# **INTRODUCTION - UNIT ROOT TESTING IN R**

### 1. INTRODUCTION

### 1.1. Downloading and Installing R

R is a widely used package for statistical analysis. The difference between R and many other statistical packages is that it is free software. R is used within a command - line interface. R is distributed by the "Comprehensive R Archive Network" (CRAN) - it is available from: http://cran.r-project.org. R can be installed by executing the downloaded file. The installation procedure is straightforward, one usually only has to specify the target directory in which to install R. After the installation, R can be started like any other application for Windows. That is, by double - clicking on the corresponding icon.

#### 1.2. Import Data in R

Importing data into R can be carried out in various ways. Below, the command read.table is used:

mydatats1<-read.table("E:/Loukia/Teaching/TimeSeries/Notes-R/datats.txt")

y <- mydatats1\$V1

Let us create a time series object using the function "ts" from a vector - single time-series or a matrix - multivariate time-series. The data consist of the Johnson & Johnson quarterly earnings per share from 1960:1 to 1980:4, i.e. 84 quarters.

j=ts(y, frequency=4, start = c(1960,1))

#### 1.3. Basic Plots in R

A time series plot can be created using the plot() command. The plot() command allows for a number of optional arguments: the option type="I" sets the plot-type to "lines", the option lwd=2 (line width) controls the thickness of the plotted line, the option col="red" controls the colour used, the options xlab and ylab are used for labeling the axes, while main specifies the title of the plot.

For example, the time series plot for the Johnson & Johnson data is given by:

plot(j, type="l", col='red', lwd=1, main="Time Series plot of Johnson & Johnson data", ylab="Quarterly earnings per share")

Time Series plot of Johnson & Johnson



Figure 1: Time series plot of J&J

Every R function has a corresponding help file, which can be accessed by typing a question mark and the command. It contains further details about the function and available options, references and examples of usage. For example, by typing ?plot into the console opens the help file for the plot() command.

The hist() command creates a histogram of the data vector. Below, the time series plot of the J&J data, of the log of J&J and of the first differences of the log of J&J data, together with their corresponding histograms are presented. The log of J&J time series and the first differences of the log of J&J data are computed by:

```
lj=log(j) # compute the logartithm of the J&J data
```

dlj=diff(lj) # compute the first differences of the log of J&J data

A nice graph can be produced by using the command par(mfrow=c(rows,col)), which splits the graph into (rows x col) subplots:

par(mfrow=c(3,2)) # set up the graphics plot(j,type="l", col='red', lwd=1,main="Time Series plot of Johnson & Johnson", ylab="Quarterly earnings per share") hist(j, nclass=15, main="Histogram of Johnson & Johnson") plot(lj,type="l", col='red', lwd=1,main="Log of Johnson & Johnson", ylab="Log of Quarterly earnings per share") hist(lj, nclass=15, main="Histogram of log of Johnson & Johnson") plot(dlj,type="l", col='red', lwd=1,main="Differences of log of Johnson & Johnson") hist(dlj, nclass=15, main="Histogram of differences of log of J&J")



Figure 2: Time series plots and histograms for the J&J, log of J&J and the differences of log(J&J)

The autocorrelation and partial autocorrelation plots are useful to examine if there is dependence between lagged values of the analyzed series. The acf() command and the pacf() command create an autocorrelation and a partial autocorrelation plot, respectively. Below, the autocorrelation and partial autocorrelation plots of the J&J data, the log of J&J and of the first differences of the log of J&J data are presented using the command par(mfrow=c(3,2)).

par(mfrow=c(3,2)) # set up the graphics acf(j, 48, main="ACF of J&J") # autocorrelation function plot pacf(j, 48, main="PACF of J&J") # partial autocorrelation function acf(lj, 48, main="ACF of log of J&J")
pacf(lj, 48, main="PACF of log of J&J")
acf(dlj, 48, main="ACF of differences of log of J&J")
pacf(dlj, 48, main="PACF of differences of log of J&J")



