean, hence enabling us to reject Ho and validate our previous inference.Because of this, no transformation to the data will be needed. When observing both ACF & PACF function we infer that, at a 95% confidence level, past information could are not helpful to guide future forecasting efforts. To formally assess the series autocorrelation, we opted to realise a Box-Ljung Test with the following hypotheses:

*Ho: Uncorrelated data* *H1: Correlated data*

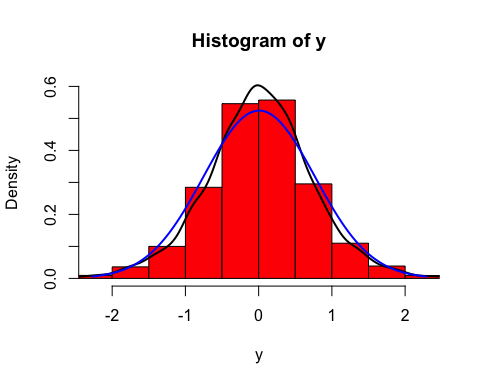
**Testing for White-Noise(WN)/Autocorrelation:**

##   
## Box-Ljung test  
##   
## data: y  
## X-squared = 537.69, df = 20, p-value < 2.2e-16

The extremely small p-value < 2.2e-16 reported confirmed our graphical analysis and we are now able to infer our time series is not a WN process of any kind (also no SWN nor GWN). Consequently, we can also conclude that a linear model is needed to model/predict future values of this time series. Ultimately, requiring (or not) a non-linear model to analyse this time series will remain unknown since we must first isolate the linear part of the model to make such inference.

Now, we will try and identify if the series’ marginal distribution is normally distributed by recurring to both graphical tools and normality tests (i.e. Shapiro and Jarque-Bera Tests).

### Testing for Normality:

 According to our graphical analysis, the time series 6 joint distribution cannot be normally distributed. We decided to verify our intuition with formal tests.

##   
## Shapiro-Wilk normality test  
##   
## data: y  
## W = 0.98338, p-value < 2.2e-16

##   
## Jarque-Bera test for normality  
##   
## data: y  
## JB = 580.58, p-value < 2.2e-16

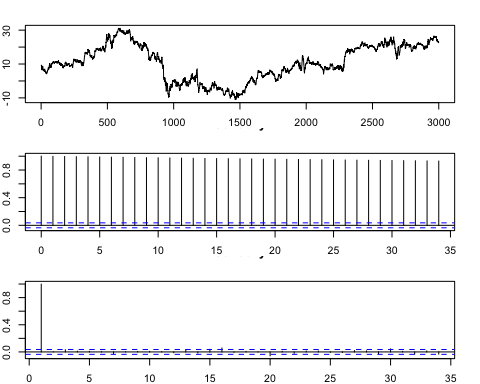
Both formal tests follow the same hypothesis structure:

*Ho: the data is normally distributed* *H1: the data is not normally distributed*

Both Jarque-Bera and Shapiro tests reported extremely small p-values < 2.2e-16, indicating we should reject the null hypothesis at a 95% confidence level. Therefore, we were able to validate the conclusions based on the team’s eye test.

## Time Series 7

### Graphical Representation of Time Series and respective ACF & PACF:



### Computation of basic descriptive statistics:

## [1] 10.88196

## [1] 10.28588

## [1] -0.1702387  
## attr(,"method")  
## [1] "moment"

## [1] 2.00975  
## attr(,"method")  
## [1] "moment"

The series plot suggests this dataset is not stationary around the mean (it exhibits a clear sustained upward trend), a result hinted by the series 10.88196 mean. Regarding the variance, the graph shows no clear and sustainable variance trend across time, hence we deemed series 7 variance stationary. To verify our deductions, we conducted a formal unit root test: the Augmented Dickey Fuller (ADF) with the following hypothesis:

*Ho: the process is not stationary.(i.e. We need a unit root)* *H1: the process is stationary.(i.e. No differences needed)*

\*\*Stationarity Formal Test:

## [1] 1

The ADF test indicates 1 difference is required for the data to become stationary in the mean, hence enabling us to accept Ho and validate our previous inference. This also means the data will need transformation. When observing both ACF & PACF function we infer that, at a 95% confidence level, past information prove to be extremely helpful to guide future forecasting efforts. To formally assess the series autocorrelation, we opted to realise a Box-Ljung Test with the following hypotheses:

*Ho: Uncorrelated data* *H1: Correlated data*

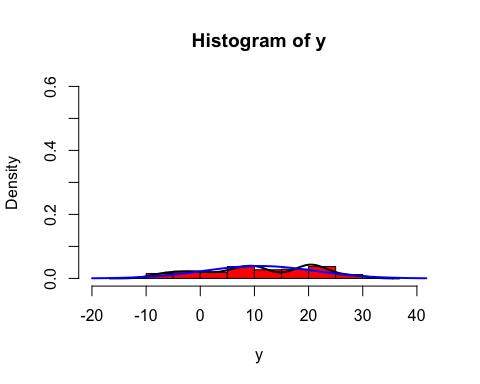
**Testing for White-Noise(WN)/Autocorrelation:**

##   
## Box-Ljung test  
##   
## data: y  
## X-squared = 57372, df = 20, p-value < 2.2e-16

The extremely small p-value < 2.2e-16 reported confirmed our graphical analysis and we are now able to infer our time series is not a WN process of any kind (also no SWN nor GWN). Consequently, we can also conclude that a linear model is needed model/predict future values of this time series. Ultimately, requiring (or not) a non-linear model to analyse this time series will remain unknown since we must first isolate the linear part of the model to make such inference.

Now, we will try and identify if the series’ marginal distribution is normally distributed by recurring to both graphical tools and nornality tests (i.e. Shapiro and Jarque-Bera Tests).

### Testing for Normality:

 According to our graphical analysis, the time series 7 joint distribution cannot be normally distributed. We decided to verify our intuition with formal tests.

##   
## Shapiro-Wilk normality test  
##   
## data: y  
## W = 0.96711, p-value < 2.2e-16

##   
## Jarque-Bera test for normality  
##   
## data: y  
## JB = 136.75, p-value < 2.2e-16

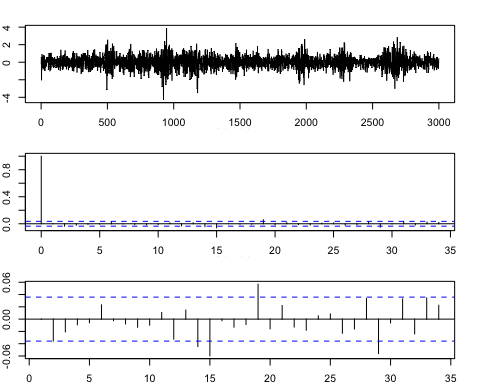
Both formal tests follow the same hypothesis structure:

*Ho: the data is normally distributed* *H1: the data is not normally distributed*

Both Jarque-Bera and Shapiro tests reported extremely small p-values < 2.2e-16, indicating we should reject the null hypothesis at a 95% confidence level. Therefore, we were able to validate the conclusions based on the team’s eye test.

## Time Series 7 (1 Difference Transformation):

### Graphical Representation of Time Series and respective ACF & PACF:



### Computation of basic descriptive statistics:

## [1] 0.004649883

## [1] 0.7023493

## [1] -0.1525146  
## attr(,"method")  
## [1] "moment"

## [1] 4.823746  
## attr(,"method")  
## [1] "moment"

The series plot suggest this dataset is stationary around both the mean but the variance (there are not significant lags out of bound). This is also supported by a mean value of 0.0004. To verify our deductions, we conducted a formal unit root test: the Augmented Dickey Fuller (ADF) with the following hypothesis:

*Ho: the process is not stationary.(i.e. We need a unit root)* *H1: the process is stationary.(i.e. No differences needed)*

\*\*Stationarity Formal Test:

## [1] 0

The ADF test indicates 0 required differences, hence enabling us to reject Ho and validate our previous inference. Furthermore, we are also capable of claiming no further data transformation will be needed. When analysing both ACF & PACF function we observe that, at a 95% confidence level, past information appears to not contribute to futuristic predictions when it comes to the variance (i.e. lags are within the confidence interval limit boundaries and, those who are not, do not have a high impact on the R^2). To formally assess the series autocorrelation, we opted to realise a Box-Ljung Test with the following hypotheses:

*Ho: Uncorrelated data* *H1: Correlated data*

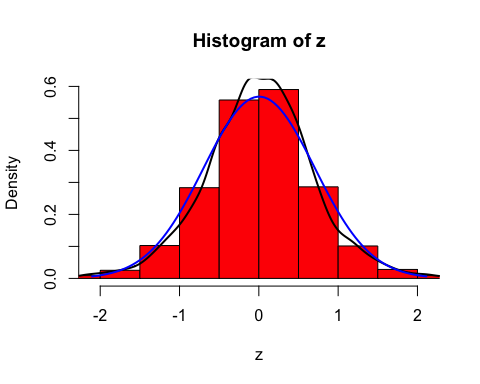
**Testing for White-Noise(WN)/Autocorrelation:**

##   
## Box-Ljung test  
##   
## data: z  
## X-squared = 40.012, df = 20, p-value = 0.004978

The test p-value= 0.0049 suggests that, at a 95% confindence level, we should not reject the null hypothesis, therefore the data is uncorrelated and the past is not useful for forecasting purposes. Therefore, we feel comfortable suggesting this series has white noise (WN) process. In addition, we can also conclude that a linear model is not helpful to model/predict future values of this time series. On the other hand, a non linear model will be useful to model/predict future values.

Now, we will try and identify if the series’ marginal distribution is normally distributed by recurring to both graphical tools and normality tests (i.e. Shapiro and Jarque-Bera Tests).

### Testing for Normality:

 According to our graphical analysis, the time series 7 (transformed) joint distribution does not seem normally distributed given the black (our series) and blue (normal distribution reference line N ~ (0,1)) do not overlap and the Bell curve way lower than what exhibited by the time series 7.

##   
## Shapiro-Wilk normality test  
##   
## data: z  
## W = 0.98584, p-value < 2.2e-16

##   
## Jarque-Bera test for normality  
##   
## data: z  
## JB = 428.72, p-value < 2.2e-16

Both formal tests follow the same hypothesis structure:

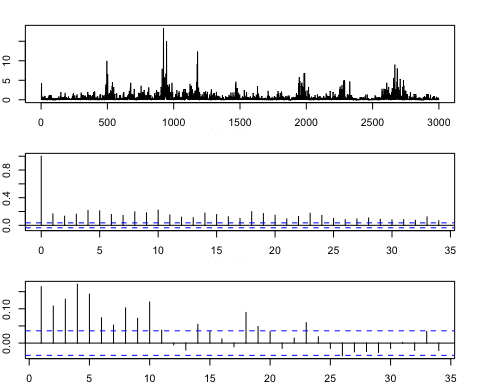
*Ho: the data is normally distributed* *H1: the data is not normally distributed*

Both Jarque-Bera and Shapiro tests (p-value < 2.2e-16) indicate we should reject the null hypothesis at a 95% confidence level. Therefore, we were able to validate the conclusions based on the team’s eye test. As per consequence,this series is not normally distributed. Because of that, in this series, we do not have Gaussian White Noise (GWN).

We are finally going to test for Strict White Noise:

### Testing for Strict White Noise:

# Graphically:  
  
par(mar=c(2,2,2,2)) # to adjust graphic size  
  
par(mfrow=c(3,1)) # analysis of the squared data  
ts.plot(z^2)   
acf(z^2)  
pacf(z^2)



#Formal Tests:  
  
Box.test(z^2,lag=20, type="Ljung")

##   
## Box-Ljung test  
##   
## data: z^2  
## X-squared = 1632.1, df = 20, p-value < 2.2e-16

*Ho: uncorrelated data* *H1: correlated data*

When testing for strict white noise, the series plot suggests this dataset does not have strict white noise as there are many lags outside of bound in both, the ACF and PACF. Our intuition is confirmed by the formal Box-Ljung test for the z^2. In fact, given a p-value < 2.2e-16, we reject the null hypothesis and find our data to be correlated.

###### THE END