

# STAT443 - Forecasting - Final Project

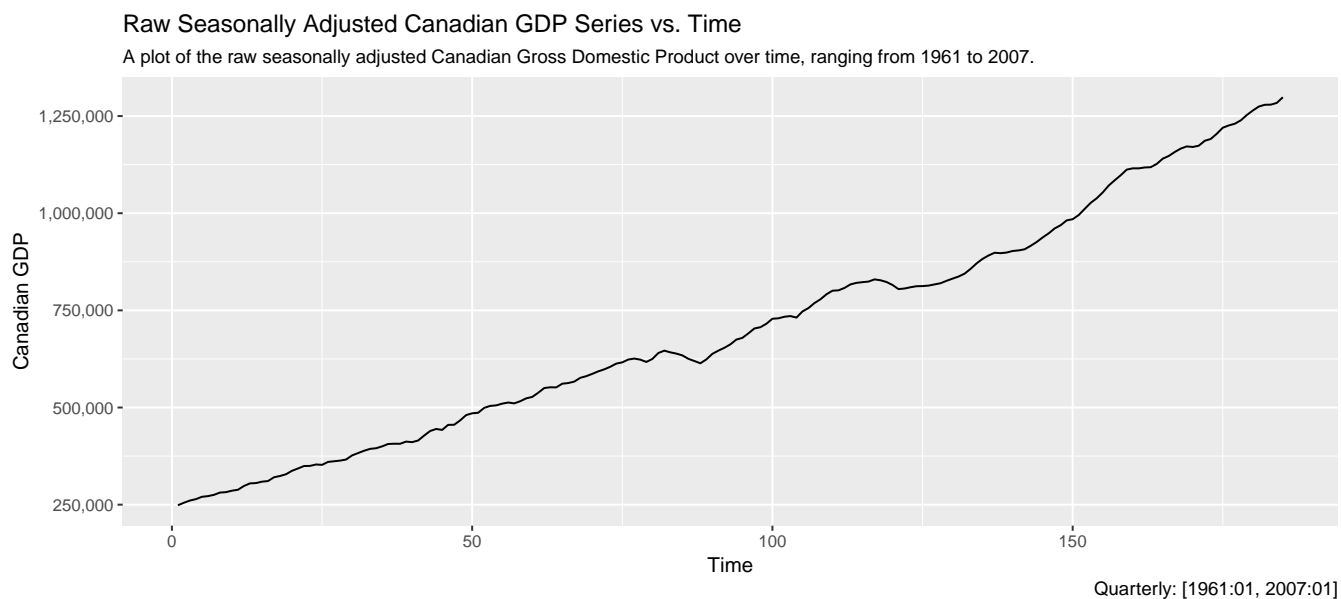
*Christopher Risi*

*July 24, 2018*

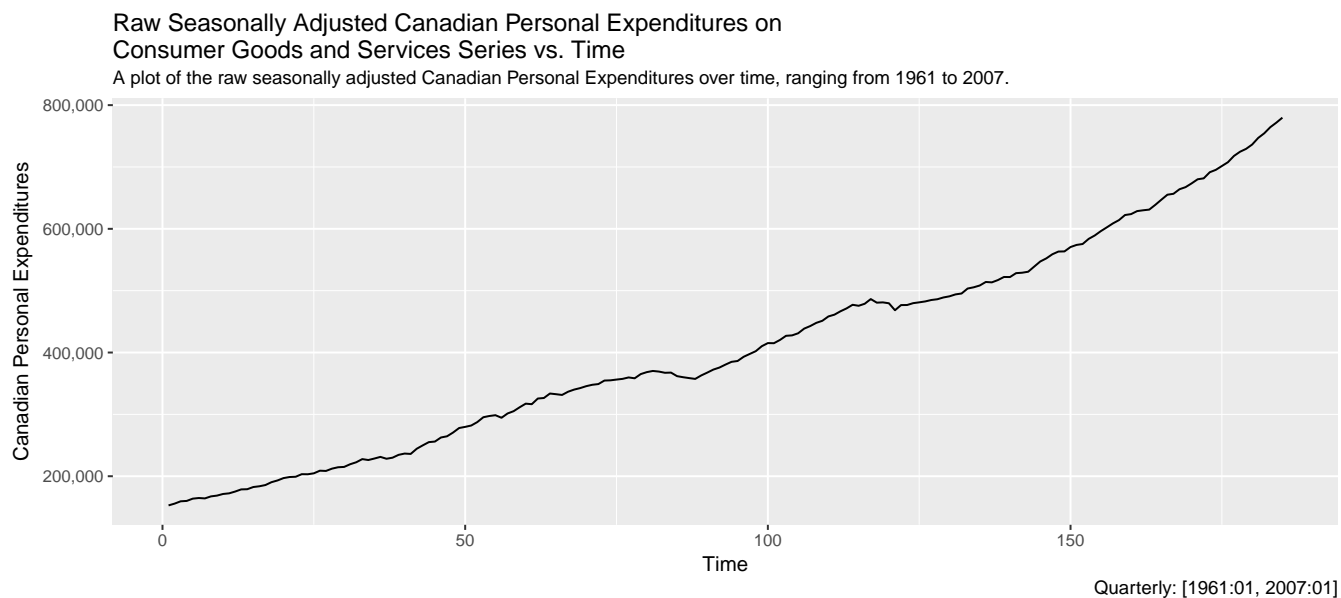
## Question 1

Let  $W_{1t}$  be the raw seasonally adjusted Canadian GDP series in the attached file “GDP\_CONS\_CANADA.” Let  $W_{2t}$  be the raw time series that you worked with in Assignment 1. Plot both raw time series.

### Plot of $W_{1t}$



### Plot of $W_{2t}$



## Question 2

Define  $X_t = \ln(W_{1t})$ . Describe your data series. Construct a measure of the business cycle  $Y_t$  using both the trend stationary (TS) and the difference stationary (DS) approaches. Plot  $Y_t$  for both cases.

$W_t$  is a raw time series that can be decomposed into  $W_t = T_t + S_t + Y_t$  where  $T_t$ —the trend—and  $S_t$ —the seasonal component—are non-stationary and  $Y_t$ —the cycle—is stationary. Our objective is to extract the cycle from  $W_t$  and try to model it using proper time series models. We will use the Log-Linear Law to define  $X_t = \ln(W_{1t})$  which states that  $X_t$  obeys linear laws. Additionally we can say that  $X_t \sim Normal$  by the Gaussian Law.

Notice that our data is seasonally adjusted, which means that in our case  $S_t = 0$  so we are just modeling  $X_t = \ln(W_{1t}) = T_t + Y_t$  which can be done using two different methods. From here we assume that  $W_t = W_0 e^{\mu t}$  and take two different approaches to our modeling of  $X_t$ . One we model, the trend stationary (TS) approach, we take  $T_t = \alpha + \mu t$  and the residuals of our models are  $Y_t$ , so  $X_t = \alpha + \mu t + Y_t(\text{residuals})$ . Our second approach, we take the differences of  $X_t$  to model our trend and get  $\Delta X_t = \mu + Y_t(\text{residuals})$ , the residuals of the difference stationary (DS) model are, again, our cycle  $Y_t$ .

### Trend Stationary Model

The results from our Trend Stationary model are:

```
## [1] "Model Results:"

##
## Call:
## lm(formula = X_1t ~ seq_along(W$X_1t), data = W)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.191639 -0.043878 -0.008175  0.059031  0.108496
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.261e+01  9.711e-03 1298.21  <2e-16 ***
## seq_along(W$X_1t) 8.263e-03  9.055e-05   91.26  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.06577 on 183 degrees of freedom
## Multiple R-squared:  0.9785, Adjusted R-squared:  0.9784
## F-statistic: 8328 on 1 and 183 DF, p-value: < 2.2e-16

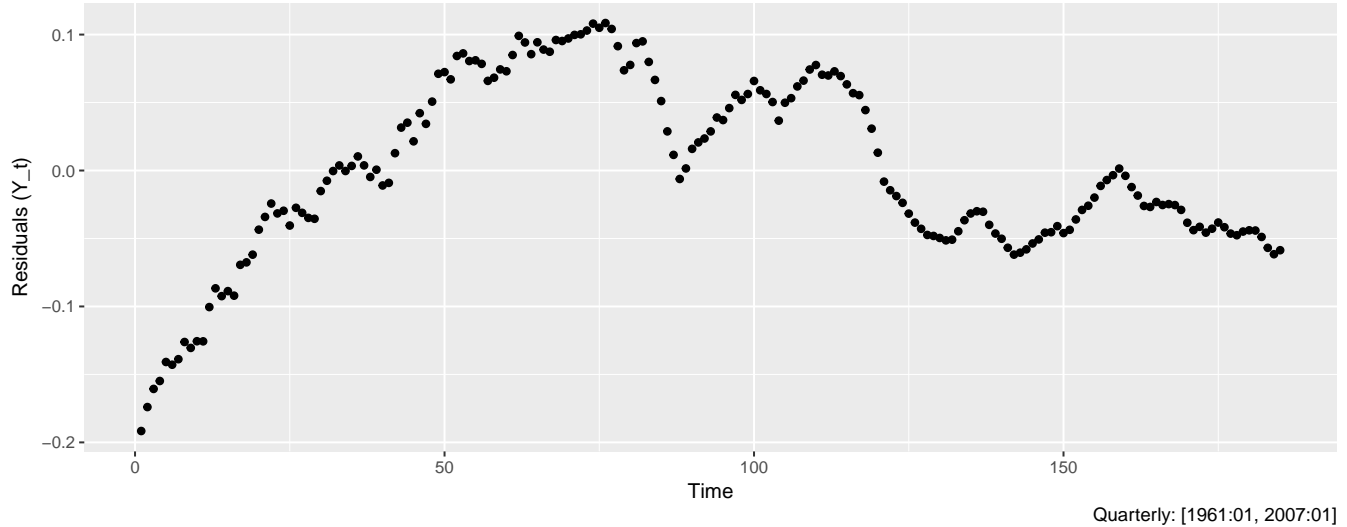
## [1] "RSS:"
## [1] 0.7916669
```

$$\hat{X}_t = \underset{(t)}{\hat{\alpha}} + \underset{(91.2)}{\hat{\mu}} t = 12.61 + 0.008263t$$

$n = 185$      $F - ratio = 8328$      $RSS = 0.7916669$      $R^2 = 0.9785$

Residuals vs. Time of our Trend Stationary (TS) Model of  $X_{1t}$

This plot shows the residuals over time, of our log-linear time series  $X_{1t}$ , using a Trend Stationary approach, data ranging from 1961 to 2007.



