Patrick Kurkiewicz
VEE Time Series
Summer 2014 Session

R code is in the appendix.

Introduction

The 10-Yr Treasury bond is an important lending benchmark and measure of economic conditions. A Treasury is a portion of government debt that is sold on the market. Various factors can affect the price of a Treasury (which is inversely related to the yield). For example, a government with poor creditworthiness will have lower priced treasuries. Periods of high inflation or economic growth are correlated with higher Treasury yields as investors are more likely to invest in securities that can potentially provide a greater rate of return. The 10-yr Treasury is also a benchmark for mortgage lending rates.

Following is a search for a pattern in the monthly average yield of the 10-year Treasury from 1974 through 2014. Data was obtained through the US Department of the Treasury's website.

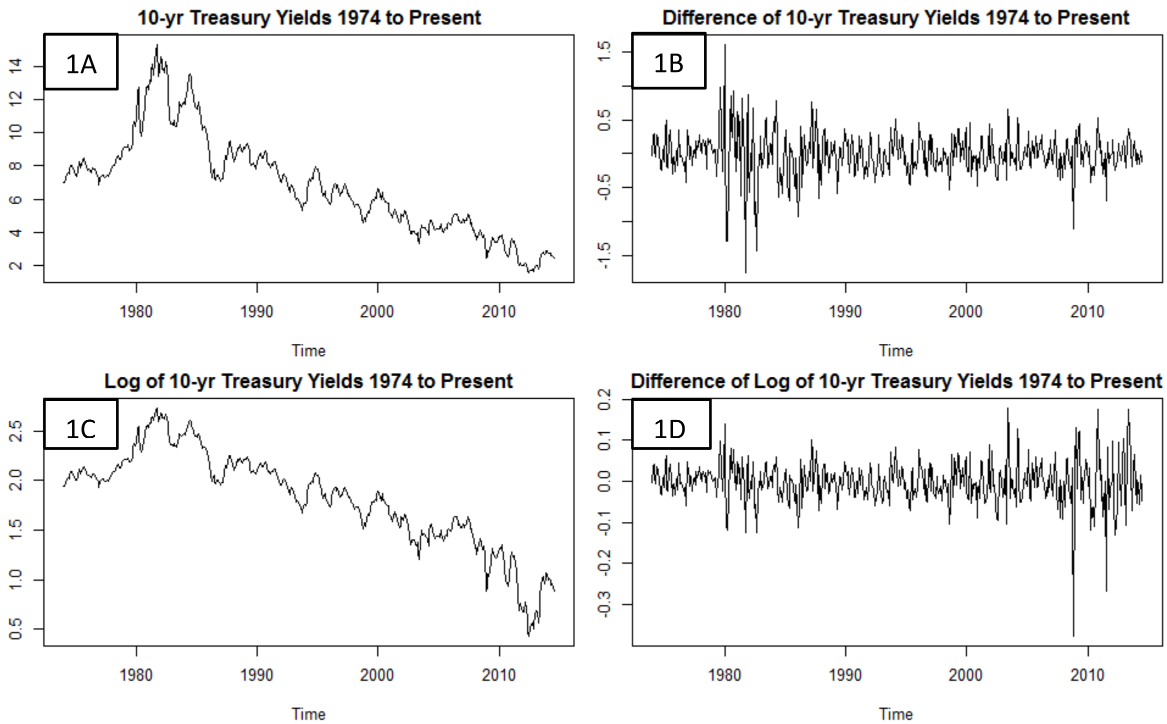
Exploration of the Data

Surely there are some complex predictors and dependencies that underlie the 10-yr Treasury yield stochastic process seen in figure 1A. An ideal model might include variables for seasonal patterns as well as regression parameters for various events that have an effect on the yield. However, for the illustration of the methods learned in this course, I am going to stick to a stationary time series model.

