STAT 443 Final Project, Spring 2017

Daniel Matheson, 20270871

Abstract

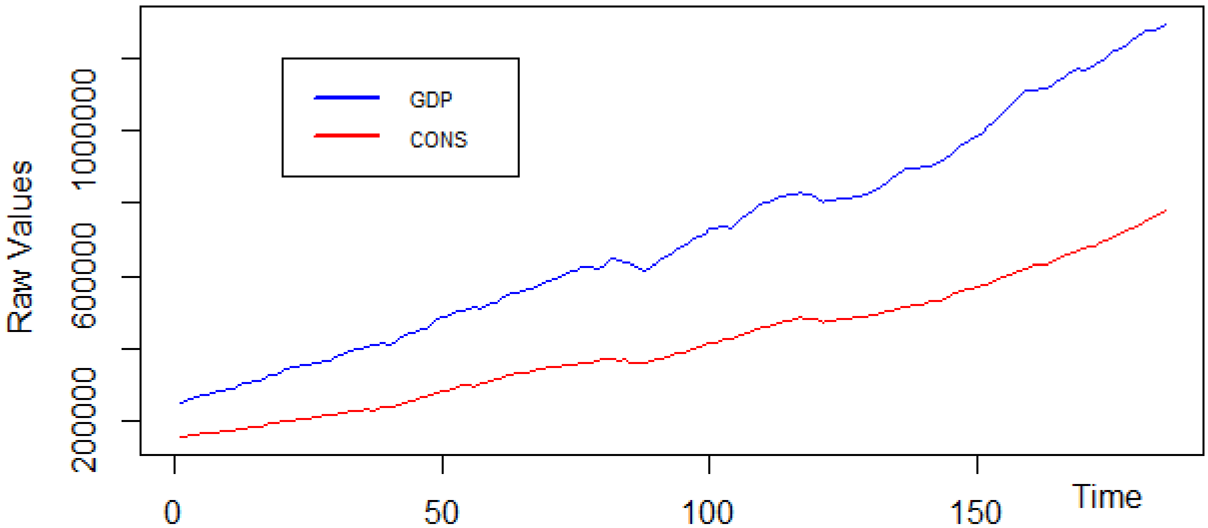
In this report, we examine the quarterly seasonally adjusted Canadian GDP data (GDP), and the quarterly Canadian personal expenditures on consumer goods and services data (CONS); for the time period from 1961 to the first quarter of 2007. Focusing our attention on the GDP data, we use two models to extract the cycle Y_t from the data; trend stationary (TS) and difference stationary (DS). We then show that the optimal AR(p) fits for Y_t are AR(2) for TS and AR(1) for DS (but we force $p \geq 2$). We find 95% confidence intervals for the forecasted growth rates. We then find the optimal ARMA(p,q) models to fit the TS and DS derived Y_t cycles, using Box-Jenkins identification, and then perform several tests to examine the validity and fit of these models.

Lastly, we analyse monthly returns from the S & P index ranging from 1939 to 1992. Specifically, we focus on its residuals and the properties thereof.

Contents

1 Looking at the Data

Before we begin, we have a preliminary look at the raw data for GDP and CONS:



We will turn our attention to the GDP data specifically, and denote it W_{1t}. For our analysis, we will work with $X_t = \ln(W_{1t})$ in order to take advantage of the Gaussian Law; that covariance stationary processes converge asymptotically to a Gaussian distribution as the size of the sample $n \to \infty$.
Once we have performed the transformation $X_t = \ln(W_{1t})$, we can invoke the *Decomposition Law*;

$$X_t = T_t + Y_t + S_t$$
$$\text{such that } T_t \text{ is the time trend component of } X_t,$$
$$Y_t \text{ is the cycle trend component of } X_t,$$
$$\text{and } S_t \text{ is the seasonal trend component of } X_t$$

Because W_{1t} was seasonally adjusted GDP data; S_t here is zero; and hence we have

$$X_t = T_t + Y_t$$

1.1 Extracting the Cycle Y_t

Our goal is to now extract Y_t from the data given, and we have two models of doing so:

1.1.1 Difference Stationary Model

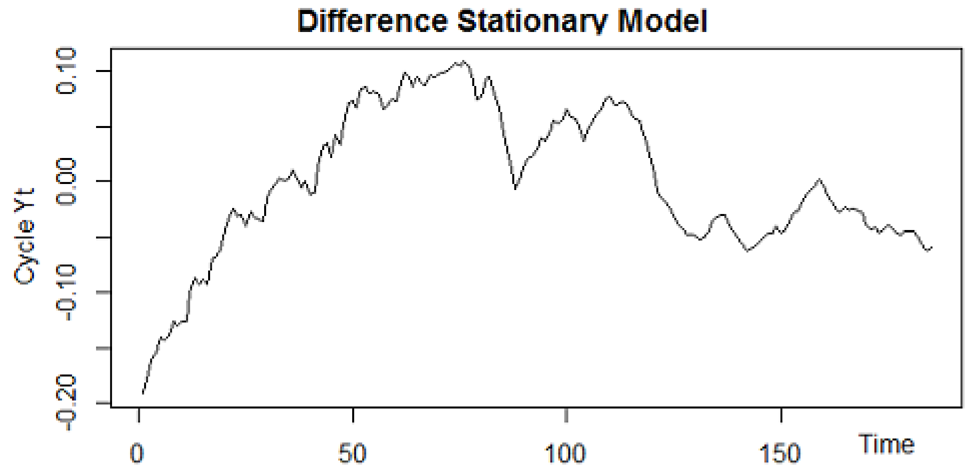
For the Difference Stationary model, we assume that X_t is of the form:

$$\Delta X_t = \mu + Y_t$$
$$\text{such that } \Delta X_t = X_t - X_{t-1}$$

And so if we now perform a linear regression of X_t onto a constant μ, then Y_t will be the residuals of this regression.
When we perform this regression, we find that:

$$\underset{(t)}{X_t} = \underset{(13.84)}{0.00899} + Y_t$$
$$n = 184,\ F = \text{N/A},\ \text{RSS} = 0.0142,\ R^2 = \text{N/A}$$

Hence if we now take the cycle Y_t to be the residuals under this regression, we get the plot of Y_t:



1.1.2 Trend Stationary Model

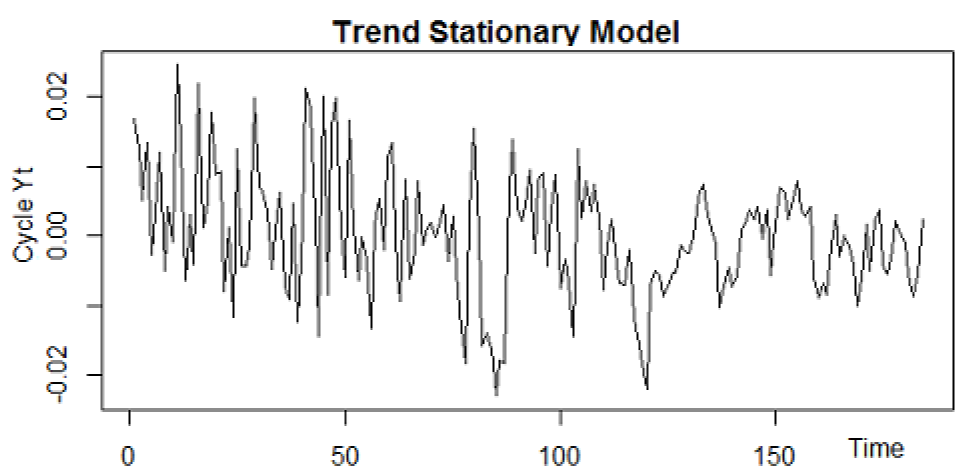
In the Trend Stationary model, we assume that X_t is of the form:

$$X_t = \alpha + \mu t + Y_t$$

And so if we perform a linear regression of X_t onto t, with an intercept term; then Y_t will be the residual terms of this regression. When we perform this regression, we find that

$$\underset{(t)}{X_t} = \underset{(1298.21)}{12.61} + \underset{(91.26)}{0.00827t} + Y_t$$
$$n = 185,\ F = 8328,\ \text{RSS} = 0.7916669,\ R^2 = 0.9784$$

Hence if we now take the cycle Y_t to be the residuals under this regression, the plot of Y_t is:



2 Fitting Y_t with an AR(p) model

We will now try to determine which AR(p) could fit this GDP cycle data Y_t the best.

In order to decide which value of p gives us the best fit, we must use some sort of metric to compare the candidate models. We have decided to use the Bayesian Information Criterion, defined as:

$$\text{BIC}(k) = \ln(\hat{\sigma}_k^2) + \frac{\ln(n) \cdot k}{n}$$

such that k is the number of parameters in the AR(p=k) model,

$\hat{\sigma}_k^2$ represents the estimated variance of the residuals for the AR(k) model,

and n is the number of data points

We will calculate BIC(k) for both the Trend Stationary and Difference Stationary models to choose a candidate AR(p) model.
And from the chosen estimated model - for Trend Stationary only - we will calculate the autocorrelation functions $\rho(k)\ \forall\ k = 0, \dots 6$, infinite moving average weights $\psi_k = \phi_{kk}\ \forall\ k = 0, \dots 6$, and the standard deviation of Y_t, $\gamma(0)^{\frac{1}{2}}$.

2.1 Difference Stationary Y_t

For the Y_t from the Difference Stationary model, we get the following values of BIC$(k), k = 1, \dots, 10$:

k	1	2	3	4	5	6	7	8	9	10
BIC(k)	-9.555	-9.530	-9.517	-9.489	-9.462	-9.438	-9.415	-9.396	-9.413	-9.386

Where the minimum value (-9.555) corresponds to $k = 1$, i.e an AR(1) model. However, we will restrict ourselves to $p \geq 2$, in which case $k = 2$ is the new minimum with BIC$(k) = -9.530$, i.e an AR(2) model. So we conclude that Y_t from the Difference Stationary model is best fit by an AR(2) model.