## Difference Stationary Model

The results from our Difference Stationary model are:

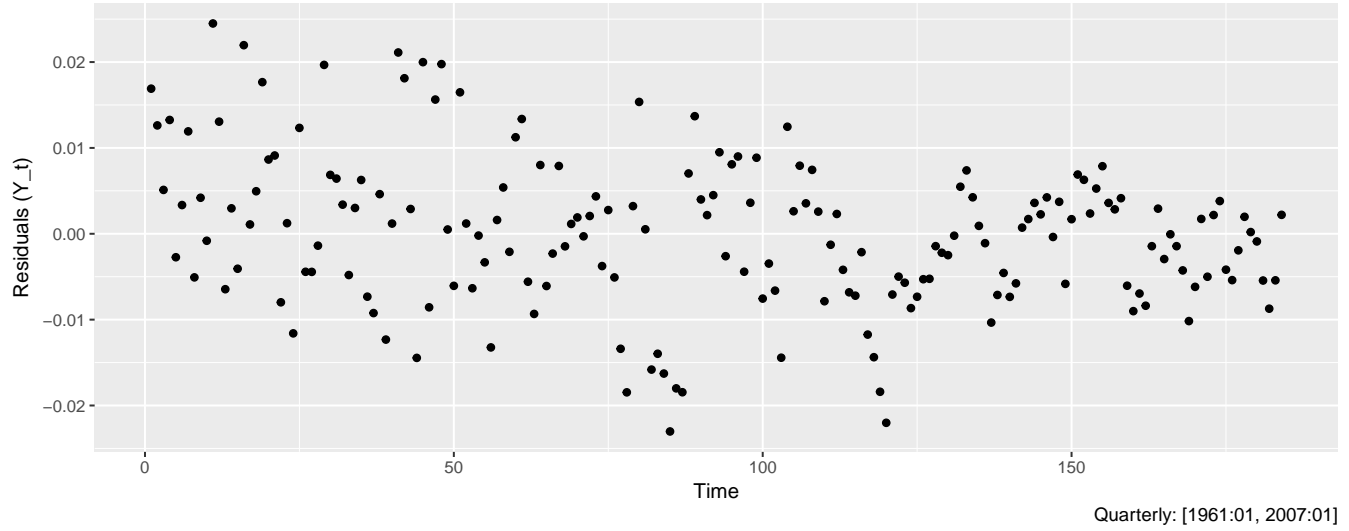
```
## [1] "Model Results:"
##
## Call:
## lm(formula = DeltaX_1t ~ 1, data = W)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.0230228 -0.0057192  0.0000702  0.0043923  0.0244942
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.0089863  0.0006494   13.84  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.00881 on 183 degrees of freedom
## (1 observation deleted due to missingness)
##
## [1] "RSS:"
##
## [1] 0.01420224
```

$$\Delta \hat{X}_t = \hat{\mu}_{13.84} = 0.0089863$$

$n = 184$      $t - value = 13.84$      $RSS = 0.01420224$      $R^2 = 0$

### Residuals vs. Time of our Difference Stationary (DS) Model of $X_{1t}$

This plot shows the residuals over time, of our log-linear time series  $X_{1t}$ , using a Difference Stationary approach, data ranging from 1961 to 2007.



### Question 3

Fit an  $AR(p)$  process to both the TS and DS  $Y_t$  using the Bayesian Information Criterion to decide on the appropriate value of  $p$ . If  $p < 2$  set  $p = 2$  for both the TS series or the DS series as required. (This is to force you to do some work,  $AR(1)$  or white noise is too easy!) From the estimated TS model for  $Y_t$ , calculate  $\gamma(0)^{1/2}$ , the infinite moving average weights  $\psi_k$ , and the autocorrelation functions  $\rho(k)$  for  $k = 0, 1, 2, \dots, 6$ .

#### Bayesian Information Criterion Table for both TS & DS Modeling Approaches

k	0.00	1.00	2.00	3.00	4.00	5.00	6.00	7.00	8.00	9.00
Yt_TS	-5.45	-9.43	-9.52	-9.50	-9.49	-9.46	-9.43	-9.41	-9.39	-9.37
Yt_DS	-9.47	-9.56	-9.54	-9.52	-9.49	-9.47	-9.44	-9.42	-9.40	-9.42
TS_BIC_Rank	10.00	5.00	1.00	2.00	3.00	4.00	6.00	7.00	8.00	9.00
DS_BIC_Rank	5.00	1.00	2.00	3.00	4.00	6.00	7.00	8.00	10.00	9.00

As you can see from the BIC Summary table above for the Trend Stationary model we have an  $AR(2)$  model that has the best fit. For the Difference Stationary model it actually showed that an  $AR(1)$  model is the best fit however an  $AR(2)$  was a close second and due to the requirements for the project we will assume going forward that an  $AR(2)$  was the best fit for the DS model aswell.

#### Summary of Selected Models

##### TS Model

```
##
## Call:
## arima(x = Y_t.TS.Residuals, order = c(2, 0, 0), include.mean = FALSE)
##
## Coefficients:
##          ar1          ar2
##      1.3332   -0.3376
## s.e.  0.0700    0.0704
##
## sigma^2 estimated as 6.908e-05:  log likelihood = 621.03,  aic = -1236.06
```

$$\hat{Y}_{t,TS} = \underset{(s.e.)}{\hat{\phi}_1} Y_{t-1} + \underset{(0.0700)}{\hat{\phi}_2} Y_{t-2} = 1.3332Y_{t-1} - 0.3376Y_{t-2}$$

##### DS Model

```
##
## Call:
## arima(x = Y_t.DS.Residuals, order = c(2, 0, 0), include.mean = FALSE)
##
```

```
## Coefficients:
##          ar1      ar2
##      0.3110  0.0582
## s.e.  0.0739  0.0742
##
## sigma^2 estimated as 6.865e-05:  log likelihood = 620.81,  aic = -1235.62
```

$$\hat{Y}_{t,DS} = \underset{(s.e.)}{\hat{\phi}_1} Y_{t-1} + \underset{(0.0742)}{\hat{\phi}_2} Y_{t-2} = 0.3110Y_{t-1} + 0.0582Y_{t-2}$$

The following three subsections only pertain to the TS Model:

### Calculation of $\gamma(0)^{1/2}$

```
##          [,1]
## [1,] 0.109349
```

We have found that  $\gamma(0) = 0.109349$ .

### Calculation of the Infinite Moving Average Weights

k	0	1.00	2.00	3.00	4.00	5.00	6.00
Psi_k	1	1.33	1.44	1.47	1.47	1.47	1.46

The  $\psi_k$ 's are calculated using the following formulas:

$$\begin{aligned}\psi_0 &= 1 \\ \psi_1 &= \phi_1\psi_0 \\ \psi_2 &= \phi_1\psi_1 + \phi_2\psi_0 \\ \psi_3 &= \phi_1\psi_2 + \phi_2\psi_1 \\ \psi_4 &= \phi_1\psi_3 + \phi_2\psi_2 \\ \psi_5 &= \phi_1\psi_4 + \phi_2\psi_3 \\ \psi_6 &= \phi_1\psi_5 + \phi_2\psi_4\end{aligned}$$

### Calculation of the Autocorrelations

k	0	1.000	2.000	3.000	4.000	5.000	6.000
Rho_k	1	0.997	0.991	0.985	0.979	0.972	0.966

Our  $\rho_k$ 's are taken directly from the `ARMAacf(estimated, lag.max= 6)` function output.

#### Question 4

Using both the DS and TS  $Y_t$  (with  $p \geq 2$ ), forecast the growth rate as  $E_T[\Delta X_{T+k}]$  for  $k = 0, 1, \dots, 8$ , where  $T$  is the last observation in your sample. Provide separate graphs of your TS and DS forecasts, and in the graphs plot 95% confidence intervals for  $\Delta X_{T+k}$ .