A stationary time series model, however, should not be used on a process that does not appear to be stationary. A stationary process is one where "the probability laws that govern its behavior do not change over time" (Cryer & Chan, 2008). From the history of Treasury yields pictured in 1a, it appears that the process that governed the yields between the mid 70's and mid 80's was different from the process that has governed yields since 2010. Additionally, there appears to be a descending trend with time. Differencing the monthly values results in a more stationary time series (figure 1b), but it appears that when the prices were higher, the monthly differences were greater. The taking the log of the time series (figure 1c) results in a process that is more stable than the original series between 1980 and 2000, when prices were higher, but less stable since because of the log operator's increased sensitivity to numbers close to 0. Also the series, like that in 1a has a descending trend.

The differenced log transformation (figure 1d) of the Treasury yield was used for analysis because it appears the most stationary of them all. There does however, appear to be a number of extreme values following the "financial meltdown" of 2008. Since then, the oscillations in the plot have been larger than at any other time. This is likely partly due to the sensitivity of the log function at numbers closer to

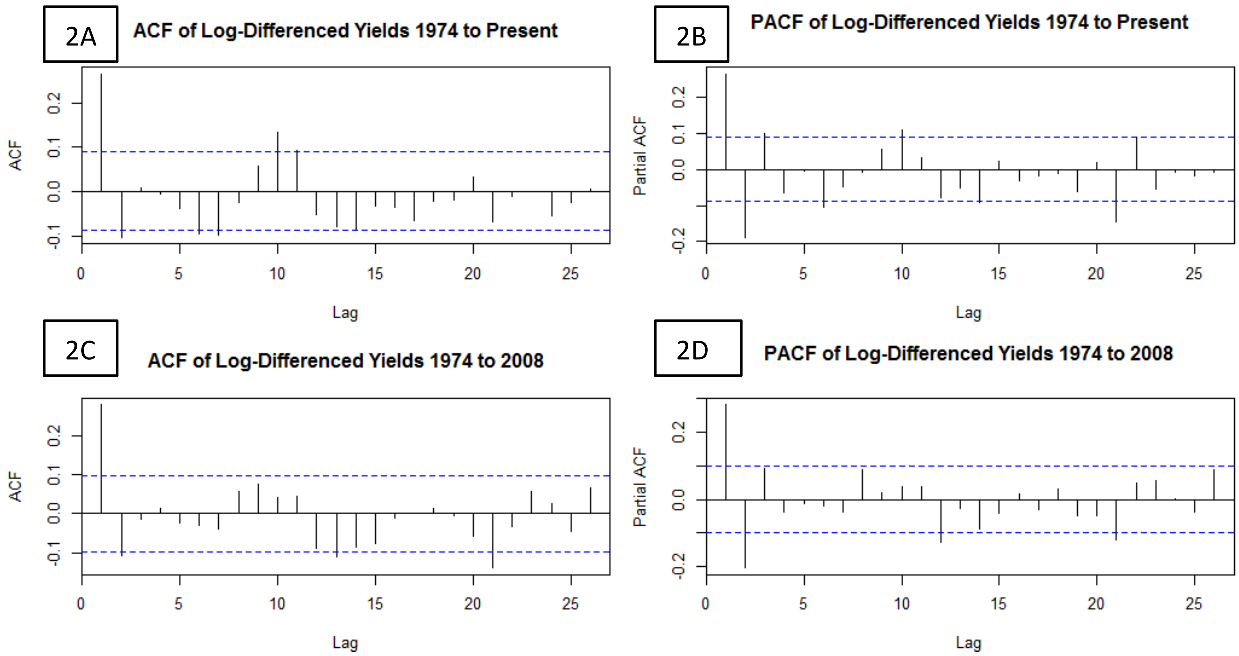
0, which is where the Treasury yield has been these last few years. For this reason, we will explore the effect of removing post-2008 data from the model building process.



Searching for an Appropriate Model

The plots of the autocorrelation functions (figures 2A and 2C) indicate that the ARIMA series has a highly significant positive first-order moving-average parameter and possibly a negative second-order parameter. It should also be noted that an increase in the values of the ACF is seen at a lag of approximately 12 months, indicating a possible yearly seasonal moving average pattern.