2.2 Trend Stationary Y_t

For the Y_t from the Trend Stationary model, we get the following values of BIC$(k), k = 1, \dots 10$:

k	1	2	3	4	5	6	7	8	9	10
BIC(k)	-9.435	-9.524	-9.500	-9.488	-9.460	-9.433	-9.410	-9.388	-9.368	-9.389

Where the minimum value (-9.524) corresponds to $k = 2$, i.e an AR(2) model.
So we conclude that Y_t from the Trend Stationary model is best fit by an AR(2) model:

$$\underset{(t)}{Y_t} = \underset{(18.43)}{1.333} Y_{t-1} - \underset{(-4.57)}{0.338} Y_{t-2} + a_t$$

$$n = 185,\ F = 5355,\ \text{RSS} = 0.0120,\ R^2 = 0.9832$$

To find $\rho(k), k = 0, \dots, 6$ for Y_t, we use the built-in R function $ARMAacf$ which takes parameters $\hat{\phi}_1, \hat{\phi}_2$ from the regression above.
And we calculate the values of $\psi_k, k = 0, \dots, 6$ using the fact that

$$\psi_k = \hat{\phi}_1 \psi_{k-1} + \hat{\phi}_2 \psi_{k-2} = 1.333\hat{\phi}_1 - 0.338\psi_{k-2}$$

such that $\psi_0 = 1$ and $\phi_{kk} = 0\ \forall\ k < 0$

The results are:

k	0	1	2	3	4	5	6
$\rho(k)$	1	0.982	0.954	0.923	0.892	0.861	0.831
ϕ_{kk}	1	1.290	1.351	1.338	1.303	1.262	1.219

And finally, to find $\gamma(0)^{\frac{1}{2}}$ we use the formula for $\gamma(0$ of an AR(2):

$$\gamma(0) = \frac{\hat{\sigma^2}}{1 - \hat{\phi}_1\rho(1) - \hat{\phi}_2\rho(2)} = 0.002053715$$
$$\Rightarrow \gamma(0)^{\frac{1}{2}} = 0.04532$$

3 Forecasting Growth Rates

Using the cycle Y_t from both Trend Stationary and Difference Stationary models, we will forecast the growth rates $\Delta X_{T+k}\ \forall\ k = 0, \ldots, 8$ where T is the last observation in the sample.

3.1 Difference Stationary

To forecast the growth rates in the case where X_t is considered to be Difference Stationary:

$$\Delta X_t = \mu + Y_t$$
$$\underset{(t)}{\Delta X_t} = \underset{(13.84)}{0.0089863} + Y_t$$
$$n = 184, \text{RSS} = 0.00881$$

Then we have the following results which we can use to find a 95% confidence interval for ΔX_{T+k}:

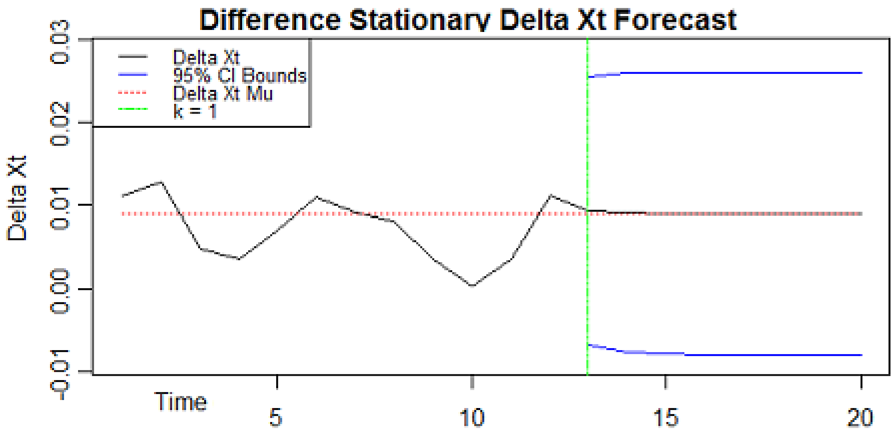
$$E_T\left[\Delta X_{T+k}\right] = \mu + E_T\left[Y_{T+k}\right]$$
$$\text{Var}_T\left(\Delta X_{T+k}\right) = \text{Var}_T\left(Y_{T+k}\right) = \hat{\sigma}^2 \sum_{j=0}^{k-1} \psi_j^2 \text{ such that } \psi_0^2 = 1$$
$$\text{where } Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + a_t$$
$$\underset{(t)}{Y_t} = \underset{(3.997)}{0.29695} Y_{t-1} + \underset{(0.780)}{0.05743} Y_{t-2} + a_t$$
$$n = 185, F = 10.42, \text{RSS} = 0.008275, R^2 = 0.104 \text{ (very low } R^2\text{)}$$

We restricted the choice of AR(p) to $p \geq 2$ so $\hat{\phi}_2$ is actually not significant with $R^2 = 0.104$ for the regression telling us it's not a good fit, and a t value of 0.780; or a p-value of 0.436; so $H_0 : \hat{\phi}_2 = 0$ is not rejected. Using these formulas, we found the following results:

k	$E_T\left[Y_{T+k}\right]$	$E_T\left[\Delta X_{T+k}\right]$	$\text{Var}_T\left(\Delta X_{T+k}\right)$	95% Confidence Interval
0	0.00221	0.01119	0	0.01119
1	0.00034	0.00933	6.772×10^{-5}	$[-0.00680, 0.02546]$
2	0.00023	0.00922	7.36984×10^{-5}	$[-0.00761, 0.02604]$
3	0.00009	0.00907	7.51344×10^{-5}	$[-0.00792, 0.02606]$
4	0.00004	0.00903	7.53806×10^{-5}	$[-0.00799, 0.02604]$
5	0.00002	0.00900	7.54273×10^{-5}	$[-0.00802, 0.02603]$
6	0.00001	0.00899	7.54359×10^{-5}	$[-0.00803, 0.02602]$
7	0.00000	0.00899	7.54375×10^{-5}	$[-0.00803, 0.02601]$
8	0.00000	0.00899	7.54378×10^{-5}	$[-0.00804, 0.02601]$

We notice that $E_T\left[\Delta X_{T+k}\right]$ is approaching it's forecasted mean $\hat{\mu} \approx 0.00899$, as expected; and the confidence intervals are also converging and remaining stable at around $[-0.0008, 0.026]$.
Below, we present a graph for 20 points of differenced GDP data
($\{\Delta X_{T+k} \mid k = -11, \ldots, 8$ such that T is the final observation $\}$). That is; we took the last 12 known

data points, and the next 8 points (beginning where the plot is split by the vertical green line) are those which were forecasted above. The blue lines represent the lower and upper 95% confidence interval bounds, and the dotted red line represents the value of $\hat{\mu}$.



3.2 Trend Stationary

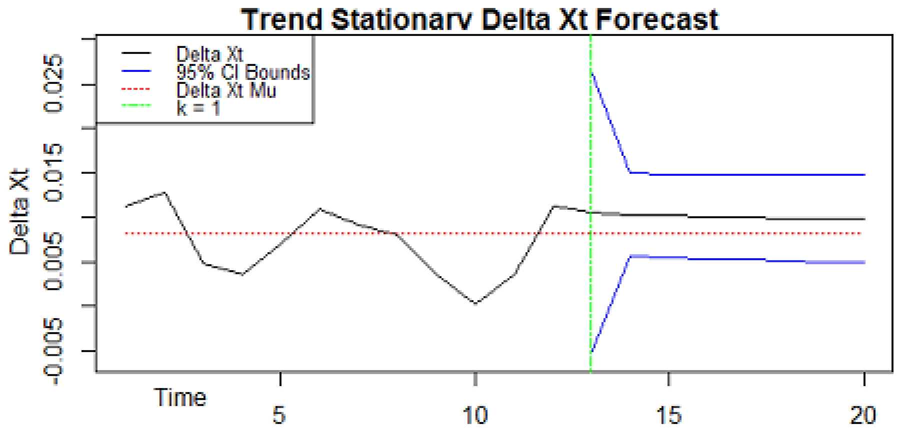
The process from the Trend Stationary approach is a little more complex. When X_t is Trend Stationary,

$$
\begin{aligned}
\underset{(t)}{X_t} &= \underset{(1298.21)}{12.61} + \underset{(91.26)}{0.00827t} + Y_t \quad \text{(from Section 1.1.2)} \\
\Rightarrow \Delta X_{T+k} &= \mu + Y_{T+k} - Y_{T+k-1} \\
\Rightarrow E_T\left[\Delta X_{T+k}\right] &= \mu + E_T\left[Y_{T+k}\right] - E_T\left[Y_{T+k-1}\right] \\
\Rightarrow \mathrm{Var}_T\left(\Delta X_{T+k}\right) &= \sigma^2 \left(1 + \sum_{j=1}^{k-1} \left(\psi_j - \psi_{j-1}\right)^2\right) \\
\text{where } Y_t &= \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + a_t \\
&= 0.29695 Y_{t-1} \quad + 0.05743 Y_{t-2} + a_t \\
(t) \quad & \quad (3.997) \qquad\qquad (0.780) \\
n &= 185, F = 10.42, \mathrm{RSS} = 0.008275, R^2 = 0.104 \text{ (very low } R^2\text{)}
\end{aligned}
$$

Using these formulas, we got the following results:

k	$E_T\left[Y_{T+k}\right]$	$E_T\left[\Delta X_{T+k}\right]$	$\mathrm{Var}_T\left(\Delta X_{T+k}\right)$	95% Confidence Interval
0	-0.05864	0.01119	0	0.01119
1	-0.05636	0.01055	6.61722×10^{-6}	$[-0.00540, 0.02649]$
2	-0.05433	0.01029	5.55950×10^{-6}	$[0.00567, 0.01491]$
3	-0.05243	0.01016	5.80360×10^{-6}	$[0.00544, 0.01488]$
4	-0.05062	0.01008	5.81380×10^{-6}	$[0.00535, 0.01481]$
5	-0.04887	0.01001	5.89510×10^{-6}	$[0.00525, 0.01477]$
6	-0.04719	0.00995	6.00800×10^{-6}	$[0.00514, 0.01475]$
7	-0.04556	0.00989	6.12650×10^{-6}	$[0.00504, 0.01474]$
8	-0.04399	0.00983	6.24120×10^{-6}	$[0.00494, 0.01473]$

From these results, it's worth questioning why we aren't getting $E_T\left[Y_{T+k}\right] \to 0$, and why the confidence intervals are strictly positive after $k = 1$. It is outside the scope of this paper to investigate this phenomenon, so we will simply present the graph of the results as we did previously in Section 3.1:



4 ADF Test: Is X_t Difference Stationary or Trend Stationary?

In this section, we will perform the Augmented Dickey-Fuller Test on X_t. This test is used to determine if the time series X_t is Difference Stationary or Trend Stationary.
To accomplish this test, we consider the following regression on ΔX_t:

$$\Delta X_t = \alpha + \mu t + \phi X_{t-1} + \theta_1 \Delta X_{t-1} + \ldots + \theta_j \Delta X_{t-j} + a_t$$

Where j is a parameter of the test. We will be using $j = 5$.
With this regression, we want to consider the null hypothesis $H_0 : \phi = 0$ which would indicate that the model is Difference Stationary.
Using built-in R functions, we find that the $p-$value under H_0 is 0.2865, so we do not reject H_0.
Hence X_t is Difference Stationary according to the Augmented Dickey-Fuller test with $j = 5$ lags.
Our results from Section 3 seem to agree with this result; the differenced X_t values appear to behave more as we would expect in the Difference Stationary models (Section 3.1).