#### Calculation of the forecasts $E_T[\Delta X_{T+k}]$

We have an  $AR(2)$  model, so to calculate  $E_T[\Delta X_{T+k}]$  we used the following formula  $E_T[\Delta X_{T+k}] = \phi_1 Y_{T+k-1} + \phi_2 Y_{T+k-2}$ . Results:

k	0	1	2	3	4	5	6	7	8
E_X_185plusk_TS	-0.059	-0.057	-0.057	-0.056	-0.056	-0.055	-0.055	-0.055	-0.054
E_X_185plusk_DS	0.0022	0.0004	0.0002	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000

#### Calculation of the 95% Confidence Intervals for $E_T[\Delta X_{T+k}]$

To calculate the Confidence Intervals for the forecasts  $k = 0, \dots, 8$  we need to:

- 1) Calculate  $var(E_T[\Delta X_{T+k}]) = \hat{\sigma}^2 \cdot \sum_{j=1}^k \psi_j^2$  for each  $k$
- 2) Calculate the lower bound using  $E_T[\Delta X_{T+k}] - 2\sqrt{var(E_T[Y_{T+k}])}$
- 3) Calculate the upper bound using  $E_T[\Delta X_{T+k}] + 2\sqrt{var(E_T[Y_{T+k}])}$

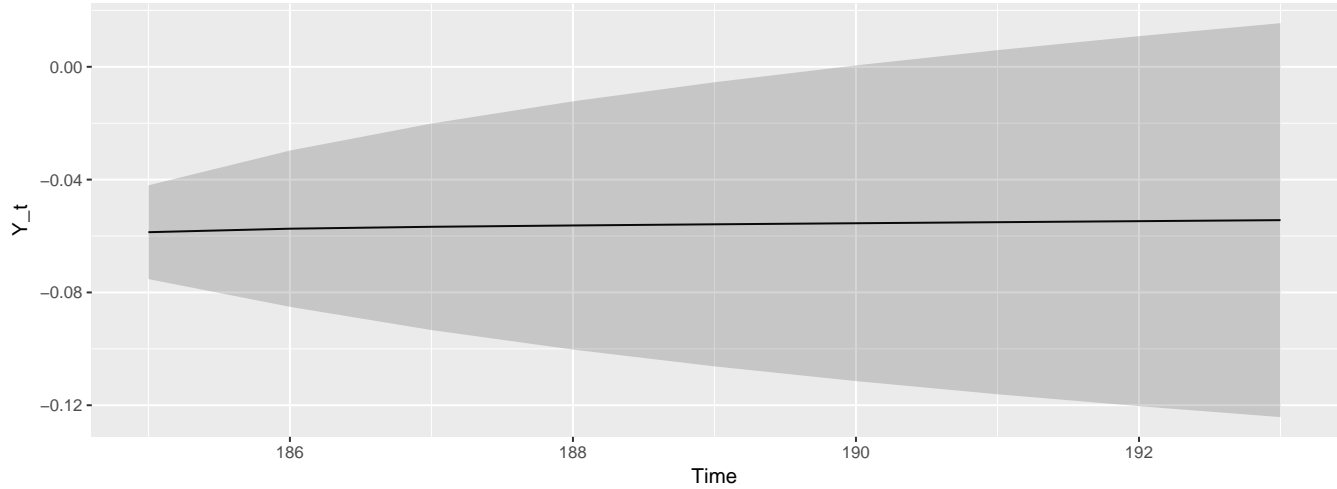
k	varEtk	EtkCIlowTS	Etk_TS	EtkCIhighTS	EtkCIlowDS	Etk_DS	EtkCIhighDS
0	0.0000691	-0.0752672	-0.0586441	-0.0420209	-0.0144166	0.0022066	0.0188298
1	0.0001919	-0.0851030	-0.0573989	-0.0296947	-0.0273333	0.0003709	0.0280750
2	0.0003351	-0.0933401	-0.0567278	-0.0201154	-0.0363687	0.0002437	0.0368560
3	0.0004843	-0.1002683	-0.0562534	-0.0122385	-0.0439176	0.0000974	0.0441123
4	0.0006343	-0.1062171	-0.0558475	-0.0054778	-0.0503252	0.0000444	0.0504141
5	0.0007832	-0.1114366	-0.0554665	0.0005036	-0.0559506	0.0000195	0.0559896
6	0.0009304	-0.1161004	-0.0550956	0.0059093	-0.0609962	0.0000086	0.0610135
7	0.0010758	-0.1203283	-0.0547297	0.0108689	-0.0655948	0.0000038	0.0656024
8	0.0012193	-0.1242043	-0.0543671	0.0154701	-0.0698355	0.0000017	0.0698389



# Forecast Visualizations

## Trend Stationary Forecasts

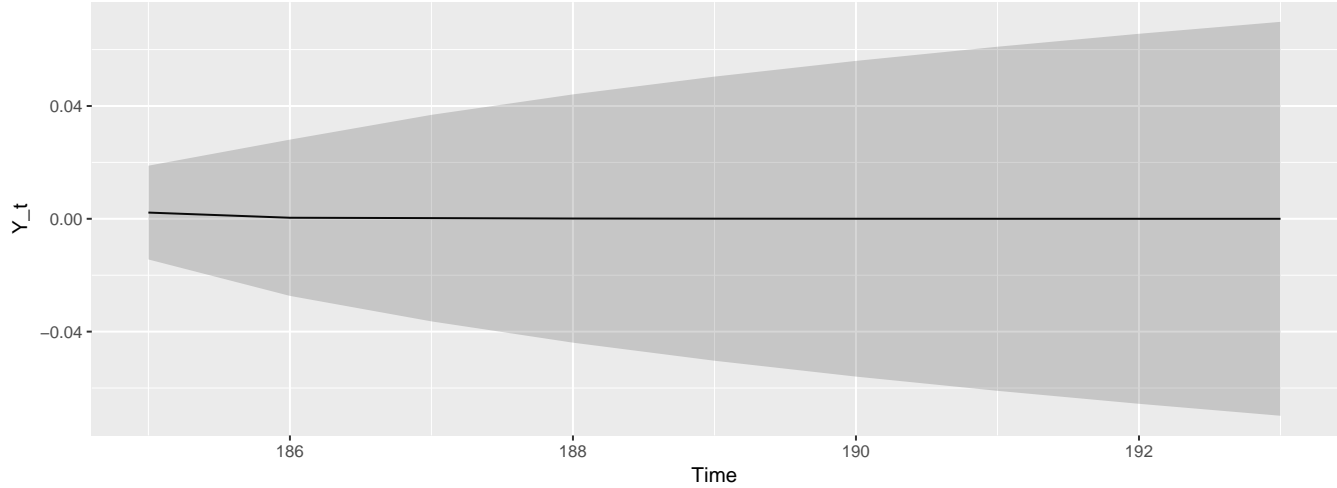
This plot shows the next 8 forecasts, of our Trend Stationary Residuals  $Y_t$ ; the shaded region indicates the area bounded by the 95% confidence interval of our forecasts.



Quarterly: [2007:01, 2009:01]

## Difference Stationary Forecasts

This plot shows the next 8 forecasts, of our Difference Stationary Residuals  $Y_t$ ; the shaded region indicates the area bounded by the 95% confidence interval of our forecasts.



Quarterly: [2007:01, 2009:01]

## Question 5

Do an augmented Dickey-Fuller test on  $X_t$ , augmenting the regression with 5 lags (or if you wish use  $AIC(k)$  or  $BIC(k)$ ). Based on the p-value, decide if  $X_t$  is Difference or Trend stationary.

### Augmented Dickey-Fuller (ADF) Test

The objective of the Augmented Dickey-Fuller test is to help us decide on whether the model is Trend Stationary or Difference Stationary. Essentially the way the test functions is to test the null hypothesis that your data has a unit root. If it does have a unit root this indicates that you should use the Difference Stationary approach otherwise, use the Trend Stationary.

We are augmenting the regression on  $X_t$  with 5 lags so our model will look like:

$$\Delta X_t = \alpha + \mu t + \phi X_{t-1} + \sum_{i=1}^5 \theta_i \Delta X_{t-i} + a_t$$

$H_0 : \phi = 0$  or  $H_0 : X_t$  is DS.  $H_1 : \phi \neq 0$  or  $H_1 : X_t$  is TS.

Our test statistic is defined by:

$$t_{stat} = \frac{\hat{\phi} - 0}{SE(\hat{\phi})} \sim t(n - 8)$$

A built in function in R does this all for us so our calculated  $t_{stat}$  actually comes from the output:

$$t_{stat} = -2.6929, \quad p - value = 0.2865$$

Our decision criteria whether or not to reject  $H_0$  is given by our p-value. Due to the fact that our p-value is much greater than 0.05 we do not reject the null hypothesis that  $\phi = 0$  therefore we believe  $X_t$  should use the Difference Stationary modeling approach.

## Question 6

Use Box-Jenkins identification to identify and estimate an  $ARMA(p, q)$  model for both the DS and TS  $Y_t$ . Plot the standardized residuals  $z_t = \frac{\epsilon_t}{\sigma}$  for both models and comment about whether the normal distribution appears to be appropriate. Perform the following diagnostic tests on both  $ARMA(p, q)$  models: 1) Box-Pierce, 2) Overfitting with  $r = 4$ , 3) Jarque-Bera, 4) a test for  $ARCH(6)$ .

### Box-Jenkins Identification

The way Box-Jenkins Identification works is by looking at the patterns in both our estimated autocorrelation function  $\hat{\rho}(k)$  and our estimated partial autocorrelation function  $\hat{\phi}_{kk}$ . We can identify the most likely model by the following guide:

Model	Rho_k	Phi_kk
AR(p)	damped exponential	cut-off at k=p
MA(q)	cut-off at k=q	damped exponential
ARMA(p,q)	damped exponential	damped exponential

### Calculation of $\hat{\rho}(k)$ & $\hat{\phi}_{kk}$

In order to calculate our  $\hat{\rho}(k) = \frac{\hat{\gamma}(k)}{\hat{\gamma}(0)}$  we need to first calculate  $\hat{\gamma}(0) = \frac{1}{n} \sum Y_t^2$  and secondly  $\hat{\gamma}(k) = \frac{1}{n} \sum Y_t Y_{t+k}$

The issue we now face though is the problem with identifying the cut-off. Randomness in the data will make  $\hat{\phi}_{kk}$  and  $\hat{\rho}(k)$  non-zero. We will need to test for the cut-off! In order to do this we will need the Standard Error to decide if  $\hat{\phi}_{kk}$  and  $\hat{\rho}(k)$  are significantly different than zero.