The PACF plots (figures 2B and 2D) also indicate a highly significant positive first-order autoregressive parameter, a highly significant negative second-order auto-regressive parameter, and possibly a positive third-order auto-regressive parameter. As with in ACF plots, we see evidence of possible yearly seasonality with a spike in the PACF values at 12 months. The spike at around month 12 in figure 2D indicates that there is a possible negative correlation of the Treasury yield in a given month year-over-year. Because the autocorrelations and partial-autocorrelations taper off more rapidly when the post-2008 data is omitted, pre-2008 data will be used for all subsequent analyses. It is possible that a separate model should be used for data following 2008.



Because a mixed AR/MA model's ACF and PACF plots can "theoretically have many non-zero values" (Cryer & Chan, 2008), observing an EACF plot can be illustrative. An EACF plot separates the autoregressive portion of the series from the moving average portion by assuming a k^{th} order autoregressive model and marking an "X" where the j^{th} order autocorrelation is significantly greater than zero (as in figure 3). The ACF plot of a theoretical j-order moving-average model has significant autocorrelations up to lag j and autocorrelations near 0 for lags greater than j. Generally the 0's form a triangular pattern, from which the uppermost, left most value is commonly chosen for the sake of parsimony. The sample EACF plot in figure 3 has a somewhat triangular pattern. Based on the ACF and PACF plots, I am inclined to choose a ARIMA(1,1,2) model for the log-differenced yields. The EACF plot, however, is not very effective for spotting seasonal parameters.

AR/MA	0	1	2	3	4	5	6	7	8	9	10	11	12	13
0	x	x	o	o	o	o	o	o	o	o	o	o	x	o
1	x	x	o	o	o	o	o	o	o	o	o	o	o	o
2	x	o	o	o	o	o	o	o	o	o	o	o	o	o
3	x	o	x	o	o	o	o	o	o	o	o	o	o	o
4	x	x	o	o	o	o	o	o	o	o	o	o	o	o
5	x	o	o	o	x	o	o	o	o	o	o	o	o	o
6	x	o	x	x	o	o	o	o	o	o	o	o	o	o
7	x	x	x	x	x	o	o	o	o	o	o	o	o	o
8	x	x	x	x	x	o	o	o	o	o	o	o	o	o
9	x	x	o	x	x	x	o	o	o	o	o	o	o	o
10	x	o	o	x	x	o	o	o	o	x	o	o	o	o
11	x	x	x	x	o	o	o	o	x	x	o	o	o	o
12	x	x	o	x	x	o	x	o	x	x	o	o	o	o
13	x	x	o	x	x	o	x	o	o	x	x	o	o	o

Figure 3: The EACF of the log-differenced Treasury yields from 1974 to 2008.

According to the textbook, the Bayesian Information Criterion (BIC) is a consistent estimator of the order of ARMA(p,q) models. This means that as sample size increases, so does the probability that the lowest BIC values correspond to models similar to the true underlying process, given that an underlying ARMA process is present. If the process is not truly an ARIMA process, or not a pure ARIMA process (as is likely true with the Treasury yield data), the BIC is still useful in selecting a model that explains much of the variation in the data.

Figure 4 contains a ranking of BIC values (lower = better fitting model) among models with up to 13 AR and 13 MA parameters. Interestingly, the best models contain a lag-12 autoregressive parameter and a lag 11 moving average parameter.