5 Fitting Y_t with an ARMA(p,q) model

We now proceed to using Box-Jenkins identification to determine some candidate ARMA(p,q) models for both Trend Stationary and Difference Stationary GDP cycles Y_t.
The goal is to choose p, q such that ARMA(p,q) best fits the data. Y_t is said to follow ARMA(p,q) if

$$Y_t = \sum_{j=1}^{p} \phi_j Y_{t-1} + \sum_{\ell=1}^{q} \theta_\ell a_{t-\ell} + a_t$$

In order to find these optimal values, we will use Box-Jenkins Identification. Box-Jenkins identification simply requires us to calculate the autocorrelation function $\rho(k)$ and partial autocorrelation function ϕ_{kk} for Y_t and observe if these functions exhibit damped exponential behavior, or have a cut-off property: that is, we can distinguish ARMA(p,q) models by the following guidelines:

Model	$\rho(k)$	ϕ_{kk}
AR(p)	damped exp.	cut-off at $k = p$
MA(q)	cut-off at $k = q$	damped exp.
ARMA(p,q)	damped exp.	damped exp.

In order to assess whether a value of $\rho(k)$ or ϕ_{kk} meets the cut-off property, we assume that we have the following asymptotic distributions:

$$\rho(k) \sim \mathcal{N}\left(0, \frac{1}{n}\right) \qquad \phi_{kk} \sim \mathcal{N}\left(0, \frac{1}{n}\right)$$

And hence for each k we test the null hypothesis $H_0^{(1)} : \rho(k) = 0$ or $H_0^{(1)} : \phi_{kk} = 0$, and

$$|\rho(k)| \leq 1.96\,\mathrm{SE}(\rho(k)) = 1.96\frac{1}{\sqrt{n}} \quad \Rightarrow \text{ accept } H_0^{(1)}$$
$$\text{or } |\phi_{kk}| \leq 1.96\,\mathrm{SE}(\phi_{kk}) = 1.96\frac{1}{\sqrt{n}} \quad \Rightarrow \text{ accept } H_0^{(2)}$$

And once we accept $H_0^{(i)}$ for some k then the cut-off property is said to have happened at $k-1$, since $\rho(k) = 0$ or $\phi_{kk} = 0$ respectively.
We first must find $\rho(k)$ and ϕ_{kk} for Y_t. To find $\rho(k)$ we again use the built-in R function *acf*, but we no longer have a model to use that can help us find ϕ_{kk} so we must do it the long way. That is, we must run the following regressions:

$$Y_t = \phi_{1k}Y_{t-1} + \ldots + \phi_{kk}Y_{t-k} + a_t$$

to get ϕ_{kk} for every value of $k = 1, 2, \ldots, 20$, where we limit the value of k at around 20 since otherwise we are observing long-range dependence.

5.1 Trend Stationary Y_t

We first perform the Box-Jenkins identification for the GDP cycle Y_t from the Trend Stationary model, because the Difference Stationary model is far more complicated, as we will see later. We found $\rho(k), \phi_{kk}$ for $k = 1, \ldots, 20$ where $\rho(0) = 1, \phi_{00} = 1$ by definition. These were the results (for the first $k = 0, \ldots, 8$ - see the plots on page 9 for the rest)

k	0	1	2	3	4	5	6	7	8
ϕ_{kk}	1	0.99736	-0.33762	-0.06497	-0.13247	0.00156	0.02675	-0.07244	-0.07985
$\rho(k)$	1	0.96560	0.92933	0.89377	0.85701	0.82241	0.78697	0.75106	0.71609

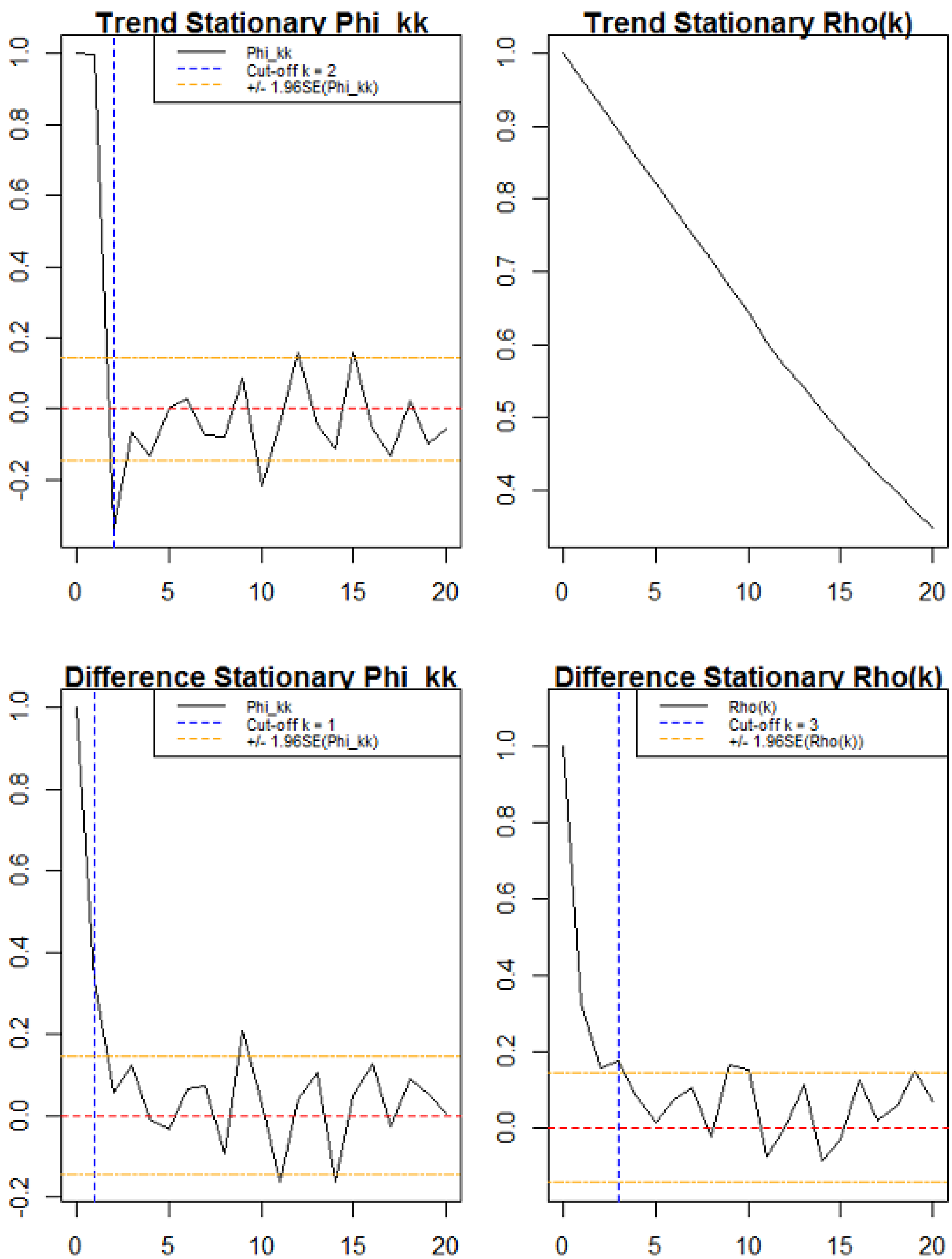
From these results it is fairly clear that the $\rho(k)$ are exhibiting damped exponential growth; as slow as it might be. And ϕ_{kk} is exhibiting a cut-off property, but it's not exactly clear by manual inspection where this cut-off is occurring. When we test each ϕ_{kk} individually we find that the cut-off occurs at $k = 2$; i.e $H_0 : \phi_{kk} = 0$ is not rejected for $k = 3$ and (most) $k > 3$. This is seen in the plots on the next page.
We clearly see that after $k = 2$, the values of ϕ_{kk} stay within the bands $\pm 1.96\frac{1}{\sqrt{n}}$, more or less; and the $\rho(k)$ values are (slowly) decaying exponentially. And so we estimate that, according to Box-Jenkins identification, the Trend Stationary Y_t follows AR(2).

5.2 Difference Stationary Y_t

We proceed in the same manner as we did for Trend Stationary Y_t above. We get:

k	0	1	2	3	4	5	6	7	8
ϕ_{kk}	1	0.32954	0.05816	0.12485	-0.00812	-0.03389	0.06506	0.07275	-0.09470
$\rho(k)$	1	0.32459	0.15393	0.17383	0.08189	0.01559	0.07245	0.10253	-0.02529

It is not clear at all from the table what it going on, so we test each $\phi_{kk}, \rho(k)$ individually under $H_0 : \phi_{kk} = 0$ or $H_0 : \rho(k) = 0$ respectively. From this, we find that ϕ_{kk} exhibits a cut-off property at $k = 1$ and $\rho(k)$ at $k = 3$. This is seen in the plots below.
We encounter a problem, however; as Box-Jenkins identification does not account for the case where both ϕ_{kk} and $\rho(k)$ exhibit a cut-off property. In order to deal with this issue; we will simply fit both an AR(1) and MA(3) to the Difference Stationary cycle Y_t, and perform model diagnostic tests on both.



6 Diagnostic Tests

We first describe all of the diagnostic tests that we will perform in Sections 6.0.1-6.0.4, and present the results for all the chosen ARMA(p,q) models in Section 6.0.5.

6.0.1 Box-Pierce Test: Are the residuals of Y_t uncorrelated?

The Box-Pierce Portmanteau Test is used to test whether the error terms $a_t = Y_t - E_{t-1}[Y_t]$ are correlated. We denote the correlation $\text{cor}(\hat{a}_t, \hat{a}_{t+k}) = \rho_a(k)$, and we want to test the null hypothesis:

$$H_0 : \rho_a(t) = 0 \ \forall \ k, 1, 2, \ldots, M$$

Where as a rule of thumb we take $M \sim \sqrt{n}$, n being the number of data points in the sample.
In order to test H_0 we use the estimator for $\rho_a(k)$:

$$\hat{\rho}_a(k) = \frac{\sum \hat{a}_t \hat{a}_{t+k}}{\sum \hat{a}_t^2}$$

$$\text{such that under } H_0 \text{ it can be shown that} \sqrt{n}\hat{\rho}_a(k) \overset{\text{approx}}{\sim} \mathcal{N}(0,1)$$

Using the fact that if $X_i \overset{\text{iid}}{\sim} \mathcal{N}(0,1), X_i^2 \sim \chi^2(1)$ and $\sum_{i=1}^N X_i^2 \sim \chi^2(N)$, the following test statistic can be derived:

$$\begin{aligned}\mathcal{Q} &= \left(\left(\sqrt{n}\hat{\rho}_a(1)\right)^2 + \left(\sqrt{n}\hat{\rho}_a(2)\right)^2 + \ldots + \left(\sqrt{n}\hat{\rho}_a(M)\right)^2\right) \\ &= n\left(\hat{\rho}_a(1) + \ldots + \hat{\rho}_a(M)\right) \overset{\text{approx}}{\sim} \chi^2(M)\end{aligned}$$

And so we reject H_0 if $\mathcal{Q} > \mathcal{Q}_{\text{crit}}$ which is the 95^{th} quantile of the $\chi^2(M)$ distribution.