From what we went over in class the approximate distribution of both  $\hat{\rho}(k), \hat{\phi}_{kk} \sim N[0, 1/n]$ . Therefore  $SE(\hat{\rho}(k)) = \frac{1}{\sqrt{n}}$  and our decision criteria for our cutoffs are:

Accept  $H_0$ , i.e.,  $\rho(k) = 0$  if:

$$|\hat{\rho}(k)| < 2 \cdot SE(\hat{\rho}(k)) = 2 \cdot \frac{1}{\sqrt{n}} = \frac{2}{\sqrt{185}} = 0.1470429$$

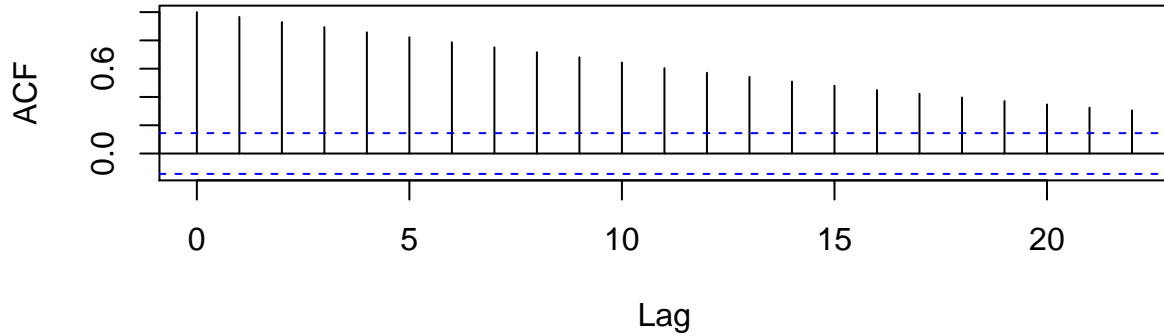
And reject  $H_0$ , i.e.,  $\rho(k) \neq 0$  if:

$$|\hat{\rho}(k)| \geq 2 \cdot SE(\hat{\rho}(k)) = 2 \cdot \frac{1}{\sqrt{n}} = \frac{2}{\sqrt{185}} = 0.1470429$$

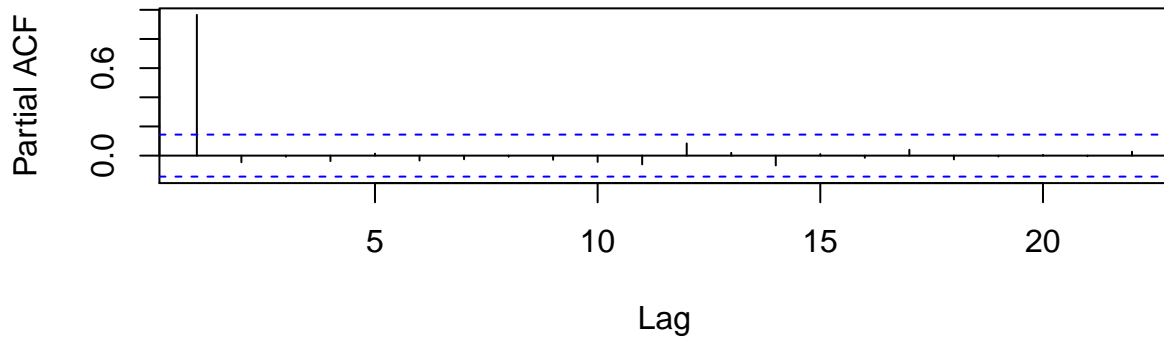
The same cut-off applies for  $\hat{\phi}_{kk}$ .

## Trend Stationary

**Series Y\_t.TS.Residuals**



**Series Y\_t.TS.Residuals**



k	0	1	2	3	4	5	6	7	8	9	10
gamma_k	0.004	0.004	0.004	0.004	0.004	0.004	0.003	0.003	0.003	0.003	0.003
rho_k	1.000	0.971	0.939	0.909	0.876	0.845	0.813	0.781	0.748	0.715	0.680
phi_kk	0.000	0.997	-0.338	-0.065	-0.132	0.002	0.027	-0.072	-0.080	0.087	-0.218

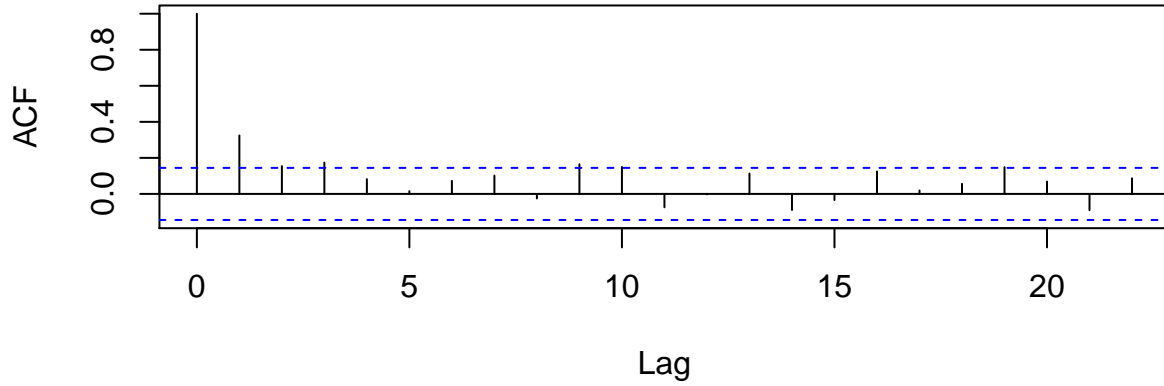
Using the  $2 \cdot SE(\hat{\rho}(k)) = 0.1470429$  it appears very clearly from the TS approach that our model has a damped exponential for  $\rho(k)$  and a cut off at  $k = 3$ . This suggests to us that we should have an  $AR(2)$  model.

$$\hat{Y}_{t,TS} = \underset{(s.e.)}{\hat{\phi}_1} Y_{t-1} + \underset{(0.0704)}{\hat{\phi}_2} Y_{t-2} = 1.3332Y_{t-1} - 0.3376Y_{t-2}$$

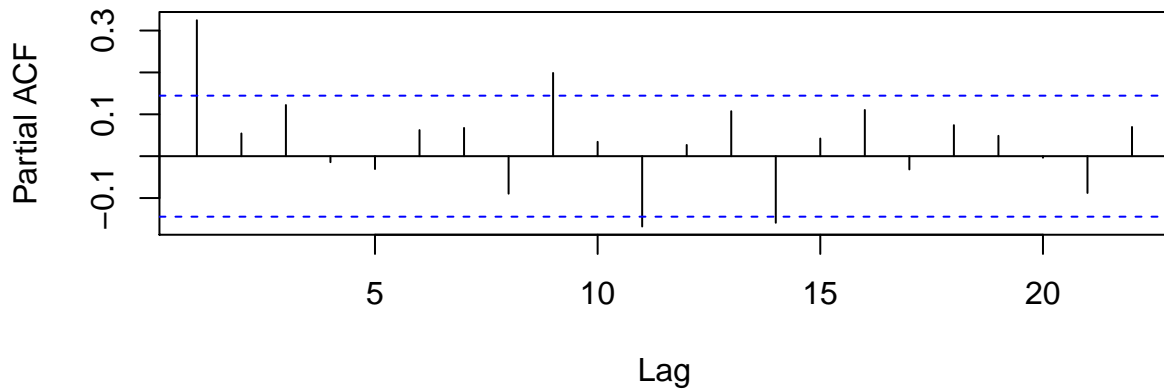
This result is consistent with what our  $BIC(k)$  method of model evaluation gave us.

## Difference Stationary

**Series Y\_t.DS.Residuals**



**Series Y\_t.DS.Residuals**



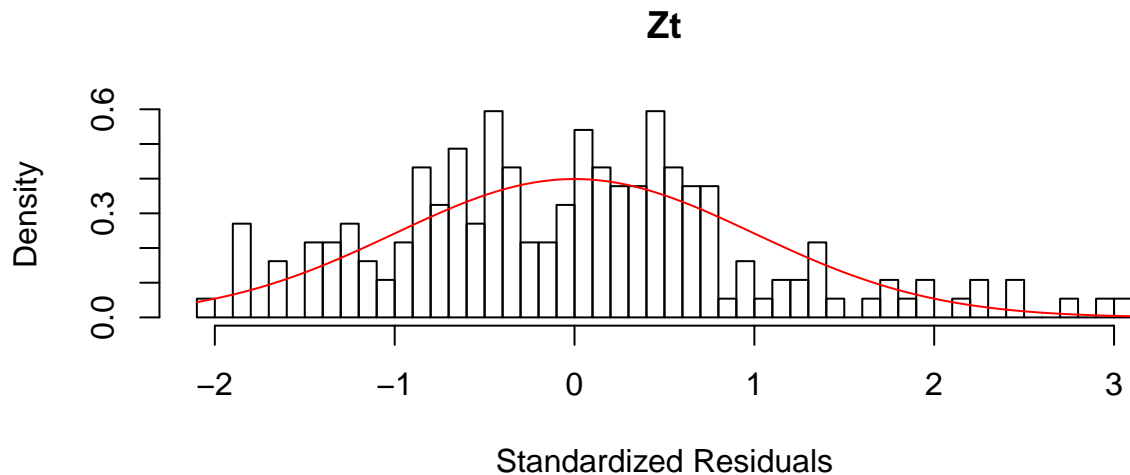
k	0	1	2	3	4	5	6	7	8	9	10
gamma_k	8e-05	3e-05	1e-05	1e-05	1e-05	0e+00	1e-05	1e-05	0e+00	1e-05	1e-05
rho_k	1.000	0.326	0.156	0.177	0.084	0.016	0.075	0.107	-0.026	0.173	0.158
phi_kk	0.000	0.330	0.058	0.125	-0.008	-0.034	0.065	0.073	-0.095	0.209	0.028