Figure 3: Autocorrelations (ACF) and partial autocorrelations (PACF) for the J&J, log of J&J and the differences of log(J&J)

Note that the lag values in the X axis are 1, 2, 3, 4, 5,... and correspond to lags 4, 8, 12, 16, 20,... because we have quarterly data, i.e. the frequency is 4. A better type of labeling can be produced by using the following set of commands:

par(mfrow=c(3,2)) # set up the graphics





Figure 4: Autocorrelations (ACF) and partial autocorrelations (PACF) for the J&J, log of J&J and the differences of log(J&J)

Note that in the autocorrelation plots presented above, the dashed lines are the approximate two standard error confidence bounds computed by  $\pm 1.96 \cdot (1/\sqrt{T})$ , where T is the number observations. If

the autocorrelation is within these bounds, it is not significantly different from zero at (approximately) 5% level of significance [Bartlett test].

The Box.test() command can be used to compute the Box-Pierce or the Ljung-Box test statistic for examining the null hypothesis that the autocorrelations of a given time series are zero. The command is: Box.test(x, lag, type = c("Box-Pierce", "Ljung-Box")), where "x" is the analyzed time series, "lag" denotes the number of lags at which the statistic will be computed, while "type" determines the Box-Pierce or the Ljung-Box test statistic. For example, by running the commands

res1=Box.test(j,48,type="Box-Pierce")

res2=Box.test(j,48,type="Ljung-Box")

will give the following results:

res1: Box-Pierce test data: j X-squared = 695.06, df = 48, p-value < 2.2e-16

res2: Box-Ljung test data: j X-squared = 853.09, df = 48, p-value < 2.2e-16

### 2. UNIT ROOT TESTING

To perform a unit-root test, the command ur.df(y, type = c("none", "drift", "trend"), lags = 1, selectlags = c("Fixed", "AIC", "BIC")) can be used; "y" is the time series to be tested for a unit root, "type" corresponds to the three fitted models, i.e. a model without constant/trend (none), a model with constant only (drift), and a model with constant and time trend (trend). "lags" denotes the maximum number of lags for endogenous variable to be included, selectlags denotes the lag selection which can be achieved according to the Akaike "AIC" or the Bayes "BIC" information criteria.

First, the package "urca" is installed (or is loaded if it is already installed):

```
install.packages("urca")
library(urca)
```

# Example 1: Unit root testing to the J&J time series

First, we can fit an autoregressive time series model to the J&J data, by selecting the complexity of the model based on AIC. Then, we perform an augmented Dickey-Fuller test of unit root, based on a model with constant and trend (see figure 2, time series plot for the J&J series):

m=ar(j)
m
m\$order
m1=ur.df(j,type="trend",lags=m\$order-1)
m1
summary(m1)

The results taken from R are presented below:

m1

######################################
The value of the test statistic is: 1.9321 16.7049 19.2758
summary(m1)
#######################################
# Augmented Dickey-Fuller Test Unit Root Test # ###################################
Test regression trend
Call:
lm(formula = z.diff ~ z.lag.1 + 1 + tt + z.diff.lag)
Residuals:
Min 1Q Median 3Q Max
-1.27266 -0.17348 0.01381 0.12299 1.18302
Coefficients:
Estimate Std. Error t value Pr(> t )
(Intercept) -0.088498 0.136166 -0.650 0.5178
Z.ldg.1 U.U/3336 U.U3/956 1.932 U.U5/3.
II U.UIUU52 U.UU53U4 I.595 U.II52 z diff log1 1 060854 0 121507 8 125 8 570 12 ***
2.011.1ag1 - 1.003034 - 0.131307 - 0.133 - 0.76-12
$2.011.1ag^{2} - 1.012.588 = 0.145600 - 0.545 1.412-05$ z diff lag3 -1 006500 = 0.143949 -6 992 1 14e-09 ***
z diff lag4 0.092346 0.141368 0.653 0.5157
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
Residual standard error: 0.4127 on 72 degrees of freedom
Multiple R-squared: 0.9262. Adjusted R-squared: 0.92
F-statistic: 150.5 on 6 and 72 DF, p-value: < 2.2e-16
Value of test-statistic is: 1.9321 16.7049 19.2758
Critical values for test statistics:
1pct 5pct 10pct
tau3 -4.04 -3.45 -3.15

phi2 6.50 4.88 4.16		
phi3 8.73 6.49 5.47		

Similar results are obtained by using EVIEWS:

Null Hypothesis: J has a unit root Exogenous: Constant, Linear Trend Lag Length: 4 (Fixed)					
			t-Statistic	Prob.*	
Augmented Dickey-Fu	ller test statisti	с	1.932145	1.0000	
Test critical values:	1% level 5% level 10% level		-4.078420 -3.467703 -3.160627		
*MacKinnon (1996) one-sided p-values.					
Augmented Dickey-Fuller Test Equation Dependent Variable: D(J) Method: Least Squares Sample (adjusted): 1961Q2 1980Q4 Included observations: 79 after adjustments					
Variable	Coefficient	Std. Error	t-Statistic	Prob.	
J(-1) D(J(-1)) D(J(-2)) D(J(-3)) D(J(-4))	0.073336 -1.069854 -1.012388 -1.006500 0.092346	0.037956 0.131507 0.145806 0.14949 0.141368	1.932145 -8.135353 -6.943389 -6.992067 0.653236	0.0573 0.0000 0.0000 0.0000	

Obviously, the null hypothesis of non-stationarity for the J&J series is not rejected. Thus the J&J is not stationary. Next, we can test for a unit root for the logarithms of the J&J. We perform an augmented Dickey-Fuller test of unit root, based on a model with constant and trend (see figure 2, time series plot for the log(J&J) series):

#Unit root testing for the log(j)
m1=ur.df(lj,type="trend",lags=5)
m1

#### summary(m1)

The results taken from R are presented below:

m1

summary(m1)

```
*****
# Augmented Dickey-Fuller Test Unit Root Test #
Test regression trend
Call:
Im(formula = z.diff \sim z.lag.1 + 1 + tt + z.diff.lag)
Residuals:
         1Q Median 3Q Max
  Min
-0.18785 -0.04923 -0.00168 0.04606 0.20635
Coefficients:
      Estimate Std. Error t value Pr(>|t|)
(Intercept) -0.037988 0.105676 -0.359 0.720323
z.lag.1 -0.179098 0.124646 -1.437 0.155214
       0.007347 0.005337 1.377 0.173026
tt
z.diff.lag1 -0.589245 0.158531 -3.717 0.000403 ***
z.diff.lag2 -0.416588 0.164498 -2.532 0.013571 *
z.diff.lag3 -0.429622 0.143892 -2.986 0.003896 **
z.diff.lag4 0.380827 0.133094 2.861 0.005558 **
z.diff.lag5 0.154810 0.106817 1.449 0.151721
---
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 '' 1
Residual standard error: 0.08266 on 70 degrees of freedom
Multiple R-squared: 0.8382,
                             Adjusted R-squared: 0.822
F-statistic: 51.8 on 7 and 70 DF, p-value: < 2.2e-16
Value of test-statistic is: -1.4369 6.8046 1.2869
Critical values for test statistics:
   1pct 5pct 10pct
tau3 -4.04 -3.45 -3.15
phi2 6.50 4.88 4.16
phi3 8.73 6.49 5.47
```

Similar results are obtained by using EVIEWS:

Null Hypothesis: LJ ha	s a unit root				
Exogenous: Constant, Linear Trend					
Lag Length: 5 (Automatic based on AIC, MAXLAG=11)					
			t-Statistic	Prob.*	
Augmented Dickey-Fu	Iller test statist	ic	-1.436854	0.8423	
Test critical values:	1% level		-4.080021		
	5% level		-3.468459		
	10% level		-3.161067		
*MacKinnon (1996) o	ne-sided p-valu	es.			
Augmented Dickey-Fuller Test Equation Dependent Variable: D(L) Method: Least Squares Sample (adjusted): 1961Q3 1980Q4 Included observations: 78 after adjustments					
Variable	Coefficient	Std. Error	t-Statistic	Prob.	
LJ(-1)	-0.179098	0.124646	-1.436854	0.1552	
D(LJ(-1))	-0.589245	0.158531	-3.716900	0.0004	
D(LJ(-2))	-0.416588	0.164498	-2.532476	0.0136	
D(LJ(-3))	-0.429622	0.143892	-2.985722	0.0039	
D(LJ(-4))	0.380827	0.133094	2.861342	0.0056	
D(LJ(-5))	0.154810	0.106817	1.449295	0.1517	
С	-0.037988	0.105676	-0.359474	0.7203	
@TREND(1960Q1)	0.007347	0.005337	1.3765 <u></u> 92	0.1730	

The null hypothesis of non-stationarity for the log(J&J) series is not rejected. Thus the log(J&J) is not stationary. Note, however, that if the number of the lagged variables in the augmented Dickey-Fuller model changes (for example, if the order is 1), the result of the unit root test is completely different. Maybe alternative models or more powerful unit root techniques could be used.

Next, we can test for a unit root for the differences of logarithms of the J&J. We perform an augmented Dickey-Fuller test of unit root, based on a model with constant (see figure 2, time series plot for the differences of log(J&J) series):

m=ar(dlj) m m\$order m1=ur.df(dlj,type="drift",lags=5)

#### summary(m1)

The results taken from R are presented below:

#### m1

summary(m1)

```
*****
# Augmented Dickey-Fuller Test Unit Root Test #
*****
Test regression drift
Call:
Im(formula = z.diff \sim z.lag.1 + 1 + z.diff.lag)
Residuals:
   Min
          1Q Median
                         3Q
                               Max
-0.189424 - 0.049737 - 0.008657 \ 0.048398 \ 0.207341
Coefficients:
      Estimate Std. Error t value Pr(>|t|)
(Intercept) 0.10117 0.02537 3.988 0.000162 ***
z.lag.1 -2.53130 0.58635 -4.317 5.11e-05 ***
z.diff.lag1 0.80961 0.53999 1.499 0.138290
z.diff.lag2 0.32762 0.45641 0.718 0.475258
z.diff.lag3 -0.22525 0.32533 -0.692 0.490988
z.diff.lag4 0.06502 0.21591 0.301 0.764209
z.diff.lag5 0.12319 0.10708 1.150 0.253900
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 '' 1
Residual standard error: 0.08322 on 70 degrees of freedom
Multiple R-squared: 0.9438,
                             Adjusted R-squared: 0.939
F-statistic: 196.1 on 6 and 70 DF, p-value: < 2.2e-16
Value of test-statistic is: -4.317 9.3192
Critical values for test statistics:
   1pct 5pct 10pct
tau2 -3.51 -2.89 -2.58
phi1 6.70 4.71 3.86
```

Similar results are obtained by using EVIEWS:

Null Hypothesis: DLJ has a unit root Exogenous: Constant Lag Length: 4 (Automatic based on AIC, MAXLAG=11)					
			t-Statistic	Prob.*	
Augmented Dickey-Fuller test statistic			-4.569037	0.0004	
Test critical values:	1% level		-3.516676		
	5% level		-2.899115		
	10% level		-2.586866		
*MacKinnon (1996) one-sided p-values. Dependent Variable: D(DLJ)					
Variable	Coefficient	Std. Error	t-Statistic	Prob.	
DLJ(-1)	-2.338380	0.511788	-4.569037	0.0000	
D(DLJ(-1))	0.600855	0.447643	1.342263	0.1837	
D(DLJ(-2))	0.054690	0.323307	0.169157	0.8661	
D(DLJ(-3))	-0.460666	0.207464	-2.220465	0.0295	
D(DLJ(-4))	-0.126099	0.105703	-1.192958	0.2368	
С	0.094353	0.022428	4.206999	0.0001	

Therefore the differences of log(J&J) seem to be a stationary process.