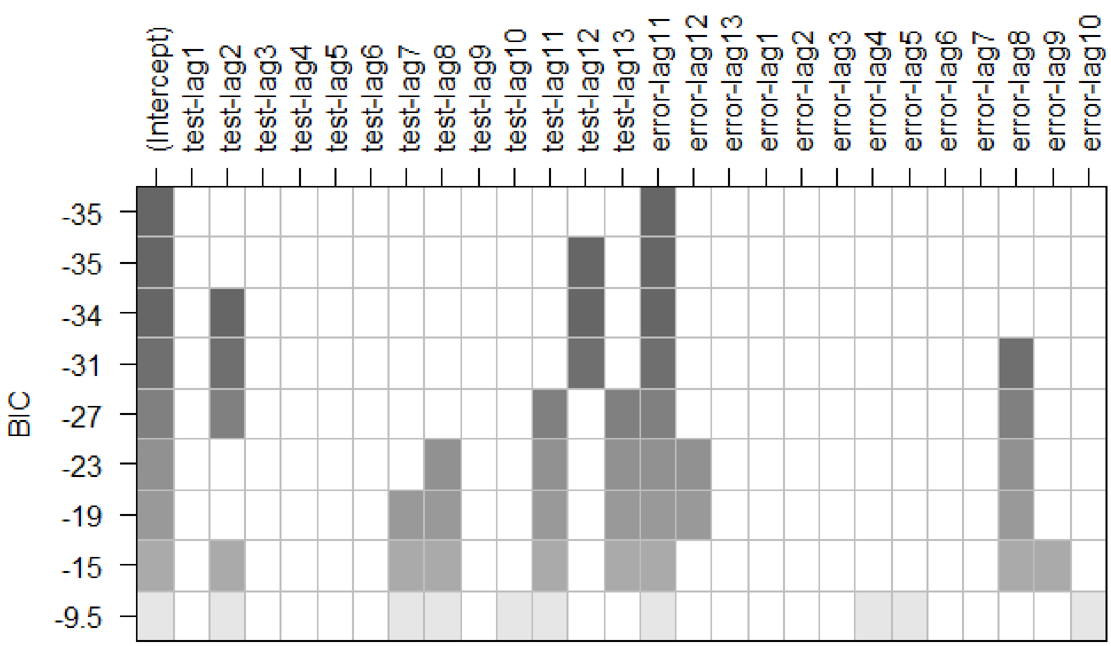


Figure 4: BIC Values for Various ARIMA subsets for fitting the log-differenced yield data from 1974 to 2008.

Because a lag 11 parameter does not really make sense for this data, another ranking of BIC parameters was generated for models containing up to 13 AR parameters and 6 MA parameters (figure 5). Here we see that the two best fitting models in terms of BIC have a second-order auto-regressive parameter and a first-order moving average parameter. The third best model (not much worse-fitting than the two best models) includes a year-over-year seasonal autoregressive parameter and first and second-order moving average parameters.