The results from the Difference Stationary data is much less obvious than the TS. To make a guess at these results I would say that  $\rho_k$  suggests a damped exponential and  $\phi_k k$  has a cut off at  $k = 2$  ignoring the large  $\phi_k k > 8$  that occur. The resulting model would then be:

$$\hat{Y}_{t,DS} = \underset{(s.e.)}{\hat{\phi}_1} Y_{t-1} = 0.3295 Y_{t-1}$$

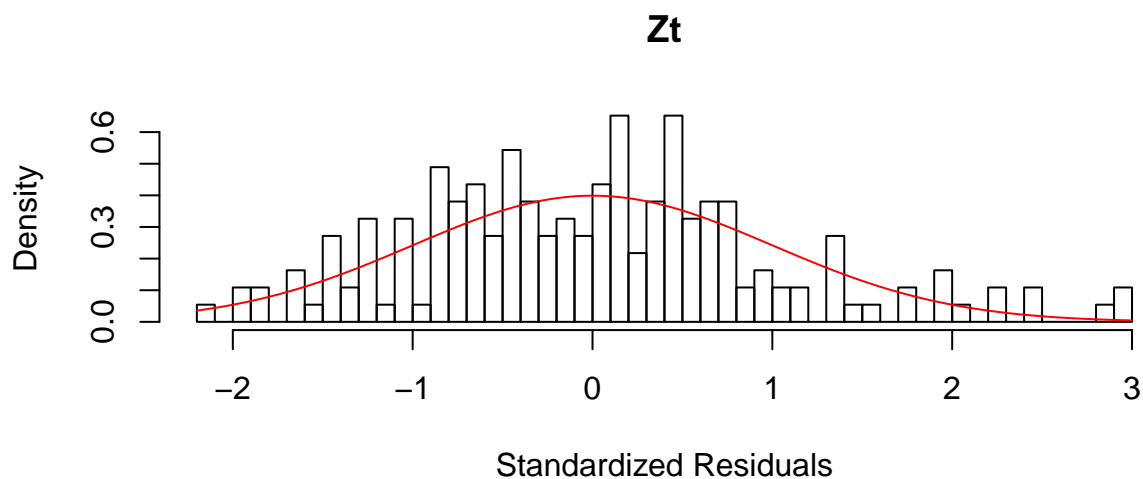
Plot the standardized residuals

Trend Stationary Residuals



This does not appear to fit the normal distribution fantastically, this could be because a lot of noise in our data. There does appear to be a bit of bimodality on either side of 0, and a bit of right skewness.

Difference Stationary Residuals



This does not appear to fit the normal distribution fantastically, this could be because a lot of noise in our data. There does appear to be a bit of right-skewness to this fit.

## 1) Box-Pierce

The Box-Pierce Test is to test whether or not there exists autocorrelation between  $\epsilon_t$  and  $\epsilon_{t-k}$  for any  $k = 1, \dots, M$  where  $M = \text{round}_{up}(\text{sqrt}(N))$ . This is a test for independence between our residual lags.

### Trend Stationary

$$H_0 : \rho_\epsilon(k) = 0 \text{ and } H_1 : \rho_\epsilon(k) \neq 0$$

```
##  
## Box-Pierce test  
##  
## data: residualsTS  
## X-squared = 32.53, df = 14, p-value = 0.003365
```

Using the built in Box-Pierce test in R we have found a p-value of 0.003365. This means we reject the null hypothesis. Our test suggests that our  $\epsilon_{t+k}$ 's are not all independent of each other. Our model is probably not a great fit.

### Difference Stationary

$$H_0 : \rho_\epsilon(k) = 0 \text{ and } H_1 : \rho_\epsilon(k) \neq 0$$

```
##  
## Box-Pierce test  
##  
## data: residualsDS  
## X-squared = 31.255, df = 14, p-value = 0.005105
```

Using the built in Box-Pierce test in R we have found a p-value of 0.005105. This means we reject the null hypothesis. Our test suggests that our  $\epsilon_{t+k}$ 's are not all independent of each other. Our model is probably not a great fit.

## 2) Overfitting with $r = 4$

The TS model suggested an  $AR(2)$  and the DS model suggested an  $AR(1)$  so we want to do a comparison using the Likelihood Ratio Test for an  $AR(2)$  versus  $AR(4)$  and an  $AR(1)$  versus  $AR(4)$ .

### Trend Stationary

Therefore  $H_0 : AR(2)$  or  $H_0 : \phi_3 = \phi_4 = 0$  and  $H_1 : AR(4)$

```
## [1] "chi-statistic"  
## [1] 3.913271  
## [1] "chi-critical value"  
## [1] 5.991465
```

For our Trend Stationary model we do not reject the Null Hypothesis that  $H_0 : \phi_3 = \phi_4 = 0$ , this suggests for our TS model we have properly the data.

### Difference Stationary

Therefore  $H_0 : AR(1)$  or  $H_0 : \phi_2 = \phi_3 = \phi_4 = 0$  and  $H_1 : AR(4)$

```
## [1] "chi-statistic"
## [1] 4.54888
## [1] "chi-critical value"
## [1] 7.814728
```

For our Difference Stationary model we do not reject the Null Hypothesis that  $H_0 : \phi_2 = \phi_3 = \phi_4 = 0$ , this suggests that an  $AR(4)$  with our DS model would be overfitting the data.

### 3) Jarque-Bera

Our Jarque-Bera statistic is defined by  $J_{stat} = n(\frac{\hat{k}_3^2}{6} + \frac{(\hat{k}_4-3)^2}{24}) \sim \chi_{0.95}^2(2)$

The Jarque-Bera statistic tests if our standardized residuals  $z_t$  are in fact Normally Distributed. Our null hypothesis  $H_0$  : our residuals are normally distributed.

### Trend Stationary

$H_0 : k_3 = 0$  and  $k_4 = 3$  and  $H_1 : k_3 \neq 0$  or  $k_4 \neq 3$

```
## [1] "K_3"
## [1] 0.5291296
## [1] "K_4"
## [1] 3.348778
## [1] "Jarque-Bera Statistic:"
## [1] 9.570347
## [1] "Jarque-Bera Critical Value:"
## [1] 5.991465
```

Our Jarque-Bera Statistic (9.570347) does exceed our Jarque-Bera Critical Value (5.991465) this means that we reject the null hypothesis that our Trend Stationary model does not exhibit skewness or that it does not have excess kurtosis. This suggests to us that our tests have not detected normality in our residuals, implying that our model probably isn't a good fit.



## Difference Stationary

```
## [1] "K_3"
## [1] 0.5265769
## [1] "K_4"
## [1] 3.387708
## [1] "Jarque-Bera Statistic:"
## [1] 9.655788
## [1] "Jarque-Bera Critical Value:"
## [1] 5.991465
```

Our Jarque-Bera Statistic (9.655788) exceeds our Jarque-Bera Critical Value (5.991465) this means that we reject the null hypothesis that our Difference Stationary model does not exhibit skewness or that it does not have excess kurtosis. This suggests to us that our tests have detected a deviation from normality in our residuals, implying that our model is not a good fit. We could already tell this from the skewness in our plot in the previous section.

## 4) a test for ARCH(6).

Autoregressive Conditional Heteroskedasticity tests are to determine if there exists non-linear correlation between our  $\epsilon_{t+k}$ 's. Our previous Box-Pierce test already determined that there existed linear correlation between our  $\epsilon_{t+k}$ 's.

## Trend Stationary

For our ARCH(6) TS test we are testing the null hypothesis that for our model:

$$\hat{Y}_{t,TS} = \underset{(s.e.)}{\hat{\phi}_1} Y_{t-1} + \underset{(0.0700)}{\hat{\phi}_2} Y_{t-2} = 1.3332Y_{t-1} - 0.3376Y_{t-2} + a_t$$

where

$$a_t = z_t(\sigma^2 + \alpha_1 a_{t-1}^2 + \dots + \alpha_6 a_{t-6}^2)^{1/2}$$

Our  $H_0 : \alpha_1 = \alpha_2 = \alpha_3 = \alpha_4 = \alpha_5 = \alpha_6 = 0$

```
## [1] "ARCH(6) Statistic:"
## [1] 67.66438
## [1] "ARCH(6) Critical Value"
## [1] 12.59159
## [1] "Statistic > Critical Value?"
## [1] TRUE
```

As it turns out our ARCH Test Statistic is greater than our Critical value, so we reject the null hypothesis that  $H_0 : \alpha_1 = \alpha_2 = \alpha_3 = \alpha_4 = \alpha_5 = \alpha_6 = 0$  for our Trend Stationary model. Our tests suggest that there exists non-linear correlation between our residuals for our TS model.

## Difference Stationary

For our ARCH(6) DS test we are testing the null hypothesis that for our model:

$$\hat{Y}_{t,DS} = \underset{(s.e.)}{\hat{\phi}_1} Y_{t-1} = 0.3295Y_{t-1} + a_t$$

where

$$a_t = z_t(\sigma^2 + \alpha_1 a_{t-1}^2 + \dots + \alpha_6 a_{t-6}^2)^{1/2}$$

Our  $H_0 : \alpha_1 = \alpha_2 = \alpha_3 = \alpha_4 = \alpha_5 = \alpha_6 = 0$

```
## [1] "ARCH(6) Statistic:"
## [1] 67.68087
## [1] "ARCH(6) Critical Value"
## [1] 12.59159
## [1] "Statistic > Critical Value?"
## [1] TRUE
```

As it turns out our ARCH Test Statistic is greater than our Critical value, so we reject the null hypothesis that  $H_0 : \alpha_1 = \alpha_2 = \alpha_3 = \alpha_4 = \alpha_5 = \alpha_6 = 0$  for our Difference Stationary model. Our tests suggest that there exists non-linear correlation between our residuals for our DS model.