Alternatively, other R commands can be used. The R package 'tseries' is used to test the null hypothesis that a series has a unit root, versus the alternative hypothesis that the process is stationary. We test the null hypothesis using the available Dickey-Fuller (DF), Augmented Dickey-Fuller (ADF) and Phillips-Perron (PP) tests; note that in each case, the general regression equation incorporates a constant and a linear trend. In the ADF test, the default number of AR components included in the model, say k, is  $[[(n-1)^{(1/3)}]]$ , which corresponds to the suggested upper bound on the rate at which the number of lags, k, should be made to grow with the sample size for the general ARMA(p,q) setup. For the PP test, the default value of k is  $[[0.04n^{(1/4)}]]$ . The distinction between the two tests (DF and PP) is that the Phillips-Perron procedure estimates the autocorrelations in the stationary process directly (using a kernel smoother) rather than assuming an AR approximation, and for this reason the Phillips-Perron test is described as semi-parametric. To implement the above unit root tests, first, we load the package 'tseries' by using the command library(tseries), and then we perform the unit root test under consideration by using one of the following commands: adf.test(y, k=0) for the DF test, adf.test(y) for the ADF test, or the command pp.test(y) for the PP test. For example:

# adf.test(y,k=4)

Augmented Dickey-Fuller Test, data:dlj,

Dickey-Fuller = -4.5649, Lag order = 4, p-value = 0.01, alternative hypothesis: stationarity.

# Example 2: Unit root testing to the VIX time series

Consider the VIX of CBOE from 2004 to 2008. The data are obtained from the CBOE web site. First, we read the data, and obtain a time series plot:

data<- read.table("E:/Loukia/Teaching/TimeSeries/Notes-R/vix08.txt",header=T) dim(data) data[1,] y=data[,7] plot(y,type="l")



Fit an AR model and run the Augmented Dickey-Fuller test:

m=ar(y) m m\$order m1=ur.df(y,type="drift",lags=10)

summary(m1)

#### m1

summary(m1)

```
# Augmented Dickey-Fuller Test Unit Root Test #
*****
Test regression drift
Call:
Im(formula = z.diff \sim z.lag.1 + 1 + z.diff.lag)
Residuals:
 Min
       1Q Median 3Q Max
-5.6122 -0.5330 -0.1247 0.4036 7.1974
Coefficients:
       Estimate Std. Error t value Pr(>|t|)
(Intercept) 0.274964 0.121815 2.257 0.0242 *
z.lag.1 -0.017346 0.007633 -2.273 0.0233 *
z.diff.lag1 -0.193620 0.031144 -6.217 7.29e-10 ***
z.diff.lag2 -0.108027 0.031777 -3.400 0.0007 ***
z.diff.lag3 -0.011997 0.031538 -0.380 0.7037
z.diff.lag4 -0.051623 0.031477 -1.640 0.1013
z.diff.lag5 -0.027070 0.031422 -0.862 0.3892
z.diff.lag6 -0.064092 0.031399 -2.041 0.0415 *
z.diff.lag7 -0.043289 0.031373 -1.380 0.1679
z.diff.lag8 -0.145963 0.031360 -4.654 3.66e-06 ***
z.diff.lag9 0.017763 0.031417 0.565 0.5719
z.diff.lag10 0.061834 0.030741 2.011 0.0445 *
---
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 '' 1
Residual standard error: 1.089 on 1057 degrees of freedom
                             Adjusted R-squared: 0.07196
Multiple R-squared: 0.08152,
F-statistic: 8.529 on 11 and 1057 DF, p-value: 1.48e-14
Value of test-statistic is: -2.2726 2.6163
Critical values for test statistics:
   1pct 5pct 10pct
tau2 -3.43 -2.86 -2.57
phi1 6.43 4.59 3.78
```

m2=ur.df(y,type="trend",lags=10)

summary(m2)