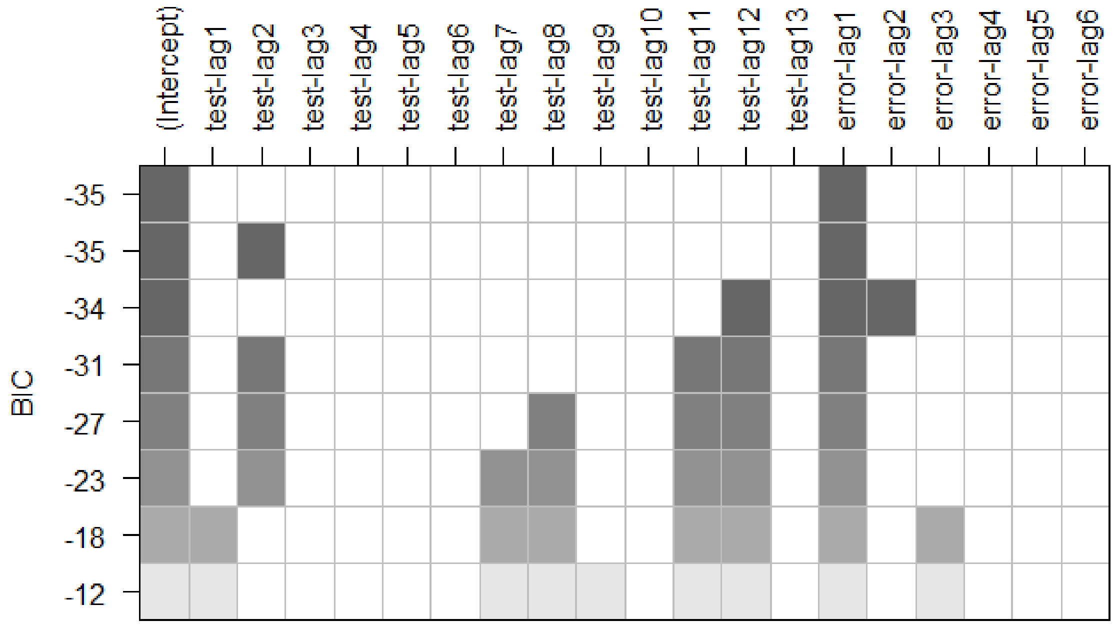


Figure 5: BIC Values for Various ARIMA subsets for fitting the log-differenced yield data from 1974 to 2008 (limited to 13 AR and 6 MA parameters).

Parameter Estimation for the Top Models

Since all the top models in terms of BIC contain a first-order MA parameter and a majority of the top models contain a seasonal autoregressive parameter models including such parameters were fit. Recall that the inclusion of a seasonal AR parameter and lower-order MA parameters was also supported by the ACF and PACF plots.

Three models were fit (table 1). Model 1 is a ARIMA $(0,1,2)x(1,1,0)_{12}$ model. It includes first and second order moving average parameters and a 12-month seasonal autoregressive parameter. The first order moving average parameter indicates a positive autocorrelation at lag 1 and a negative, and weaker, autocorrelation at lag 2. The most significant parameter is the negative seasonal autoregressive parameter, which suggests a moderately strong negative relationship of the log of first differences of Treasury yields from one year to the next. This corresponds to the significant negative value seen at lag 12 in the PACF plot.

The large standard error relative to the parameter estimate for the second order moving average parameter suggests that the true value of the parameter may be close to zero, if not zero. We see in model 2 that the removal of the second order moving average parameter has a small effect on the estimate of the first order moving average parameter. This model also has the lowest AIC of the three (1296 vs 1300 for model 1 and 1299 for model 3). For these reasons, the model was selected over model 1 as the best model for the log-differenced Treasury yields.

A third model was fit, which included a first-order autoregressive parameter. Contrary to the PACF plot the parameter is negative. The addition of the parameter also has a large effect on the value of the moving average parameter, which I suspect is due to a high correlation of the parameters with one another. For these reasons, model 3 was rejected.

Table 1: Three Fitted ARIMA Seasonal Models

Model 1: ARIMA (0,1,2)x(1,1,0)$_{12}$	**θ_1**	**θ_2**	**ϕ_1**	**Φ**
estimate	0.376	-0.125	x	-0.561
s.e.	0.050	0.049	x	0.041
Model 2: ARIMA (0,1,1)x(1,1,0)$_{12}$	**θ_1**	**θ_2**	**ϕ_1**	**Φ**
estimate	0.439	x	x	-0.557
s.e.	0.051	x	x	0.041
Model 3: ARIMA (1,1,1)x(1,1,0)$_{12}$	**θ_1**	**θ_2**	**ϕ_1**	**Φ**
estimate	0.616	x	-0.227	-0.560
s.e.	0.078	x	0.098	0.041

Residual Analysis

A Q-Q plot of the selected model's residuals (figure 6) indicates that there are more extreme error values than would be expected under normality. While it is certainly possible that the residuals do not follow a normal distribution, it is also possible that a better fitting model may reduce the number of extreme residuals.