## Question 7

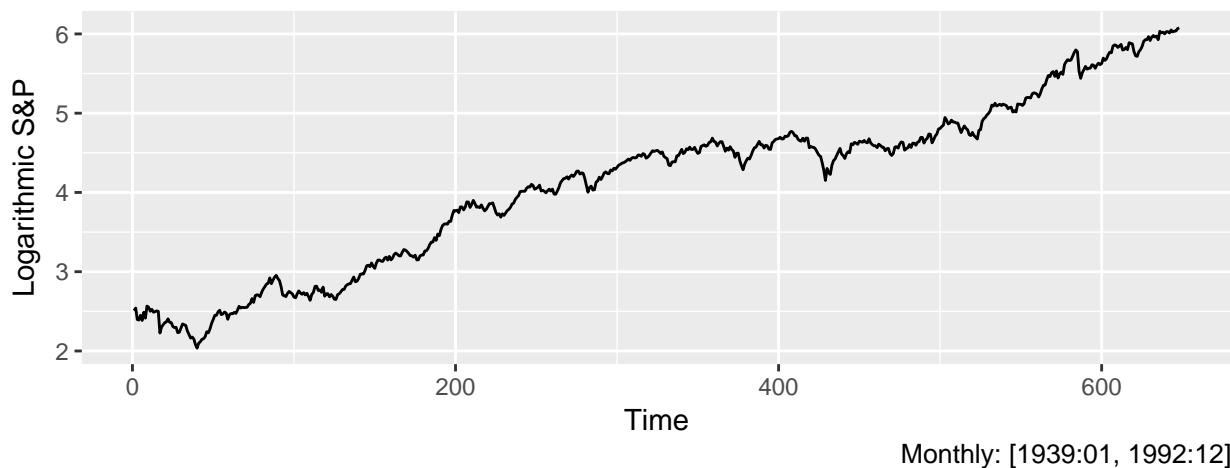
Find a monthly (or higher frequency) financial time series  $P_t$  that is either an exchange rate or a stock index. You can use the data file S&P\_Data on Learn. Using the estimated autocorrelation function  $\hat{\rho}_\alpha(k)$  for  $a_t$ , determine if this series follows a random walk as

$$\ln(P_t) = \delta + \ln(P_{t-1}) + a_t \text{ with } a_t \sim i.i.N[0, \sigma^2]$$

Test whether  $a_t$  is normally distributed. Estimate the autocorrelation function for  $a_t^2$ . Estimate a  $GARCH(1,1)$  model for  $a_t$

### Logarithmic S&P Series vs. Time

A plot of the logarithmic S&P over time, ranging from 1939 to 1992.



```
##
## Call:
## lm(formula = lnPt[-1] ~ lnPt[-length(lnPt)])
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.279323 -0.023709  0.002527  0.029122  0.146871
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.005455   0.006955   0.784    0.433
## lnPt[-length(lnPt)] 1.000014   0.001638 610.622 <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.04363 on 645 degrees of freedom
## Multiple R-squared:  0.9983, Adjusted R-squared:  0.9983
## F-statistic: 3.729e+05 on 1 and 645 DF, p-value: < 2.2e-16
```

$$\ln(P_t) = \underset{(0.784)}{\delta} + \underset{(610.622)}{\beta} \cdot \ln(P_{t-1}) + a_t = 0.0055 + 1.00\ln(P_{t-1}) + a_t$$

### Estimate of the autocorrelation function $\hat{\rho}_a(k)$

The estimated autocorrelation function  $\hat{\rho}_a(k)$  appears to follow a random walk as all  $\hat{\rho}_a(k) \approx 0$  for  $k = 1, \dots, 10$ :

```
##
## Autocorrelations of series 'at', by lag
##
##      1      2      3      4      5      6      7      8      9     10
## -0.012 -0.014  0.004  0.021  0.082 -0.056 -0.010 -0.059  0.015 -0.043
```

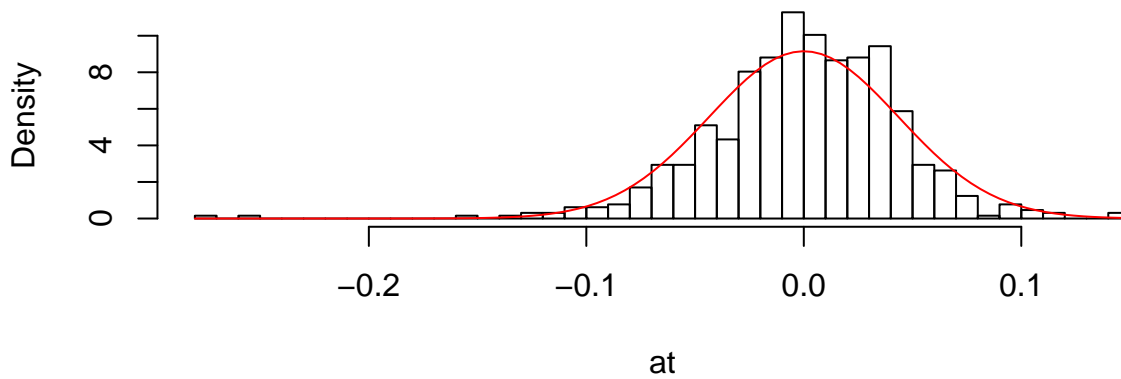
### Test whether $a_t$ is normally distributed

From the plot, the data appears to be normally distributed with a very low variance.

To test for Normality I used the Kolmogorov-Smirnov Test:

$H_0 : a_t \sim N[0, \sigma^2 = 0.0019]$  versus  $H_0 : a_t \text{ not } \sim N[0, \sigma^2 = 0.0019]$

### Histogram of at to test for Normality



Two-sample Kolmogorov-Smirnov test

data: at and x D = 0.043277, p-value = 0.5797 alternative hypothesis: two-sided

Our results show a very high p-value, so we can conclude that  $a_t$  is Normally distributed.

### Estimate the autocorrelation function for $a_t^2$

```
##
## Autocorrelations of series 'at^2', by lag
##
##      1      2      3      4      5      6      7      8      9     10
##  0.075  0.023  0.014  0.015 -0.009  0.046 -0.046  0.110  0.073  0.054
```

**Estimate a  $GARCH(1,1)$  model for  $a_t$**

	Estimate	Std. Error	t value	Pr(> t )
mu	-1.190e-17	1.618e-03	0.000	1.00000
omega	9.487e-05	3.926e-05	2.416	0.01567 *
alpha1	6.072e-02	2.026e-02	2.996	0.00273 **
beta1	8.884e-01	2.867e-02	30.988	< 2e-16 ***

$$\underset{(s.e.)}{\sigma_t^2} = \underset{(0.000039)}{0.000095} + \underset{(0.02026)}{0.06072}\epsilon_{t-1}^2 + \underset{(0.02867)}{0.8884}\sigma_{t-1}^2$$

$GARCH(1,1)$  appears to be a good fit for our variance.

## Question 8

Find another type of model for  $\ln(P_t)$  not covered in class that displays nonlinearity or conditional heteroskedasticity. Estimate the model, report your results, and run relevant diagnostics. Comment on the reliability of your results.

### SETAR Model

We will now attempt to fit a Threshold AutoRegressive Model (TAR), Tong(1990).

```
##
## Non linear autoregressive model
##
## SETAR model ( 2 regimes)
## Coefficients:
## Low regime:
##      const.L      phiL.1      phiL.2
## -0.001013748  0.021828793 -0.067568279
##
## High regime:
##      const.H      phiH.1      phiH.2
## -0.005699381 -0.070343330  0.150716911
##
## Threshold:
## -Variable: Z(t) = + (0) X(t)+ (1)X(t-1)
## -Value: 0.002527 (fixed)
## Proportion of points in low regime: 50.23%    High regime: 49.77%
##
## Residuals:
##      Min      1Q    Median      3Q      Max
## -0.2750116 -0.0230125  0.0024857  0.0291983  0.1426712
##
## Fit:
## residuals variance = 0.001847,  AIC = -4060, MAPE = 117.3%
##
## Coefficient(s):
##
##      Estimate Std. Error t value Pr(>|t|)
## const.L -0.0010137  0.0033589  -0.3018  0.7629
## phiL.1  0.0218288  0.0472565   0.4619  0.6443
## phiL.2 -0.0675683  0.0723151  -0.9344  0.3505
## const.H -0.0056994  0.0040816  -1.3964  0.1631
## phiH.1 -0.0703433  0.0690625  -1.0185  0.3088
## phiH.2  0.1507169  0.1006466   1.4975  0.1348
##
## Threshold
## Variable: Z(t) = + (0) X(t) + (1) X(t-1)
```

##

## Value: 0.002527 (fixed)

The Threshold was chosen to be the median of  $a_t$

## Appendix (Code)

```
### Question 1
library(xlsx)
library(dplyr)
library(ggplot2)
library(scales)
W <- read.xlsx("GDP_CONS_CANADA.xls", sheetName = "Sheet3")
W <- W[complete.cases(W),]

#### Plot of  $W_{1t}$ 
qplot(seq_along(W$GDP), W$GDP, geom = "line") +
  labs(x= "Time",
       y = "Canadian GDP",
       title="Raw Seasonally Adjusted Canadian GDP Series vs. Time",
       subtitle = "A plot of the raw seasonally adjusted Canadian Gross Domestic Product over time",
       caption = "Quarterly: [1961:01, 2007:01]") +
  scale_y_continuous(labels = comma)

#### Plot of  $W_{2t}$ 
qplot(seq_along(W$CONS), W$CONS, geom = "line") +
  labs(x= "Time",
       y = "Canadian Personal Expenditures",
       title="Raw Seasonally Adjusted Canadian Personal Expenditures on \nConsumer Goods and Services",
       subtitle = "A plot of the raw seasonally adjusted Canadian Personal Expenditures over time",
       caption = "Quarterly: [1961:01, 2007:01]") +
  scale_y_continuous(labels = comma)

### Question 2

W <- W %>% mutate(X_1t = log(GDP.))

#### Trend Stationary Model

model_TS <- lm(X_1t ~ seq_along(W$X_1t), data = W)
print("Model Results:")
summary(model_TS)
print("RSS:")
sum(resid(model_TS)^2)

qplot(seq_along(model_TS$residuals), model_TS$residuals) +
  labs(x= "Time",
       y = "Residuals (Y_t)",
       title="Residuals vs. Time of our Trend Stationary (TS) Model of X_1t",
       subtitle = "This plot shows the residuals over time, of our log-linear time series X_1t")
```



```

        caption = "Quarterly: [1961:01, 2007:01]") +
scale_y_continuous(labels = comma)

#### Difference Stationary Model
W$DeltaX_1t <- W$X_1t - lag(W$X_1t)
model_DS <- lm(DeltaX_1t ~ 1, data = W)
print("Model Results:")
summary(model_DS)
print("RSS:")
sum(resid(model_DS)^2)

qplot(seq_along(model_DS$residuals), model_DS$residuals) +
  labs(x= "Time",
       y = "Residuals (Y_t)",
       title="Residuals vs. Time of our Difference Stationary (DS) Model of X_1t",
       subtitle = "This plot shows the residuals over time, of our log-linear time series X",
       caption = "Quarterly: [1961:01, 2007:01]") +
  scale_y_continuous(labels = comma)

### Question 3

#### Bayesian Information Criterion Table for both TS & DS Modeling Approaches
library(knitr)
library(dplyr)
# Defining the residuals datasets
Y_t.TS.Residuals <- residuals(model_TS)
Y_t.DS.Residuals <- residuals(model_DS)

## Defining the BIC function rather than using R's.
BIC <- function(res, k, N){
  bic <- log(sum(res^2) / N)
  bic <- bic + k * log(N)/N
  bic
}

#obtaining the bic
bic.data.frame = data.frame(k = 0:9, Yt_TS = rep(NA, 10), Yt_DS = rep(NA, 10))
N = length(Y_t.TS.Residuals)
for(ii in 0:9) {
  TS.model.arima = arima(Y_t.TS.Residuals, order = c(ii, 0, 0), include.mean=FALSE) # TS_A
  DS.model.arima = arima(Y_t.DS.Residuals, order = c(ii, 0, 0), include.mean=FALSE) # DS_A
  res.arima_TS = TS.model.arima$residuals
  res.arima_DS = DS.model.arima$residuals
  bic.data.frame[ii + 1, "Yt_TS"] = BIC(res.arima_TS, ii, N)
  bic.data.frame[ii + 1, "Yt_DS"] = BIC(res.arima_DS, ii, N)
}