6.0.2 Over-fitting with $r = 4$ Test

The over-fitting test is very simple; if you estimate that your model is an AR(p), then you try to fit it to an AR(p+r) and compare the two models; if the model is an MA(q), you do the same with MA(q+r).
In order to test the comparative fits of the models, we will be using the likelihood ratio test.
That is, for the AR(p+r) = AR(p+4) process

$$Y_t = \phi_1 Y_{t-1} + \ldots + \phi_p Y_{t-p} + \phi_{p+1} Y_{t-(p+1)} + \ldots + \phi_{p+4} Y_{t-(p-4)} + a_t$$

We must test the following null hypothesis:

$$H_0 : \phi_{p+1} = \phi_{p+2} = \phi_{p+3} = \phi_{p+4} = 0$$

In order to do so, we will use the likelihood ratio test statistic:

$$\Lambda = n \ln\left(\frac{\hat{\sigma}_p^2}{\hat{\sigma}_{p+4}^2}\right) \overset{\text{approx}}{\sim} \chi^2(4)$$

where n is taken to be the number of observations minus $p+4$, $\hat{\sigma}_p^2$ is the variance coming from the AR(p) model, and $\hat{\sigma}_{p+4}^2$ from the AR(p+4) model.
Once we have calculated this test statistic, we reject H_0 if $\Lambda > \Lambda_{\text{crit}}$ which is the 95^{th} quantile of the $\chi^2(4)$ distribution.
The result is analogous for MA(q+4);

$$\begin{aligned}Y_t &= \theta_1 a_{t-1} + \ldots + \theta_{p+4} a_{t-(p+4)} + a_t \\ H_0 : &\theta_{p+1} = \theta_{p+2} = \theta_{p+3} = \theta_{p+4} = 0\end{aligned}$$

and we use the likelihood ratio test once again.

6.0.3 Are the residuals normally distributed?

We will test if the residuals are normally distributed in two ways:
(1) First we will plot the standardized residuals of the models, and compare these to the normal distribution in a qualitative manner.
(2) The other test is much more rigorous; the Jarque-Bera test:
The Jarque-Bera test is based on the skewness and kurtosis of the residuals. We take the residuals as a random variable, $A \sim (\mu, \sigma^2)$, and define $Z = \frac{A-\mu}{\sigma}$.
The idea is that if $Z \sim \mathcal{N}(\mu, \sigma^2)$, then skewness $k_3 = 0$ and kurtosis $k_4 = 3$.
To find the empirical values of skewness and kurtosis from data, we use the following formulas:

$$\hat{k}_3 = \frac{1}{n}\sum_{t=1}^{n} \hat{z}_t^3 \qquad \hat{k}_4 = \frac{1}{n}\sum_{t=1}^{n} \hat{z}_t^4$$

Let n denote the sample size, then;
Under $H_0 : k_3 = 0$, it can be shown that $\hat{t}_3 = \sqrt{\frac{n}{6}}\hat{k}_3 \overset{\text{approx}}{\sim} \mathcal{N}(0,1)$,
Under $H_0 : k_4 = 3$, it can be shown that $\hat{t}_4 = \sqrt{\frac{n}{24}}(\hat{k}_4 - 3) \overset{\text{approx}}{\sim} \mathcal{N}(0,1)$,
And finally, under $H_0 : k_3 = 0, k_4 = 3$ it can be shown that the Jarque-Bera test statistic

$$\mathcal{J} = \hat{t}_3^2 + \hat{t}_4^2 = n\left(\frac{\hat{k}_3^2}{6} + \frac{(\hat{k}_4 - 3)^2}{24}\right) \overset{\text{approx}}{\sim} \chi^2(2)$$

Then we reject H_0 if $\mathcal{J} > \mathcal{J}_{\text{crit}}$ which is the 95$^{\text{th}}$ quantile of the $\chi^2(2)$ distribution,
Note that the residuals can fail the Jarque-Bera test in two ways:
(1) Normality of residuals is rejected completely.
(2) The model is normal but didn't pass the test because of the existence of significant outliers or structural breaks.

6.0.4 ARCH(6) Test

The objective of the Autoregressive Conditional Heteroskedasticity (ARCH(6)) test is to test whether there is non-linear dependence in the residuals.
An ARCH(6) model for Y_t is written as:

$$\begin{gathered} Y_t = \beta_0 + \beta_1 X_{1t} + \ldots + \beta_k X_{kt} + a_t \\ \text{such that } a_t = z_t\left(\sigma^2 + \alpha_1 a_{t-1}^2 + \ldots + \alpha_6 a_{t-6}^2\right)^{\frac{1}{2}} \\ \text{with } z_t \overset{\text{iid}}{\sim} \mathcal{N}(0,1) \text{ and } \sigma^2 = \text{Var}(a_t) \end{gathered}$$

With this in mind, we examine the conditional variance of a_t dependent on the information set (sample known) I_t:

$$\text{Var}\left(a_t \mid I_t\right) = \sigma^2 + \alpha_1 a_{t-1}^2 + \ldots + \alpha_6 a_{t-6}^2$$

And hence if the null hypothesis $H_0 : \alpha_1 = \ldots = \alpha_6 = 0$ is true, then $\text{Var}\left(a_t \mid I_t\right) = \sigma^2$; i.e, there is no dependence between the residuals a_t.
To perform the ARCH(6) test we simply follow the following steps:
(1) Run the regression shown above on Y_t to get the residuals $\hat{a}_t$; and approximate $\text{Var}(a_t)$ with $\text{Var}(\hat{a}_t)$.
(2) Run the regression:

$$\hat{a}_t^2 = \phi_0 + \phi_1 \hat{a}_{t-1}^2 + \ldots + \phi_6 \hat{a}_{t-6}^2 + \epsilon_t$$

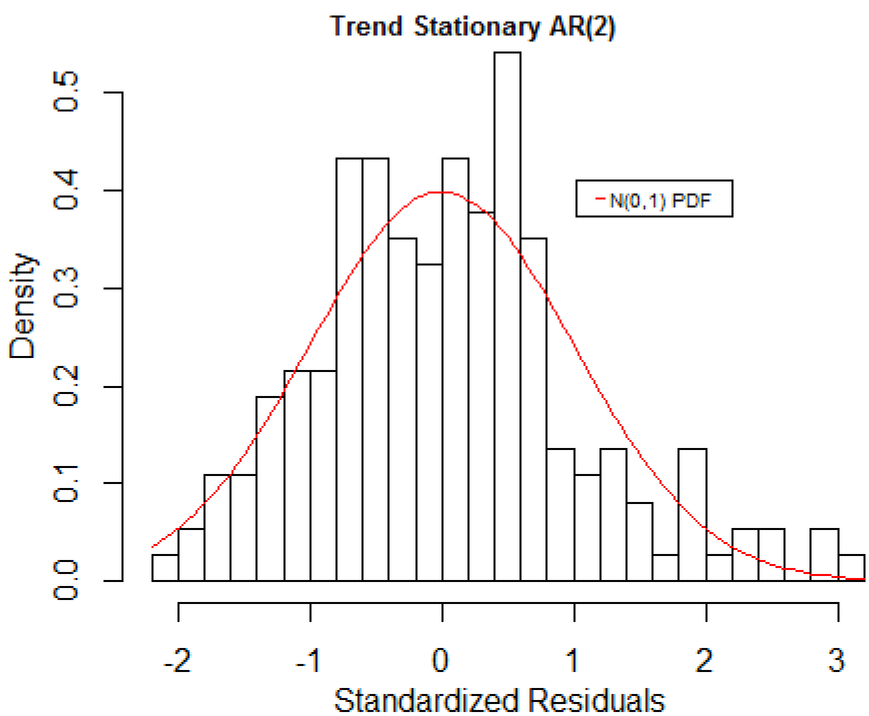
(3) Compute the coefficient of determination R^2 from this regression.
(4) And use the test statistic $nR^2 \overset{\text{approx}}{\sim} \chi^2(6)$ to reject H_0 if $nR^2 > \chi^2(6)_{\text{crit}}$.

6.1 Diagnostic Results

We now present the results of the diagnostic tests explained in Sections 6.0.1-6.0.4:
For Trend Stationary Y_t, fit to an AR(2) model:

Test	Value	Pass/Fail
Box-Pierce	$p = 0.003365$	FAIL
Overfitting	$3.95 < \chi^2(4)_{\text{crit}} = 9.49$	PASS
Jarque-Bera	$9.57 > \chi^2(2)_{\text{crit}} = 5.99$	FAIL
ARCH(6)	$66.34 > \chi^2(6)_{\text{crit}} = 12.59$	FAIL

We also have the plot of the standardized residuals, to check for normality along with Jarque-Bera:



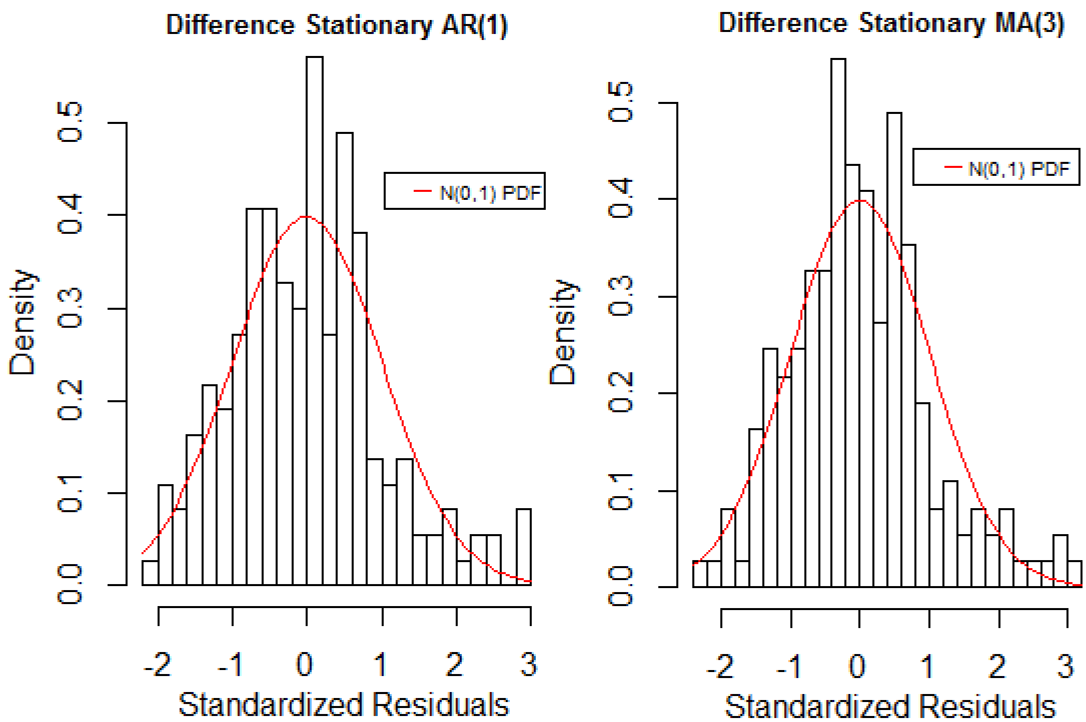
In this plot we can see that the residuals have a higher kurtosis than the $\mathcal{N}(0, 1)$ distribution, and appear to be skewed to the left; as well as having too many values in the range around $(1.75, 3.25)$. So it makes sense that the residuals were found not to be normally distributed by the Jarque-Bera test.
For the Difference Stationary Y_t, fit to an AR(1) model:

Test	Value	Pass/Fail
Box-Pierce	$p = 0.005105$	FAIL
Overfitting	$4.66 < \chi^2(4)_{\text{crit}} = 9.49$	PASS
Jarque-Bera	$67.68 > \chi^2(2)_{\text{crit}} = 5.99$	FAIL
ARCH(6)	$67.685 > \chi^2(6)_{\text{crit}} = 12.59$	FAIL

For the Difference Stationary Y_t, fit to an MA(3) model:

Test	Value	Pass/Fail
Box-Pierce	$p = 0.09984$	PASS (barely)
Overfitting	$3.64 < \chi^2(4)_{\text{crit}} = 9.49$	PASS
Jarque-Bera	$13.32 > \chi^2(2)_{\text{crit}} = 5.99$	FAIL
ARCH(6)	$63.77 > \chi^2(6)_{\text{crit}} = 12.59$	FAIL

We also have the plot of the standardized residuals, to check for normality along with Jarque-Bera:



In this plot we see the same issues as in the Trend Stationary's standardized residuals plot: the kurtosis is higher than that of the normal distribution, and the residuals are skewed to the left, with too many large values in the $\sim [2, 3]$ range.

From these results, one thing should be clear: regardless of which way we extracted the cycle Y_t from the GDP data, no ARMA(p,q) model was appropriate for fitting it. The Over-fitting test was passed every time, and the ARMA model was selected using Box-Jenkins identification. So we assert that it is safe to assume these were the best possible candidate ARMA models; but they still failed every other test, sometimes quite dramatically; with the ARCH(6) test especially, with values 5 times as large as the critical value.
We conclude that ARMA(p,q) is not a good model to fit the GDP cycle data; we should search for other candidate models.

7 S & P Time Series Analysis

In this last section, we will analyze monthly returns from the $S\&P$ index ranging from 1939 to 1992. We first take the returns, denoted P_t and consider $\ln(P_t)$. We then attempt the regression:

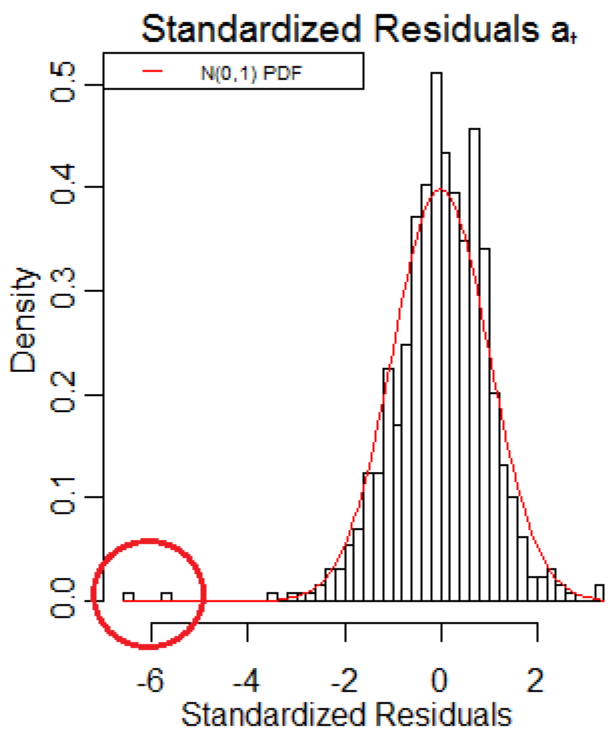
$$\begin{aligned} \ln(P_t) &= \delta + \phi \ln(P_{t-1}) + a_t \\ &= 0.005455 + 1.000014 \ln(P_{t-1}) + a_t \\ (t) \quad & \quad (0.784) \qquad (610.622) \\ & n = 648, F = 372900, \text{RSS} = 0.04363, R^2 = 0.9983 \end{aligned}$$

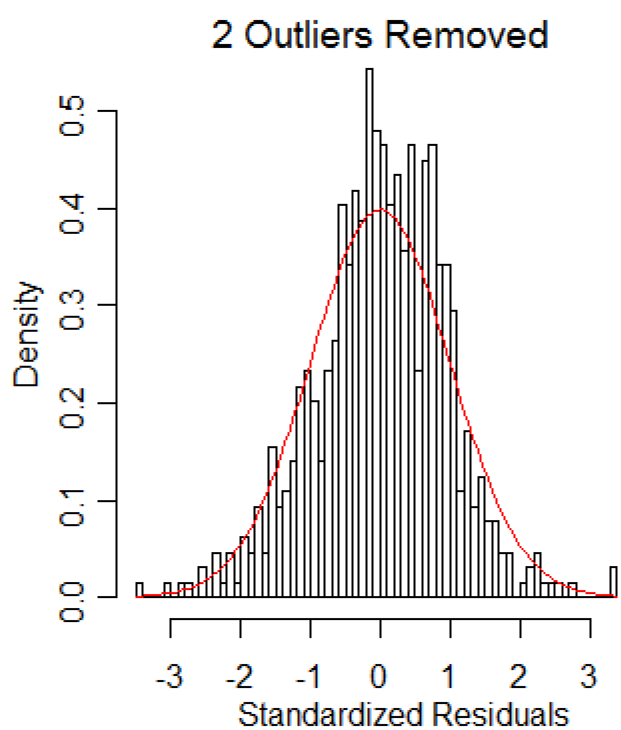
From these results it is clear that $\hat{\phi} \approx 1$; indeed, $\hat{\phi}$ has standard error $0.001638 > (\hat{\phi} - 1)$ meaning that the simple hypothesis test $H_0 : \phi = 1$ is not rejected.

The value of $\hat{\delta}$ is both extremely small, and extremely insignificant; with a t value of only 0.784; translating to a p-value of 0.433 for $H_0 : \delta = 0$. So we simply ignore δ. This means that:

$$\ln(P_t) = \ln(P_{t-1}) + a_t \Rightarrow \ln(P_t) - \ln(P_{t-1}) = a_t$$

Due to this fact; if we can show that $a_t \overset{\text{iid}}{\sim} \mathcal{N}(0, \sigma^2)$, then the series $\ln(P_t)$ follows a random walk.
It is easy to show that $E[a_t] = 0$, by simply taking the sample mean, which converges to the true expected value as $n \to \infty$ by Central Limit Theorem.
We found this sample mean to be $-1.198456 \times 10^{-18}$ which is essentially zero.
It remains to show that the residuals are independent, and Normally distributed:
For independence, we first use the Box-Pierce test, and find for the null hypothesis $H_0 : \rho_a(k) = 0 \ \forall \ k = 1, \ldots, M \approx \sqrt{648}$ we get a p-value of 0.6529 so we do not reject H_0: the residuals are uncorrelated. This is sufficient for Normally distributed data, as a multivariate normal distribution is characterized only by correlation, and not any other form of dependence.
The last task left is to check for normality of the residuals. When we performed the Jarque-Bera test on the residuals a_t, we got a test statistic value of $\mathcal{J} = 547 >> \chi^2(2)_{\text{crit}} = 5.99$, so we reject the null hypothesis $H_0 : k_3 = 0, k_4 = 3$; hence we find that the residuals are *not* normal!
However, upon inspecting the plot of the standardized residuals below, we saw that it may be possible that this test is simply failing due to some extremely large outliers (circled in red); so we remove these *two* data points of 648, and perform the Jarque-Bera test again on the residuals that remain.





The two residuals removed correspond to the changes in the *S&P* on April 1940, and October 1987. April 1940 was in the middle of WW2 which saw the crash of many stock market indices[1]. October 1987 was one of the worst market crashes in history, euphemistically named "Black Monday"[2]. This crash was so

[1] Frederic S. Mishkin & Eugene N. White. *U.S. Stock Market Crashes and Their Aftermath: Implications for Monetary Policy* https://pdfs.semanticscholar.org/562e/4d2940c0e89df4d2c42e58560bc6dc8b9377.pdf, page 5.
[2] Same source as above.

severe, that to this day, there is a superstitious belief called the "October Effect" which is the theory that stocks tend to decline during the month of October; so some investors may be nervous during October because the dates of some large historical market crashes in this month.[3]
Once these two clear outlier residuals are removed, the data looks fairly normal from a qualitative glance at the plot on the right (on the previous page); and when we perform the Jarque-Bera test again, it does indeed give us a test statistic value of only $\mathcal{J} = 1.398 < 5.99$; hence we do not reject the null hypothesis $H_0 : k_3 = 0, k_4 = 3$, hence the residuals are normally distributed, as long as we ignore these two points in the data.
Therefore we have shown that the residuals from the regression follow a Normal distribution:

$$\ln(P_t) - \ln(P_{t-1}) = a_t \overset{\text{iid}}{\sim} \mathcal{N}(0, \sigma^2)$$