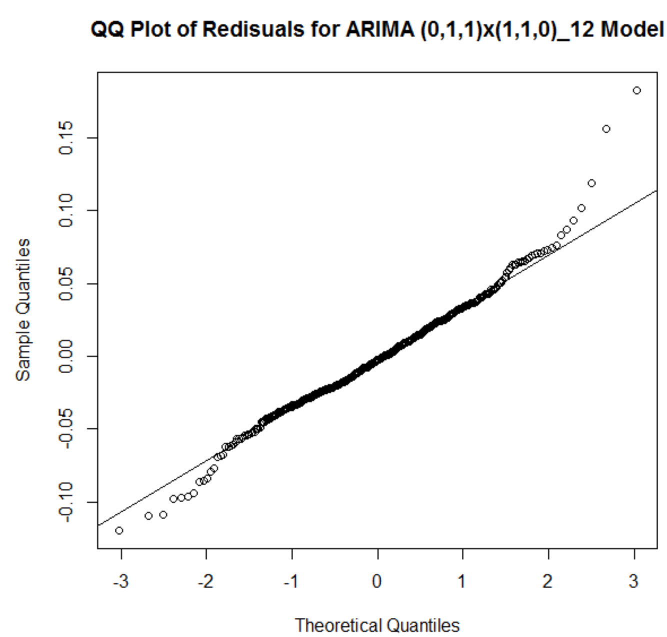


Figure 6: QQ Plot of residuals for selected model.

The PACF and ACF plots (figure 7) of the residuals for model 2 do not provide sufficient evidence to suggest any other autoregressive or moving average parameters should be added to the model. While there is a spike at lag 21, one would be hard pressed to come up with a justification for adding a lag 21 autoregressive or moving average parameter to the model. It is also the only spike outside of the confidence bounds, which do not make any adjustment for sample size.

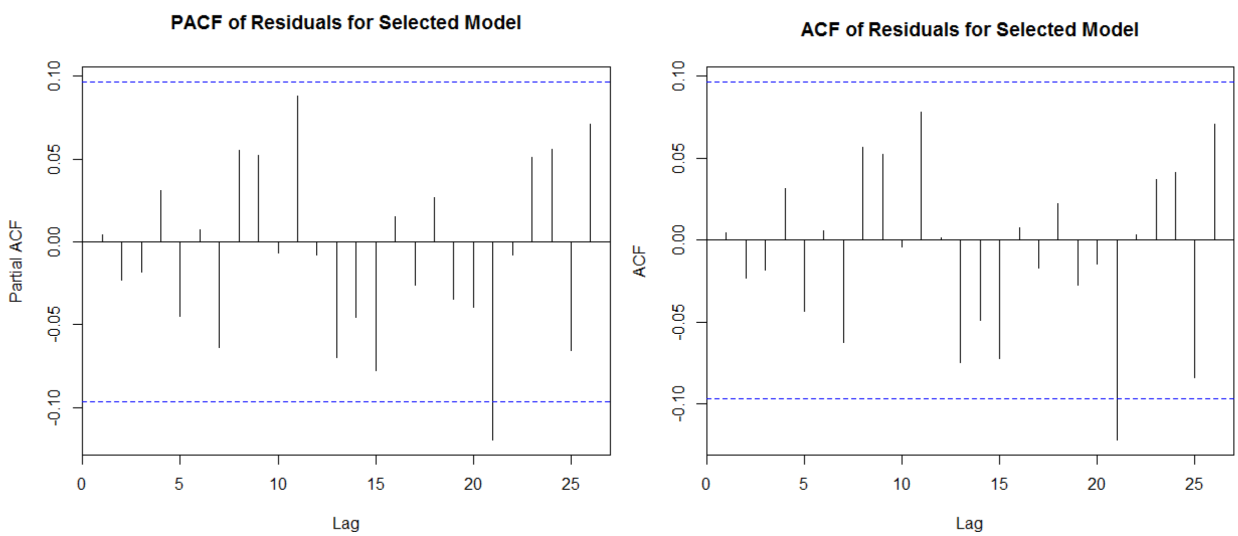


Figure 7: ACF and PACF plots of residuals for the selected model.

<u>Conclusion</u>

The best purely ARIMA model that could be fit was a $\mathbf{(0,1,1)x(1,1,0)_{12}}$ seasonal ARIMA model following differencing and a log-transformation.   This model was fit merely as an exercise in ARIMA modeling. A better model for the data would likely include regression parameters that capture factors that directly or indirectly affect the yield on the 10-yr Treasury bond. ARIMA modeling like that done in this paper can be a useful tool for generating quick, ad hoc analyses/searches for patterns when an extensive search for predictors may not feasible.