```

```

bic.data.frame <- mutate(bic.data.frame, TS_BIC_Rank = min_rank(Yt_TS), DS_BIC_Rank = min_rank(DS_TS),
kable(t(round(bic.data.frame, 2)))

#### Summary of Selected Models

##### TS Model
model.arima.TS = arima(Y_t.TS.Residuals, order = c(2, 0, 0), include.mean=FALSE)
model.arima.TS

##### DS Model
model.arima.DS = arima(Y_t.DS.Residuals, order = c(2, 0, 0), include.mean=FALSE)
model.arima.DS

##### Calculation of  $\gamma(0)^{1/2}$ 
Phi <- model.arima.TS$coef # Estimates from AR(X) model. Find X with AIC/BIC
rhos <- ARMAacf(Phi, lag.max= 6)
sigma.hat.squared <- sum(model.arima.TS$residuals^2)/N
# Extracting the theoretical acf values for k=0,...,6
gamma0 <- (sigma.hat.squared / (1 - Phi %*% rhos[2:3])) ### gamma0
sqrt(gamma0)

##### Calculation of the Infinite Moving Average Weights
# calculating psi's, the infinite MA weights
psis <- rep(NA, 9)
psis[1] <- 1
psis[2] <- Phi[1]
psis[3] <- Phi[1] * psis[2] + Phi[2]
psis[4] <- Phi[1] * psis[3] + Phi[2] * psis[2]

for( j in 5:9) {
  psis[j] <- psis[(j-1):-1:(j-2)] %*% Phi
}
kable(t(round(data.frame(k = 0:6, Psi_k = psis[1:7]), 2)))

##### Calculation of the Autocorrelations

### Question 4
#### Calculation of the forecasts  $E_T[\Delta X_{T+k}]$ 

# Calculation for the Trend Stationary Forecasts
Etk <- rep(NA, 8)
phi <- coef(model.arima.TS)
Etk[1] <- Y_t.TS.Residuals[length(Y_t.TS.Residuals)]
Etk[2] <- phi[1] * Etk[1] + phi[2] * Y_t.TS.Residuals[length(Y_t.TS.Residuals)-1]
Etk[3] <- phi[1] * Etk[2] + phi[2] * Etk[1]
Etk[4] <- phi[1] * Etk[3] + phi[2] * Etk[2]

```

```

Etk[5] <- phi[1] * Etk[4] + phi[2] * Etk[3]
Etk[6] <- phi[1] * Etk[5] + phi[2] * Etk[4]
Etk[7] <- phi[1] * Etk[6] + phi[2] * Etk[5]
Etk[8] <- phi[1] * Etk[7] + phi[2] * Etk[6]
Etk[9] <- phi[1] * Etk[8] + phi[2] * Etk[7]

Etk_TS <- Etk

# Calculation for the Difference Stationary Forecasts
Etk <- rep(NA,8)
phi <- coef(model.arima.DS)
Etk[1] <- Y_t.DS.Residuals[length(Y_t.DS.Residuals)]
Etk[2] <- phi[1] * Etk[1] + phi[2] * Y_t.DS.Residuals[length(Y_t.DS.Residuals)-1]
Etk[3] <- phi[1] * Etk[2] + phi[2] * Etk[1]
Etk[4] <- phi[1] * Etk[3] + phi[2] * Etk[2]
Etk[5] <- phi[1] * Etk[4] + phi[2] * Etk[3]
Etk[6] <- phi[1] * Etk[5] + phi[2] * Etk[4]
Etk[7] <- phi[1] * Etk[6] + phi[2] * Etk[5]
Etk[8] <- phi[1] * Etk[7] + phi[2] * Etk[6]
Etk[9] <- phi[1] * Etk[8] + phi[2] * Etk[7]

Etk_DS <- Etk

kable(t(data.frame(k = c("0", "1", "2", "3", "4", "5", "6", "7", "8"), E_X_185plusk_TS = rowMeans(Etk_TS))))

#### Calculation of the 95% Confidence Intervals for  $E_T[\Delta X_{T+k}]$ 

varEtk <- rep(NA,9)

for(j in 1:9){
  varEtk[j] = sigma.hat.squared * sum(psis[1:j]^2)
}

EtkCIlowTS <- rep(NA,9)
EtkCIhighTS <- rep(NA,9)
for(j in 1:9){
  EtkCIlowTS[j] <- Etk_TS[j] - 2 * sqrt(varEtk[j])
  EtkCIhighTS[j] <- Etk_TS[j] + 2 * sqrt(varEtk[j])
}

EtkCIlowDS <- rep(NA,9)
EtkCIhighDS <- rep(NA,9)
for(j in 1:9){
  EtkCIlowDS[j] <- Etk_DS[j] - 2 * sqrt(varEtk[j])
  EtkCIhighDS[j] <- Etk_DS[j] + 2 * sqrt(varEtk[j])
}

```

```

kable(data.frame(k = as.character(0:8), varEtk, EtkCIlowTS, Etk_TS, EtkCIhighTS, EtkCIlowD

#### Forecast Visualizations
qplot(185:(185+8), Etk_TS, geom="line") +
  labs(x= "Time",
       y = "Y_t",
       title="Trend Stationary Forecasts",
       subtitle = "This plot shows the next 8 forecasts, of our Trend Stationary Residuals",
       caption = "Quarterly: [2007:01, 2009:01]") +
  geom_ribbon(aes(ymin=Etk_TS, ymax=EtkCIhighTS), alpha=0.2, show.legend=FALSE) +
  geom_ribbon(aes(ymin=EtkCIlowTS, ymax=Etk_TS), alpha=0.2, show.legend=FALSE) +
  scale_y_continuous(labels = comma)

qplot(185:(185+8), Etk_DS, geom="line") +
  labs(x= "Time",
       y = "Y_t",
       title="Difference Stationary Forecasts",
       subtitle = "This plot shows the next 8 forecasts, of our Difference Stationary Residuals",
       caption = "Quarterly: [2007:01, 2009:01]") +
  geom_ribbon(aes(ymin=Etk_DS, ymax=EtkCIhighDS), alpha=0.2, show.legend=FALSE) +
  geom_ribbon(aes(ymin=EtkCIlowDS, ymax=Etk_DS), alpha=0.2, show.legend=FALSE) +
  scale_y_continuous(labels = comma)

### Question 5

#### Augmented Dickey-Fuller (ADF) Test

library(tseries)

### Question 6

#### Box-Jenkins Identification
model <- c("AR(p)", "MA(q)", "ARMA(p,q)")
rho_k <- c("damped exponential", "cut-off at k=q", "damped exponential")
phi_kk <- c("cut-off at k=p", "damped exponential", "damped exponential")

kable( data.frame(Model = model, Rho_k = rho_k, Phi_kk = phi_kk))

##### Calculation of  $\hat{\rho}(k)$  &  $\hat{\phi}_{kk}$ 

##### Trend Stationary

#calculation of gamma_k
gamma_k <- rep(NA, 11)
gamma_0 <- (1/length(Y_t.TS.Residuals))*sum(Y_t.TS.Residuals^2)
gamma_k[1] <- gamma_0