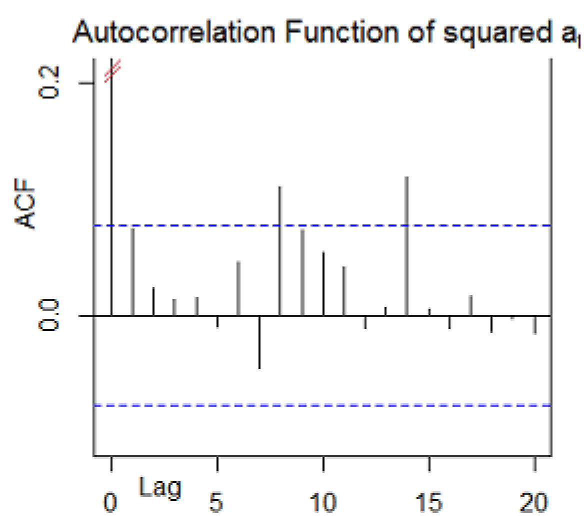
Which makes this series a random walk.
It may also be of interest for us to examine the fit of a GARCH(1,1) model to these residuals, as well as estimating the autocorrelation function of a_t^2.
a_t follows a Generalized Autoregressive Conditional Heteroskedasticity(1,1) model or GARCH(1,1) if:

$$a_t = \sigma_t Z_t$$
$$\text{such that } Z_t \overset{\text{iid}}{\sim} \mathcal{N}(0, 1) \text{ and } \sigma_t^2 = \sigma^2 + \alpha_1 a_{t-1}^2 + \beta_1 \sigma_{t-1}^2$$

Running this regression in R, we got the following values:

$$\underset{(t)}{\sigma_t^2} = \underset{(2.416)}{0.00009487} + \underset{(2.996)}{0.06072} a_{t-1}^2 + \underset{(30.988)}{0.8884} \sigma_{t-1}^2$$

What we can take from this is that the variance at time t is not constant; which should be expected from the random walk which we normally think of as a Brownian motion which at time t has variance $t\sigma^2$.
Lastly, we present the estimated autocorrelation function for a_t^2 below. Note that we cut the graph off for legibility; $\rho(0) = 1$. What we notice is that there is a cut-off at $k = 0$ (blue bands represent critical values), i.e the squared residuals could be said to follow an "MA(0)"; that is, there is no relationship between them in terms of an ARMA model; and they are, by definition of the autocorrelation function, uncorrelated.