Bibliography

Cryer, J., & Chan, K.-S. (2008). *Time Series Analysis with Applications in R 2nd Edition.* Iowa City: Springer.

US Department of the Treasury. (2014, 10 01). *Daily Treasury Yield Curve Rates*. Retrieved October 1, 2014, from US Department of the Treasury: http://www.Treasury.gov/resource-center/data-chart-center/interest-rates

Appendix

```
#READING IN THE DATA
dat = read.csv('FinalProj.csv')
int=dat[1:488, ]

save(int, file="int.rdata")
load("int.rdata")

int$DATE=as.Date(int$DATE,"%m/%d/%Y")
```

##EXPLORING THE DATA
#PLOTS OF TRANSFORMATIONS FOR STATIONARITY

```
dev.new()
par(mfrow = c(2, 2))
par(mar=c(4,2,2,1))
plot(VALUE~DATE,data=int,type="l",xlab="Time",
ylab="Monthly Avg. Interest Rate",main="10-yr Treasury Yields 1974 to Present" )
plot(diff(VALUE)~DATE[1:(length(DATE)-1)],data=int,type="l",xlab="Time",
ylab="Monthly Avg. Interest Rate",main="Difference of 10-yr Treasury Yields 1974 to Present" )
plot(log(VALUE)~DATE,data=int,type="l",xlab="Time",
ylab="Monthly Avg. Interest Rate",main="Log of 10-yr Treasury Yields 1974 to Present" )
plot(diff(log(VALUE))~DATE[1:(length(DATE)-1)],data=int,type="l",xlab="Time",
ylab="Monthly Avg. Interest Rate",main="Difference of Log of 10-yr Treasury Yields 1974 to Present" )
```

##SEARCHING FOR AN APPROPRIATE MODEL
#ACF AND PACF PLOTS

```
dev.new()
par(mfrow = c(2, 2))
par(mar=c(4,4,4,1))
acf(diff(log(int$VALUE)), ci.type= 'ma', plot=TRUE, main = 'ACF of Log-Differenced Yields 1974 to
Present')
pacf(diff(log(int$VALUE)), plot=TRUE, main = 'PACF of Log-Differenced Yields 1974 to Present')
acf(diff(log(int$VALUE[1:410])), ci.type= 'ma',  plot=TRUE, main = 'ACF of Log-Differenced Yields 1974 to
2008')
pacf(diff(log(int$VALUE[1:410])), plot=TRUE, main = 'PACF of Log-Differenced Yields 1974 to 2008')
```

#EACF PLOT

```
eacf(diff(log(int$VALUE[1:410])),ar.max=13,ma.max=13)
```

#BIC PLOTS

```
dev.new()
res=armasubsets(y=diff(log(int$VALUE[1:410])),nar=13,nma=13,y.name='test',ar.method='ols')
plot(res)
dev.new()
res=armasubsets(y=diff(log(int$VALUE[1:410])),nar=13,nma=6,y.name='test',ar.method='ols')
plot(res)
```

```
#PARAMETER ESTIMATION
arima(log(int$VALUE[1:410]),order=c(0,1,2),seasonal=list(order=c(1,1,0),period=12))
arima(log(int$VALUE[1:410]),order=c(0,1,1),seasonal=list(order=c(1,1,0),period=12))
arima(log(int$VALUE[1:410]),order=c(1,1,1),seasonal=list(order=c(1,1,0),period=12))

#RESIDUAL ANALYSIS
dev.new()
model=arima(log(int$VALUE[1:410]),order=c(0,1,1),seasonal=list(order=c(1,0,0),period=12))
qqnorm(residuals(model),main='QQ Plot of Redisuals for ARIMA (0,1,1)x(1,1,0)_12 Model')
qqline(residuals(model))

dev.new()
par(mfrow = c(1, 2))
par(mar=c(4,4,4,1))
pacf(residuals(model), ci.type='ma',main = 'PACF of Residuals for Selected Model')
acf(residuals(model),ci.type='ma', main = 'ACF of Residuals for Selected Model')
```