```

```

for(i in 1:10){
  gamma_k[i+1] <- (1/length(na.omit(lag(Y_t.TS.Residuals,i))))*sum(Y_t.TS.Residuals[-c(1:i)])
}

rho_k <- gamma_k/gamma_k[1]

#obtaining the phi_kk
phi_kk = data.frame(k = 1:10, Phi_kk_TS = rep(NA, 10), Phi_kk_DS = rep(NA, 10))
for(ii in 1:10) {
  TS.model.arima = arima(Y_t.TS.Residuals, order = c(ii, 0, 0), include.mean=FALSE) # TS_Arima
  DS.model.arima = arima(Y_t.DS.Residuals, order = c(ii, 0, 0), include.mean=FALSE) # DS_Arima
  phi.arima_TS = TS.model.arima$coef
  phi.arima_DS = DS.model.arima$coef
  phi_kk[ii , "Phi_kk_TS"] = phi.arima_TS[length(phi.arima_TS)]
  phi_kk[ii , "Phi_kk_DS"] = phi.arima_DS[length(phi.arima_DS)]
}

acf(Y_t.TS.Residuals)

pacf(Y_t.TS.Residuals)

kable(t(data.frame(k = as.character(0:10), gamma_k = round(as.vector(gamma_k),3), rho_k =

arima(Y_t.TS.Residuals, order = c(2, 0, 0), include.mean=FALSE)

##### Difference Stationary

#calculation of gamma_k
gamma_k <- rep(NA, 11)
gamma_0 <- (1/length(Y_t.DS.Residuals))*sum(Y_t.DS.Residuals^2)
gamma_k[1] <- gamma_0

for(i in 1:10){
  gamma_k[i+1] <- (1/length(na.omit(lag(Y_t.DS.Residuals,i))))*sum(Y_t.DS.Residuals[-c(1:i)])
}

acf(Y_t.DS.Residuals)

pacf(Y_t.DS.Residuals)

rho_k <- gamma_k/gamma_k[1]

kable(t(data.frame(k = as.character(0:10), gamma_k = round(as.vector(gamma_k),5), rho_k =

arima(Y_t.DS.Residuals, order = c(1, 0, 0), include.mean=FALSE)

```

```

#### Plot the standardized residuals

##### Trend Stationary Residuals
residualsTS <- arima(Y_t.TS.Residuals, order = c(2, 0, 0), include.mean=FALSE)$residuals
sigma.hat.squared <- sum(residualsTS^2)/length(residualsTS)
sigma.hatTS <- sqrt(sigma.hat.squared)
Zt <- (residualsTS-mean(residualsTS))/sigma.hatTS
hist(Zt, breaks = 50, freq = F, main = "Zt", xlab = "Standardized Residuals")
curve(dnorm(x),col = "red", add = T)

##### Difference Stationary Residuals
residualsDS <- arima(Y_t.DS.Residuals, order = c(1, 0, 0), include.mean=FALSE)$residuals
sigma.hat.squared <- sum(residualsDS^2)/length(residualsDS)
sigma.hatDS <- sqrt(sigma.hat.squared)
Zt <- (residualsDS-mean(residualsDS))/sigma.hatDS
hist(Zt, breaks = 50, freq = F, main = "Zt", xlab = "Standardized Residuals")
curve(dnorm(x),col = "red", add = T)

#### 1) Box-Pierce

##### Trend Stationary

#### Box-Piece test
# use built-in function:
M <- sqrt(length(residualsTS))
M <- 14

Box.test(x = residualsTS, type = "Box-Pierce",lag= M)
# returns p-value when run

##### Difference Stationary

#### Box-Piece test
# use built-in function:
M <- sqrt(length(residualsDS))
M <- 14

Box.test(x = residualsDS, type = "Box-Pierce",lag= M)
# returns p-value when run

#### 2) Overfitting with $r= 4$

##### Trend Stationary
model.arima.a <- arima(Y_t.TS.Residuals, order = c(4, 0, 0), include.mean=FALSE)
sigma.hat.a <- sum(model.arima.a$residuals^2)/N
print("chi-statistic")

```

```

Chi_2 = N * log(sigma.hatTS^2 / sigma.hat.a)
Chi_2
print("chi-critical value")
qchisq(0.95, 2)

##### Difference Stationary

model.arima.a <- arima(Y_t.DS.Residuals, order = c(4, 0, 0), include.mean=FALSE)
sigma.hat.a <- sum(model.arima.a$residuals^2)/N
print("chi-statistic")
Chi_2 = N * log(sigma.hatDS^2 / sigma.hat.a)
Chi_2
print("chi-critical value")
qchisq(0.95, 3)

#### 3) Jarque-Bera

##### Trend Stationary
##### Jarque Bera Test
# can fail this test in 2 ways:
# 1. normality is rejected completely
# 2. the model is normal but didn't pass JB test because of
# the existence of significant outliers or structure breaks

#####
# if an ARMA model passes all diagnostics except for JB due
# to outliers, it's still a good model
#####

# find standardized residuals (same as above)
Zt <- (residualsTS - mean(residualsTS))/sigma.hatTS
Zt_TS <- Zt
# calculate K_3, K_4,
K_3 <- (1/length(Zt))*sum(Zt^3)
K_4 <- (1/length(Zt))*sum(Zt^4)
print("K_3")
K_3
print("K_4")
K_4
# find J_Stat and J_crit.
J_Stat <- length(Zt)*(((K_3^2)/6)+((K_4-3)^2)/24)
print("Jarque-Bera Statistic:")
J_Stat
J_Crit <- qchisq(0.95,2)
print("Jarque-Bera Critical Value:")
J_Crit

```

```
##### Difference Stationary
```

```
##### Jarque Bera Test
```

```
# can fail this test in 2 ways:
```

```
# 1. normality is rejected completely
```

```
# 2. the model is normal but didn't pass JB test because of
```

```
# the existence of significant outliers or structure breaks
```

```
#####
```

```
# if an ARMA model passes all diagnostics except for JB due
```

```
# to outliers, it's still a good model
```

```
#####
```

```
# find standardized residuals (same as above)
```

```
Zt <- (residualsDS - mean(residualsDS)) / sigma.hatDS
```

```
Zt_DS <- Zt
```

```
# calculate K_3, K_4,
```

```
K_3 <- (1/length(Zt)) * sum(Zt^3)
```

```
K_4 <- (1/length(Zt)) * sum(Zt^4)
```

```
print("K_3")
```

```
K_3
```

```
print("K_4")
```

```
K_4
```

```
# find J_Stat and J_crit.
```

```
J_Stat <- length(Zt) * (((K_3^2)/6) + ((K_4-3)^2)/24)
```

```
print("Jarque-Bera Statistic:")
```

```
J_Stat
```

```
J_Crit <- qchisq(0.95, 2)
```

```
print("Jarque-Bera Critical Value:")
```

```
J_Crit
```

```
#### 4) a test for ARCH(6).
```

```
##### Trend Stationary
```

```
### ARCH(6) test
```

```
## again, use standardized residuals (name them std.residuals to use this code)
```

```
std.residuals <- Zt_TS
```

```
N <- length(std.residuals)
```

```
std.res.sq <- std.residuals^2
```

```
ARCH.model <- lm(std.res.sq[-(1:6)] ~ std.res.sq[-c((1:5), N)]
```

```
      + std.res.sq[-c(1:4, (N-1), N)] + std.res.sq[-c(1:3, (N-2):N)]
```

```
      + std.res.sq[-c(1:2, (N-3):N)] + std.res.sq[-c(1, (N-4):N)]
```

```
      + std.res.sq[-((N-5):N)] - 1)
```

```
R.squared <- summary(ARCH.model)$r.squared
```

```
ARCH.test.stat <- N * R.squared
```



```

ARCH.test.crit <- qchisq(0.95,6)
ARCH.Null.Hypothesis <- (ARCH.test.stat > ARCH.test.crit) # if true reject H_0
print("ARCH(6) Statistic:")
ARCH.test.stat
print("ARCH(6) Critical Value")
ARCH.test.crit
print("Statistic > Critical Value?")
ARCH.Null.Hypothesis

##### Difference Stationary

### ARCH(6) test
## again, use standardized residuals (name them std.residuals to use this code)
std.residuals <- Zt_DS
N <- length(std.residuals)
std.res.sq <- std.residuals^2
ARCH.model <- lm(std.res.sq[-(1:6)]~std.res.sq[-c((1:5), N)]
                + std.res.sq[-c(1:4,(N-1),N)] + std.res.sq[-c(1:3, (N-2):N)]
                + std.res.sq[-c(1:2, (N-3):N)] + std.res.sq[-c(1, (N-4):N)]
                + std.res.sq[-((N-5):N)] - 1)
R.squared <- summary(ARCH.model)$r.squared
ARCH.test.stat <- N * R.squared
ARCH.test.crit <- qchisq(0.95,6)
ARCH.Null.Hypothesis <- (ARCH.test.stat > ARCH.test.crit) # if true reject H_0
print("ARCH(6) Statistic:")
ARCH.test.stat
print("ARCH(6) Critical Value")
ARCH.test.crit
print("Statistic > Critical Value?")
ARCH.Null.Hypothesis

### Question 7

library(fGarch)
library(xlsx)
library(dplyr)
library(ggplot2)
library(scales)
P <- read.xlsx("S&P_data.xls", sheetName = "Sheet3")
P <- P[complete.cases(P),]
lnPt <- log(P$P)
qplot(x=seq_along(lnPt), y=lnPt, geom="line") +
  labs(x= "Time",
       y = "Logarithmic S&P",
       title="Logarithmic S&P Series vs. Time",
       subtitle = "A plot of the logarithmic S&P over time, ranging from 1939 to 1992.",

```

```

caption = "Monthly: [1939:01, 1992:12]") +
scale_y_continuous(labels = comma)

model <- lm(lnPt[-1] ~ lnPt[-length(lnPt)])
summary(model)

#### Estimate of the autocorrelation function  $\hat{\rho}_a(k)$ 
at <- model$residuals
acf(at, plot=FALSE)[1:10]

#### Test whether  $a_t$  is normally distributed
hist(at, breaks = 50, freq = F, main = "Histogram of at to test for Normality", xlab = "at",
curve(dnorm(x, mean=0, sd=sd(at)), col = "red", add = T))

#x <- rnorm(length(at), mean = 0, sd=sd(at))

#ks.test(at, x)

#### Estimate the autocorrelation function for  $a_t^2$ 
acf(at^2, plot=FALSE)[1:10]

#### Estimate a GARCH(1,1) model for  $a_t$ 
model.GARCH <- garchFit(formula = ~garch(1, 1), data = at)
#model.GARCH
coef(model.GARCH)

### Question 8

#### SETAR Model

library(tsDyn)

summary(setar(as.ts(at), m=2, thDelay=1, th=median(at), model="TAR"))

```