[3]http://www.investopedia.com/terms/o/octobereffect.asp

8 Appendix: R Code

```r
setwd("C:/Users/Daniel/Dropbox/Spring 2017/STAT 443/Final Project")
#################################
#    STAT 443: Final Project    #
#   Daniel Matheson: 20270871   #
#         Spring 2017           #
#################################
library(tseries)
library(ggplot2)
###############################
####      Question 1      ####
###############################
input.Data <- read.csv("GDP_CONS_CANADA.csv")
w1t <- as.numeric(as.character(input.Data$GDP))
w2t <- as.numeric(as.character(input.Data$CONS))

plot(w1t, xlab = "Time", ylab = "Raw Values",
     ylim = c(min(min(w1t), min(w2t)) - 500, max(max(w1t),max(w2t)) + 500),
     type = "l", col = "blue", lwd = 1)
lines(w2t, col = "red", lwd = 1)
legend(20, 1200000, legend = c("GDP", "CONS"), lty = c(1,1),lwd=c(2.5,2.5),
       col=c("blue","red"), cex = 0.7)
# CONS = Canadian Personal Expenditures on Consumer Goods and Service

###############################
####      Question 2      ####
###############################
xt <- log(w1t)
## Trend Stationary Method
tSeq <- 1:length(xt)
TSmodel <- lm(xt ~ tSeq + 1)
TSyt <- TSmodel$residuals

## Difference Stationary Method
xt.diff <- diff(xt)
DSmodel <- lm(xt.diff ~ 1)
DSyt <- DSmodel$residuals
plot(TSyt, type = "l", xlab = "Time",
     ylab = "Cycle Yt", main = "Difference Stationary Model")
plot(DSyt, type = "l", xlab = "Time",
     ylab = "Cycle Yt", main = "Trend Stationary Model")

###############################
####      Question 3      ####
###############################
BIC <- function(res, k, N) {
  bic = log(sum(res^2) / N)
  bic = bic + log(N) * k / N
  bic}

# DS Method
Q3DSBIC <- rep(0,10)
for (k in 1:10){
  Q3DSmodel <- arima(DSyt, order = c(k,0,0), include.mean = F)
  Q3DS.resid <- Q3DSmodel$residuals
  Q3DSBIC[k] <- BIC(Q3DS.resid, k, length(DSyt))}
```

```
# TS Method
Q3TSBIC <- rep(0,10)
for (k in 1:10){
  Q3TSmodel <- arima(TSyt, order = c(k,0,0), include.mean = F)
  Q3TS.resid <- Q3TSmodel$residuals
  Q3TSBIC[k] <- BIC(Q3TS.resid , k, length(TSyt))}

indices <- 1:10
BICmins <- c(indices[Q3DSBIC == min(Q3DSBIC)], indices[Q3TSBIC == min(Q3TSBIC)])
# DS Method: p = 1 (set to p = 2)
# TS Method: p = 2

# Final Q3 model, TS with p = 2:
TSyt.N<- length(TSyt)
Q3model <-  lm(TSyt[-(1:2)] ~TSyt[-c(1,TSyt.N)] + TSyt[-((TSyt.N-1):TSyt.N)] -1)
Q3phi <- Q3model$coef

# rho(k) and psi_k for k = 0,1,2,3,4,5,6:
Q3rho <- ARMAacf(ar = Q3phi, pacf = F, lag.max = 6)
Q3psi <- rep(NA,7)
Q3psi[1] <- 1
Q3psi[2] <- Q3phi[1] * Q3psi[1]
Q3psi[3] <- Q3phi[1] * Q3psi[2] + Q3phi[2] * Q3psi[1]
for( j in 3:7) {
  Q3psi[j] <- Q3psi[c((j-1),(j-2))] %*% Q3phi[1:2]}
# gamma(0)^{1\2}
Q3.resid <- Q3model$residuals
Q3sigma.sq <- sum((Q3.resid)^2)/length(TSyt)
Q3gamma.zero <- Q3sigma.sq/(1 - Q3phi[1] * Q3rho[2] - Q3phi[2] * Q3rho[3])
Q3gamma.zero.sqrt <- sqrt(Q3gamma.zero)

################################
####        Question 4        ####
################################
# DS Method: under estimation that Y_t ~ AR(2)
DSyt.N = length(DSyt)
Q4DS.model <- lm(DSyt[-(1:2)] ~DSyt[-c(1,DSyt.N)] + DSyt[-((DSyt.N-1):DSyt.N)] -1)
Q4DS.mu <- DSmodel$coefficients["(Intercept)"]
Q4DS.phi <- Q4DS.model$coef
Q4DS.N <- length(xt.diff)
Q4DS.E <- rep(0,9) # E_t[X_t+k], k from 0 to 8
Q4DS.Y <- rep(0,9)  # E_t[Y_t+k], k from 0 to 8
Q4DS.Y[1] <- DSyt[Q4DS.N]
Q4DS.Y[2] <- Q4DS.phi %*% DSyt[c(Q4DS.N,(Q4DS.N-1))] # Y_T+1

for (j in 3:9){
  Q4DS.Y[j] <- Q4DS.phi %*% Q4DS.Y[c(j-1,j-2)]}

Q4DS.E[1] <- xt.diff[Q4DS.N]
Q4DS.E[2:9] <- Q4DS.mu + Q4DS.Y[2:9]

Q4DS.psi <- rep(0,9)
Q4DS.psi[1] <- 1
Q4DS.psi[2] <- Q4DS.phi[1] * Q4DS.psi[1]
for( j in 3:9) {
  Q4DS.psi[j] <- Q4DS.phi %*% Q4DS.psi[c((j-1),(j-2))]}

Q4DS.sigma.sq <- sum((Q4DS.model$residuals)^2)/(Q4DS.N - 2)
Q4DS.var <- rep(0,9) # Var(delta X_t+k) for k = 0 ... 8
```

```r
Q4DS.var[1] <- 0
for (k in 2:9){
Q4DS.var[k] <- Q4DS.sigma.sq * sum((Q4DS.psi[c(1:(k-1))])^2)}

Q4DS.CI.U <- rep(0,9) # 95% confidence interval bounds
Q4DS.CI.L <- rep(0,9)
Q4DS.CI.U <- Q4DS.E + 1.96 * sqrt(Q4DS.var)
Q4DS.CI.L <- Q4DS.E - 1.96 * sqrt(Q4DS.var)
Q4DS.CI <- cbind(round(Q4DS.CI.L,5), round(Q4DS.CI.U,5))
colnames(Q4DS.CI) <- c("Lower Bound", "Upper Bound")

Q4DS.E.vec <- rep(0,20)
Q4DS.E.vec[13:20] <- Q4DS.E[-1]
Q4DS.E.vec[1:12] <- xt.diff[(length(xt.diff)-11):length(xt.diff)]
Q4DS.CI.L.vec <- rep(0,20)
Q4DS.CI.L.vec[1:12] <- rep(NA,12)
Q4DS.CI.L.vec[13:20] <- Q4DS.CI.L[-1]
Q4DS.CI.U.vec <- rep(0,20)
Q4DS.CI.U.vec[1:12] <- rep(NA,12)
Q4DS.CI.U.vec[13:20] <- Q4DS.CI.U[-1]

plot(Q4DS.E.vec, type = "l", ylim = c(min(Q4DS.CI.L)*1.1, max(Q4DS.CI.U)*1.1),
     xlab = "Time (Final Obs. = T = 12)", ylab = "Delta Xt",
     main = "Difference Stationary Delta Xt Forecast")
lines(rep(Q4DS.mu,20), col = "red", lty =3)
lines(Q4DS.CI.L.vec, col = "blue")
lines(Q4DS.CI.U.vec, col = "blue")
abline(v = 13, col = "green", lty = 6)
legend(x = "topleft", legend = c("Delta Xt", "95% CI Bounds", "Delta Xt Mu", "k = 1"),
       lty = c(1,1,3,6),lwd= c(0.5,0.5,0.5,0.5),col= c("black","blue","red", "green"), cex = 0.8,
       seg.len = c(1,1,1,1), text.width = 3.2)

# TS Method
Q4TS.model <- lm(TSyt[-(1:2)] ~TSyt[-c(1,TSyt.N)] + TSyt[-((TSyt.N-1):TSyt.N)] -1)
Q4TS.mu <- TSmodel$coefficients["tSeq"]
Q4TS.phi <- Q4TS.model$coef
Q4TS.N <- length(xt.diff)
Q4TS.E <- rep(0,9)   # E_t[X_t+k], k from 0 to 8
Q4TS.Y <- rep(0,10)  # E_t[Y_t+k], k from -1 to 8

Q4TS.Y[1] <- TSyt[(Q4TS.N)]   #Yt[184]
#Yt[185] : last Yt since length(delta X_t) < length(Yt) for TS:
Q4TS.Y[2] <- TSyt[(Q4TS.N+1)]

for (j in 3:10){
  Q4TS.Y[j] <- Q4TS.phi %*% Q4TS.Y[c((j-1),(j-2))]}

Q4TS.E[1] <- xt.diff[Q4TS.N] # k = 0
Q4TS.E[2:9] <- Q4TS.mu + Q4TS.Y[-c(1,2)] - Q4TS.Y[2:9]
Q4TS.psi <- rep(0,8)
Q4TS.psi[1] <- 1
Q4TS.psi[2] <- Q4TS.phi[1] * Q4TS.psi[1]
Q4TS.psi[3] <- Q4TS.phi[1] * Q4TS.psi[2] + Q4TS.phi[2] * Q4TS.psi[1]
for( j in 3:8) {
  Q4TS.psi[j] <- Q4TS.psi[c((j-1),(j-2))] %*% Q4TS.phi[1:2]}

Q4TS.sigma.sq <- sum((Q4TS.model$residuals^2))/(Q4TS.N-2)
Q4TS.var <- rep(0,9) # Var(delta X_t+k) for k = 1 ... 8
Q4TS.var[1] <- 0
```

```
Q4TS.var[2] <- Q4TS.sigma.sq
for (k in 3:9){
  Q4TS.var[k] <- Q4TS.sigma.sq * sum( (Q4TS.psi[2:(k-1)] - Q4TS.psi[1:(k-2)])^2)}

Q4TS.CI.U <- rep(0,9) # 95% confidence interval bounTS
Q4TS.CI.L <- rep(0,9)
Q4TS.CI.U <- Q4TS.E + 1.96 * sqrt(Q4TS.var)
Q4TS.CI.L <- Q4TS.E - 1.96 * sqrt(Q4TS.var)
Q4TS.CI <- cbind(round(Q4TS.CI.L,5), round(Q4TS.CI.U,5))
colnames(Q4TS.CI) <- c("Lower Bound", "Upper Bound")
Q4TS.E.vec <- rep(0,20)
Q4TS.E.vec[13:20] <- Q4TS.E[-1]
Q4TS.E.vec[1:12] <- xt.diff[(length(xt.diff)-11):length(xt.diff)]
Q4TS.CI.L.vec <- rep(0,20)
Q4TS.CI.L.vec[1:12] <- rep(NA,12)
Q4TS.CI.L.vec[13:20] <- Q4TS.CI.L[-1]
Q4TS.CI.U.vec <- rep(0,20)
Q4TS.CI.U.vec[1:12] <- rep(NA,12)
Q4TS.CI.U.vec[13:20] <- Q4TS.CI.U[-1]
plot(Q4TS.E.vec, type = "l", ylim = c(min(Q4TS.CI.L)*1.1, max(Q4TS.CI.U)*1.1),
     xlab = "Time (Final Obs. = T = 12)", ylab = "Delta Xt",
     main = "Trend Stationary Delta Xt Forecast")
lines(rep(Q4TS.mu,20), col = "red", lty =3)
lines(Q4TS.CI.L.vec, col = "blue")
lines(Q4TS.CI.U.vec, col = "blue")
abline(v = 13, col = "green", lty = 6)
legend(x = "topleft", legend = c("Delta Xt", "95% CI Bounds", "Delta Xt Mu", "k = 1"),
       lty = c(1,1,3,6),lwd= c(0.5,0.5,0.5,0.5),col= c("black","blue","red", "green"),
       cex = 0.8, seg.len = c(1,1,1,1), text.width = 3.2)

#################################
####       Question 5       ####
#################################
# Here the H0 is that the series is difference stationary.
Q5ADF <- adf.test(xt,k=5)
# p = 0.2865
# We do not reject H_0: x_t is difference stationary.

#################################
####       Question 6       ####
#################################
# DS Method
DSkmax <- 20
Q6DSrho <- acf(DSyt, lag.max = DSkmax, plot = F)$acf
Q6DSpsi <- rep(0,DSkmax+1)
Q6DSpsi[1] <- 1
for (k in 1:DSkmax){
  Q6DSmodels <- arima(DSyt, order = c(k,0,0), include.mean = F)
  Q6DSpsi[(k+1)] <- Q6DSmodels$coef[length(Q6DSmodels$coef)]}

# TS Method
TSkmax <- 20
Q6TSrho <- acf(TSyt, lag.max = TSkmax, plot = F)$acf
Q6TSpsi <- rep(0,TSkmax+1)
Q6TSpsi[1] <- 1
for (k in 1:TSkmax){
  Q6TSmodels <- arima(TSyt, order = c(k,0,0), include.mean = F)
  Q6TSpsi[(k+1)] <-  Q6TSmodels$coef[length(Q6TSmodels$coef)]}
```

```
#################################################
# NOTE ON IDENTIFICATION OF CUT-OFF:
# k is ahead by 1 since psi_0/rho(0) are included
# then we take the index previous to the one which
# passes H_0, so this is index (k-2)
# with max(k-2,0) in case the cut-off happens immediately
#################################################
#### Identifying p,q in ARMA(p,q) for DS:
Q6DSp <- 0 # will return 0 if H_0: psi_k = 0 is rejected for every k
for (k in 1:(DSkmax+1)){
  if(abs(Q6DSpsi[k]) < 2/sqrt(length(DSyt))){
    Q6DSp <- max(k-2,0)
    break}}
Q6DSq <- 0 # will return 0 if H_0: rho_k = 0 is rejected for every k
for (k in 1:(DSkmax+1)){
  if(abs(Q6DSrho[k]) < 2/sqrt(length(DSyt))){
    Q6DSq <- max(k-2,0)
    break}}

#### Identifying p,q in ARMA(p,q) for TS:
Q6TSp <- 0 # will return 0 if H_0: psi_k = 0 is rejected for every k
for (k in 1:TSkmax){
  if(abs(Q6TSpsi[k]) < 2/sqrt(length(TSyt))){
    Q6TSp <- max(k-2,0)
    break}}
Q6TSq <- 0 # will return 0 if H_0: rho_k = 0 is rejected for every k
for (k in 1:length(Q6TSrho)){
  if(abs(Q6TSrho[k]) < 2/sqrt(length(TSyt))){
    Q6TSq <- max(k-2,0)
    break}}

Q6matrix <- matrix(c(Q6DSp, Q6TSp, Q6DSq, Q6TSq), nrow = 2, ncol = 2)
rownames(Q6matrix) <- c("DS", "TS")
colnames(Q6matrix) <- c("p","q")
##
#### Recall: rho(k) determines q, psi_k determines p
#### Plots to show cut-off properties:
##
DSplot.x <- rep(0,21)
DSplot.x[2:21] <- 1:20
par(mfrow = c(1,2), mar = c(2,2,1,1))
plot(x = DSplot.x, y =Q6DSpsi, type = "l", main = "Difference Stationary Phi_kk")
abline(h = 0, col = "red", lty = 2)
abline(v = Q6DSp, col = "blue", lty = 2)
abline(h = -1.96*sqrt(1/length(DSyt)), lty = 6, col = "orange")
abline(h = 1.96*sqrt(1/length(DSyt)), lty = 6, col = "orange")