#### m2

#### summary(m2)

```
# Augmented Dickey-Fuller Test Unit Root Test #
Test regression trend
Call:
Im(formula = z.diff \sim z.lag.1 + 1 + tt + z.diff.lag)
Residuals:
 Min
       1Q Median 3Q Max
-5.6281 -0.5331 -0.1150 0.4262 7.1234
Coefficients:
       Estimate Std. Error t value Pr(>|t|)
(Intercept) 0.2603461 0.1220682 2.133 0.033172 *
z.lag.1
       -0.0231909 0.0084572 -2.742 0.006207 **
       0.0001916 0.0001198 1.600 0.109987
tt
z.diff.lag1 -0.1902722 0.0311915 -6.100 1.49e-09 ***
z.diff.lag2 -0.1054542 0.0317940 -3.317 0.000942 ***
z.diff.lag3 -0.0100074 0.0315391 -0.317 0.751078
z.diff.lag4 -0.0496545 0.0314776 -1.577 0.114991
z.diff.lag5 -0.0252114 0.0314203 -0.802 0.422507
z.diff.lag6 -0.0622962 0.0313956 -1.984 0.047488 *
z.diff.lag7 -0.0417138 0.0313650 -1.330 0.183823
z.diff.lag8 -0.1446717 0.0313477 -4.615 4.41e-06 ***
z.diff.lag9 0.0184760 0.0313970 0.588 0.556347
z.diff.lag10 0.0622955 0.0307192 2.028 0.042821 *
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 '' 1
Residual standard error: 1.088 on 1056 degrees of freedom
Multiple R-squared: 0.08374, Adjusted R-squared: 0.07333
F-statistic: 8.043 on 12 and 1056 DF, p-value: 1.422e-14
Value of test-statistic is: -2.7422 2.5997 3.8655
Critical values for test statistics:
   1pct 5pct 10pct
tau3 -3.96 -3.41 -3.12
phi2 6.09 4.68 4.03
phi3 8.27 6.25 5.34
```

m3=ur.df(y,type="none",lags=10)

summary(m3)

#### m3

#### summary(m3)

```
# Augmented Dickey-Fuller Test Unit Root Test #
*****
Test regression none
Call:
Im(formula = z.diff \sim z.lag.1 - 1 + z.diff.lag)
Residuals:
  Min 1Q Median 3Q Max
-5.7480 -0.4964 -0.0723 0.4372 7.2926
Coefficients:
       Estimate Std. Error t value Pr(>|t|)
z.lag.1 -0.000774 0.002091 -0.370 0.711372
z.diff.lag1 -0.205749 0.030736 -6.694 3.52e-11 ***
z.diff.lag2 -0.119423 0.031434 -3.799 0.000153 ***
z.diff.lag3 -0.022077 0.031280 -0.706 0.480471
z.diff.lag4 -0.061325 0.031242 -1.963 0.049922 *
z.diff.lag5 -0.035995 0.031232 -1.152 0.249386
z.diff.lag6 -0.072692 0.031227 -2.328 0.020107 *
z.diff.lag7 -0.051163 0.031238 -1.638 0.101756
z.diff.lag8 -0.153488 0.031243 -4.913 1.04e-06 ***
z.diff.lag9 0.011601 0.031359 0.370 0.711506
z.diff.lag10 0.056654 0.030714 1.845 0.065378.
---
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
Residual standard error: 1.091 on 1058 degrees of freedom
Multiple R-squared: 0.07711, Adjusted R-squared: 0.06752
F-statistic: 8.037 on 11 and 1058 DF, p-value: 1.406e-13
Value of test-statistic is: -0.3701
Critical values for test statistics:
   1pct 5pct 10pct
tau1 -2.58 -1.95 -1.62
```