legend(x = "topright", legend = c("Phi_kk", "Cut-off k = 1", "+/- 1.96SE(Phi_kk)"),
       lty = c(1,2,2), col = c("black", "blue", "orange"), cex = 0.7)
plot(x = DSplot.x, y =Q6DSrho, type = "l", main = "Difference Stationary Rho(k)",
     ylim = c(-1.96*sqrt(1/length(DSyt)),max(Q6DSrho)*1.1))
abline(h = 0, col = "red", lty = 2)
abline(v = Q6DSq, col = "blue", lty =2)
abline(h = -1.96*sqrt(1/length(DSyt)), lty = 6,col = "orange")
abline(h = 1.96*sqrt(1/length(DSyt)), lty = 6,col = "orange")
legend(x = "topright", legend = c("Rho(k)", "Cut-off k = 3", "+/- 1.96SE(Rho(k))"),
       lty = c(1,2,2), col = c("black", "blue", "orange"), cex = 0.7)

TSplot.x <- rep(0,21)
TSplot.x[2:21] <- 1:20
```

```
par(mfrow = c(1,2), mar = c(2,2,1,1))
plot(x = TSplot.x , y = Q6TSpsi, type = "l", main = "Trend Stationary Phi_kk")
abline(h = 0, col = "red", lty = 2)
abline(v = Q6TSp, col = "blue", lty = 2)
abline(h = -1.96*sqrt(1/length(TSyt)), lty = 6,col = "orange")
abline(h = 1.96*sqrt(1/length(TSyt)), lty = 6,col = "orange")
legend(x = "topright", legend = c("Phi_kk", "Cut-off k = 2", "+/- 1.96SE(Phi_kk)"),
       lty = c(1,2,2), col = c("black", "blue", "orange"), cex = 0.7)
plot(x = TSplot.x, y= Q6TSrho, type = "l", main = "Trend Stationary Rho(k)", xlab = "k")

#### ARMA models
Q6DS.model1 <- arima(DSyt, order = c(1,0,0), include.mean = F,
                     optim.control = list(maxit = 10000))
Q6DS.model2 <- arima(DSyt, order = c(0,0,3), include.mean = F,
                     optim.control = list(maxit = 10000))
Q6TS.model <- arima(TSyt, order = c(2,0,0), include.mean = F,
                    optim.control = list(maxit = 10000))
# AR(1) DS
Q6DS.resid1 <- Q6DS.model1$residuals
Q6DS.sigmasq1 <- sum((Q6DS.resid1)^2)/length(DSyt)
Q6DS.std.resid1 <- Q6DS.resid1/sqrt(Q6DS.sigmasq1)
# MA(3) DS
Q6DS.resid2 <- Q6DS.model2$residuals
Q6DS.sigmasq2 <- sum((Q6DS.resid2)^2)/length(DSyt)
Q6DS.std.resid2 <- Q6DS.resid2/sqrt(Q6DS.sigmasq2)
Q6TS.resid <- Q6TS.model$residuals
Q6TS.sigmasq <- sum((Q6TS.resid)^2)/length(TSyt)
Q6TS.std.resid <- Q6TS.resid/sqrt(Q6TS.sigmasq)

####### Plots of std. resid
# Means of residuals are extremely close to 0, so ignore when plotting
par(mfrow = c(1,2), mar = c(4,4,2,2))
hist(Q6DS.std.resid1, breaks = 30, freq = F,
     xlab = "Standardized Residuals", main = "Difference Stationary AR(1)",
     cex.main = 0.8)
curve(dnorm(x),col = "red", add = T)
legend(x = "topright", legend = "N(0,1) PDF",
       lty = 1,lwd=0.5,col="red", cex = 0.6,
       seg.len = 0.5, text.width = 1.35, x.intersp = 0.3)
hist(Q6DS.std.resid2, breaks = 30, freq = F,
     xlab = "Standardized Residuals", main = "Difference Stationary MA(3)",
     cex.main = 0.8)
curve(dnorm(x),col = "red", add = T)
legend(x = "topright", legend = "N(0,1) PDF",
       lty = 1,lwd=0.5,col="red", cex = 0.6,
       seg.len = 0.5, text.width =1.35, x.intersp = 0.3)

par(mfrow= c(1,1), mar = c(4,4,2,2))
hist(Q6TS.std.resid, breaks = 30, freq = F,
     xlab = "Standardized Residuals", main = "Trend Stationary AR(2)",
     cex.main = 0.8)
curve(dnorm(x),col = "red", add = T)
legend(x = "topright", legend = "N(0,1) PDF",
       lty = 1,lwd=0.5,col="red", cex = 0.6,
       seg.len = 0.5, text.width =1.35, x.intersp = 0.3)

####### Box Pierce Tests
# DS AR(1)
# sqrt(length(Q6DS.resid1)) = 13.56 -> round to 14
```

```
DS1.BoxPierce <- Box.test(x = Q6DS.resid1, type = "Box-Pierce",lag=14)
# p = 0.005105; reject H_0  -> residuals correlated
# DS MA(3)
DS2.BoxPierce <- Box.test(x = Q6DS.resid2, type = "Box-Pierce",lag=14)
# p = 0.09984; do not reject H_0 -> residuals uncorrelated (but close)
# TS
TS.BoxPierce <- Box.test(x = Q6TS.resid, type = "Box-Pierce",lag=14)
# p = 0.003365; reject H_0 -> residuals correlated


####### Overfitting with r = 4
Q6DS.model1of <- arima(DSyt, order = c(5,0,0), include.mean = F,
                       optim.control = list(maxit = 10000))
Q6DS.model2of <- arima(DSyt, order = c(0,0,7), include.mean = F,
                       optim.control = list(maxit = 10000))
Q6TS.modelof <- arima(TSyt, order = c(6,0,0), include.mean = F,
                      optim.control = list(maxit = 10000))
Q6DS.sigmasq1.of <- Q6DS.model1of$sigma2
Q6DS.sigmasq2.of <- Q6DS.model2of$sigma2
Q6TS.sigmasq.of <- Q6TS.modelof$sigma2
Q6DS.Lambda1 <- (length(DSyt) - 5) * log(Q6DS.sigmasq1/Q6DS.sigmasq1.of)
Q6DS.Lambda2 <- (length(DSyt) - 7) * log(Q6DS.sigmasq2/Q6DS.sigmasq2.of)
Q6TS.Lambda <- (length(TSyt) - 5) * log(Q6TS.sigmasq/Q6TS.sigmasq.of)
of.crit <- qchisq(0.95, 4)

# if true then reject H_0:
Q6DS.of.test1 <- (Q6DS.Lambda1 > of.crit)
Q6DS.of.test2 <- (Q6DS.Lambda2 > of.crit)
Q6TS.of.test <- (Q6TS.Lambda > of.crit)
# all false -> do not reject any of H_0. overfitting fails.

####### 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
##############################################################

# DS AR(1)
DS.resid1 <- Q6DS.resid1
DS.sigma.sq1 <- Q6DS.sigmasq1
DS.Z1 <- (DS.resid1 - mean(DS.resid1))/sqrt(DS.sigma.sq1)
DS.n1 <- length(DS.Z1)
DS.k31 <- 1/DS.n1 * sum((DS.Z1)^3)
DS.k41 <- 1/DS.n1 * sum((DS.Z1)^4)
DS.J1 <- DS.n1 * ( (DS.k31^2)/6 + (DS.k41-3)^2/24)
chisq.crit <- qchisq(0.95,2)
DS.J.test1 <- (DS.J1 > chisq.crit) # if true, reject H_0

# DS MA(3)
DS.resid2 <- Q6DS.resid2
DS.sigma.sq2 <- Q6DS.sigmasq2
DS.Z2 <- (DS.resid2 - mean(DS.resid2))/sqrt(DS.sigma.sq2)
DS.n2 <- length(DS.Z2)
DS.k32 <- 1/DS.n2 * sum((DS.Z2)^3)
```

```
DS.k42 <- 1/DS.n2 * sum((DS.Z2)^4)
DS.J2 <- DS.n2 * ( (DS.k32^2)/6 + (DS.k42-3)^2/24)
DS.J.test2 <- (DS.J2 > chisq.crit) # if true, reject H_0

# TS
TS.resid <- Q6TS.resid
TS.sigma.sq <- Q6TS.sigmasq
TS.Z <- (TS.resid - mean(TS.resid))/sqrt(TS.sigma.sq)
TS.n <- length(TS.Z)
TS.k3 <- 1/TS.n * sum((TS.Z)^3)
TS.k4 <- 1/TS.n * sum((TS.Z)^4)
TS.J <- TS.n * ( (TS.k3^2/6) + (TS.k4-3)^2/24)
TS.J.test <- (TS.J > chisq.crit) # if true, reject H_0

####### ARCH(6) test for independence of resid
# DS AR(1)
ARCH.q <- 6
N1 <- length(Q6DS.resid1)
Q6DS.resid1 <- Q6DS.resid1^2
ARCH.DS.model1 <- lm(Q6DS.resid1[-(1:6)]~Q6DS.resid1[-c((1:5), N1)]
                     +  Q6DS.resid1[-c(1:4,(N1-1),N1)] + Q6DS.resid1[-c(1:3, (N1-2):N1)]
                     + Q6DS.resid1[-c(1:2, (N1-3):N1)] + Q6DS.resid1[-c(1, (N1-4):N1)]
                     + Q6DS.resid1[-((N1-5):N1)] - 1)
DS.Rsq1 <- summary(ARCH.DS.model1)$r.squared
DS.ARCH.test1 <- N1 * DS.Rsq1
DS.ARCH.crit <- qchisq(0.95,6)
DS.ARCH.H01 <- (DS.ARCH.test1 > DS.ARCH.crit) # if true reject H_0

# DS MA(3)
N2 <- length(Q6DS.resid2)
Q6DS.resid2 <- Q6DS.resid2^2
ARCH.DS.model2 <- lm(Q6DS.resid2[-(1:6)]~Q6DS.resid2[-c((1:5), N2)]
                     + Q6DS.resid2[-c(1:4,(N2-1),N2)] + Q6DS.resid2[-c(1:3, (N2-2):N2)]
                     + Q6DS.resid2[-c(1:2, (N2-3):N2)] + Q6DS.resid2[-c(1, (N2-4):N2)]
                     + Q6DS.resid2[-((N2-5):N2)] - 1)
DS.Rsq2 <- summary(ARCH.DS.model2)$r.squared
DS.ARCH.test2 <- N2 * DS.Rsq2
DS.ARCH.crit <- qchisq(0.95,6)
DS.ARCH.H02 <- (DS.ARCH.test2 > DS.ARCH.crit) # if true reject H_0

# TS
N3 <- length(Q6TS.resid)
Q6TS.resid <- Q6TS.resid^2
ARCH.TS.model <- lm(Q6TS.resid[-(1:6)]~Q6TS.resid[-c((1:5), N3)]
                    + Q6TS.resid[-c(1:4,(N3-1),N3)] + Q6TS.resid[-c(1:3, (N3-2):N3)]
                    + Q6TS.resid[-c(1:2, (N3-3):N3)] + Q6TS.resid[-c(1, (N3-4):N3)]
                    + Q6TS.resid[-((N3-5):N3)] - 1)
TS.Rsq <- summary(ARCH.TS.model)$r.squared
TS.ARCH.test <- N3 * TS.Rsq
TS.ARCH.crit <- qchisq(0.95,6)
TS.ARCH.H0 <- (TS.ARCH.test > TS.ARCH.crit) # if true reject H_0

################################
####       Question 8       ####
################################
Q8data <- read.csv("SP_data_for_Q8.csv")
P <- Q8data$P
lnP <- log(P)
N <- length(lnP)
```

```r
Q8.model <- lm(lnP[-1] ~ 1 + lnP[-N])
at <- Q8.model$residuals
at.N <- length(at)

at.rho <- acf(at, lag.max = at.N, plot = F)$acf
at.psi <- rep(0,10)
for (k in 1:10){
  at.models <- arima(at, order = c(k,0,0), include.mean = F)
  at.psi[k] <-  at.models$coef[k]}
# sqrt(at.N) ~ 25
at.BoxPierce <- Box.test(x = at, type = "Box-Pierce",lag=25)
# p-value = 0.6529; do not reject H_0; uncorrelated

## Checking if a_t is normally distributed:
at.sigmasq <- sum((Q8.model$residuals)^2)/(at.N-1)
at.std <- (at - mean(at))/sqrt(at.sigmasq)

# Jarque-Bera Test for Normality
at.k3 <- 1/at.N * sum((at.std)^3)
at.k4 <- 1/at.N * sum((at.std)^4)
at.J <- at.N * ( (at.k3^2/6) + (at.k4-3)^2/24)
chisq.crit <- qchisq(0.95,2)


at.J.test <- (at.J > chisq.crit) # if true, reject H_0
# true -> reject H_0 -> residuals not normal, or there are outliers:
# at.J = 547

# outliers are indices 16, 585:
ato.std <- at.std[-c(16,585)]
ato.N <- at.N - 2

# plots
par(mfrow= c(1,2), mar=c(4,4,2,2))
hist(at.std, breaks = 50, freq = F,
     xlab = "Standardized Residuals", main = expression("Standardized Residuals a"[t]))
curve(dnorm(x),col = "red", add = T)
legend(x = "topleft", legend = "N(0,1) PDF",
       lty = 1,lwd=0.5,col="red", cex = 0.6,
       seg.len = 0.5, text.width = 3)
hist(ato.std, breaks = 50, freq = F,
     xlab = "Standardized Residuals", main = expression("2 Outliers Removed"))
curve(dnorm(x),col = "red", add = T)
# looks normal

# JB test without outliers:
ato.k3 <- 1/ato.N * sum((ato.std)^3)
ato.k4 <- 1/ato.N * sum((ato.std)^4)
ato.J <- ato.N * ( (ato.k3^2/6) + (ato.k4-3)^2/24)

# if true, reject H_0
ato.J.test <- (ato.J > chisq.crit) # value 1.398
# false: do not reject H_0, residuals are normally distributed
# when only 2 of 647 observations removed
# -> will assume normal then

# autocorrelation function for a_t^2:
at.sq <- at^2
par(mfrow = c(1,1), mar = c(4,4,4,2))
```

```
at.sq.rho <- acf(at.sq, lag.max = 20, plot = T,
                 main = expression("Autocorrelation Function of squared a"[t]))$acf
# none of them are significant other than rho(0)
# remove rho(0) to fit into a smaller space and for visibility

### Fitting to GARCH(1,1):
require(fGarch)
# Here Yt is the appropriate time series you are trying to fit.
at.GARCH <- garchFit(formula = ~garch(1, 1), data = at)
```