

Personal Savings Analysis

Andrew Brown, Melissa Hooke, Frances Hung, Mai Nguyen, Brenner Ryan

12/19/2018

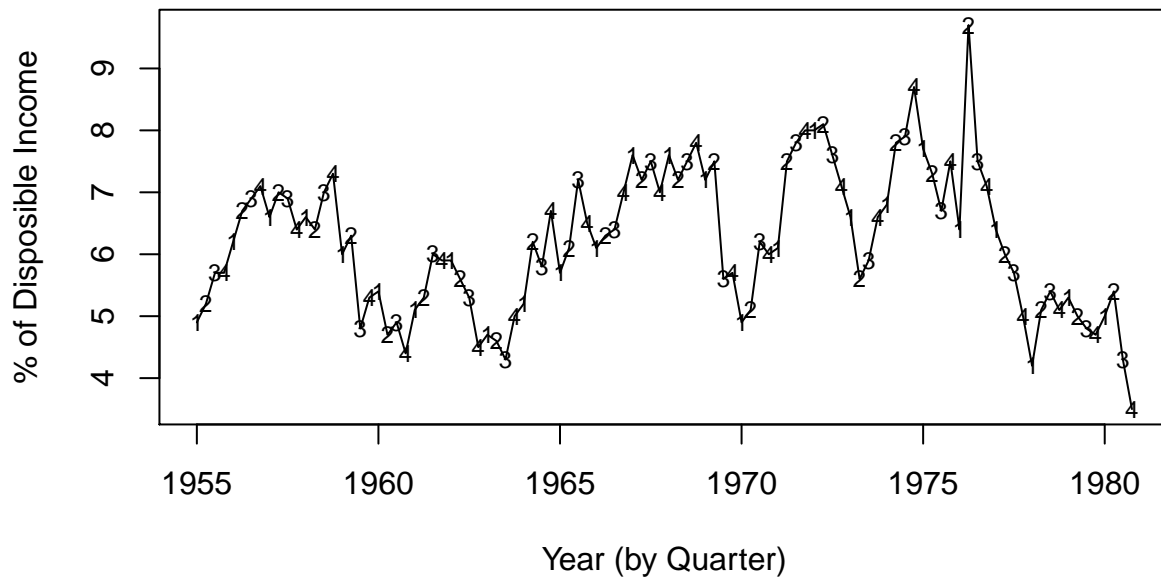
Abstract: This paper explores personal savings data in the United States from 1955 - 1980; a span of 26 years. We first examined the trend of our data by using various time series analysis tools such as model specification, model fitting and diagnostics of our fitted models. Through these in-depth analyses, we found three candidate models: $AR(1)$, $AR(1) \times AR(1)_4$, and $ARIMA(0, 1, 0) \times ARIMA(1, 0, 1)_6$. Further examination concluded that $AR(1) \times AR(1)_4$ is the most fitting model for the data set. With this model, we are able to form a greater understanding of personal savings data, gain insights into the U.S.'s economic performance and perform predictions on future savings rates.

Introduction

Time Series Exploration

The original time series, as pulled from DataMarket (<https://datamarket.com/>), spans 26 years of personal savings as percent of disposable income in the United States. Each year of the time series is divided into financial quarters, amounting to 104 total observations, which we plot in the time series below:

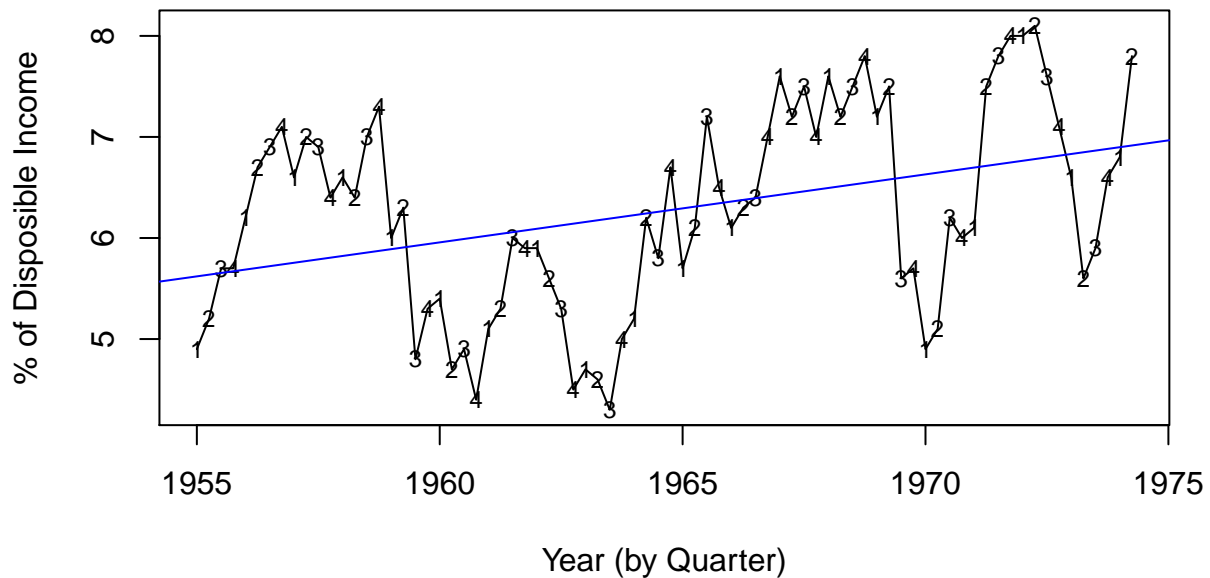
Time Series of Personal Savings in the US (1955–1980)



From our original time series plot, we see that there is a general upward trend with some sudden volatility in the 1970s that may be linked to the economic crash in the early 70s and oil energy crisis in 1979. Since the 1970s were marked by high inflation and growing expenses due to rising interest rates, we made the decision to remove the last 5 years of the time series since they would likely follow a different time series trend than the rest of the data.

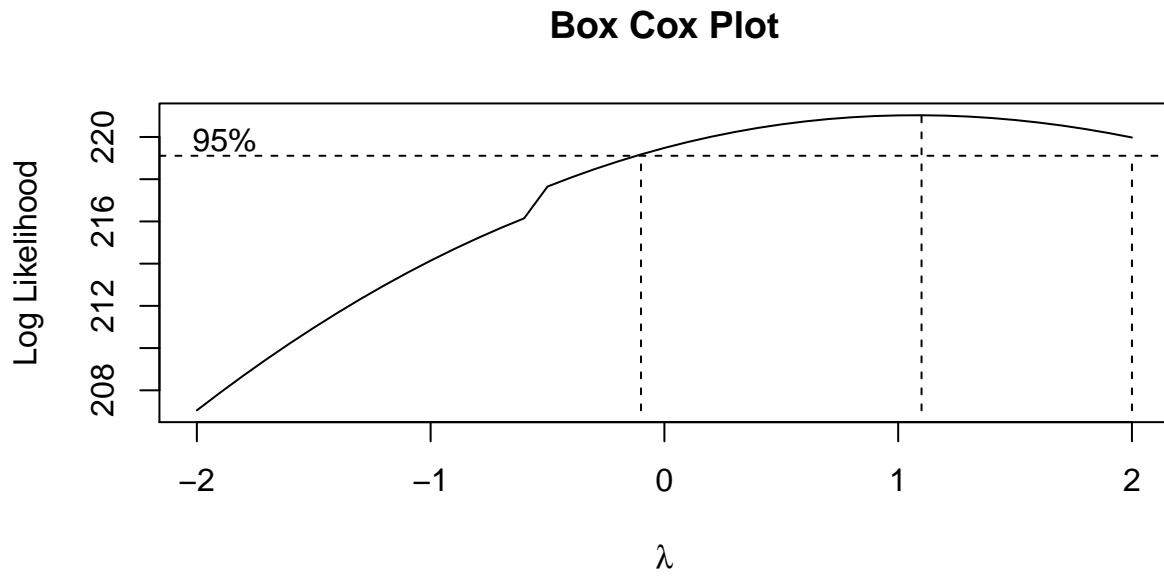
In addition, we set aside the last 6 observations in order to use them as test points to compare with our forecasts at the end of our analysis. The resulting series of 78 observations is plotted below:

Time Series of Personal Savings in the US (1955–1974)

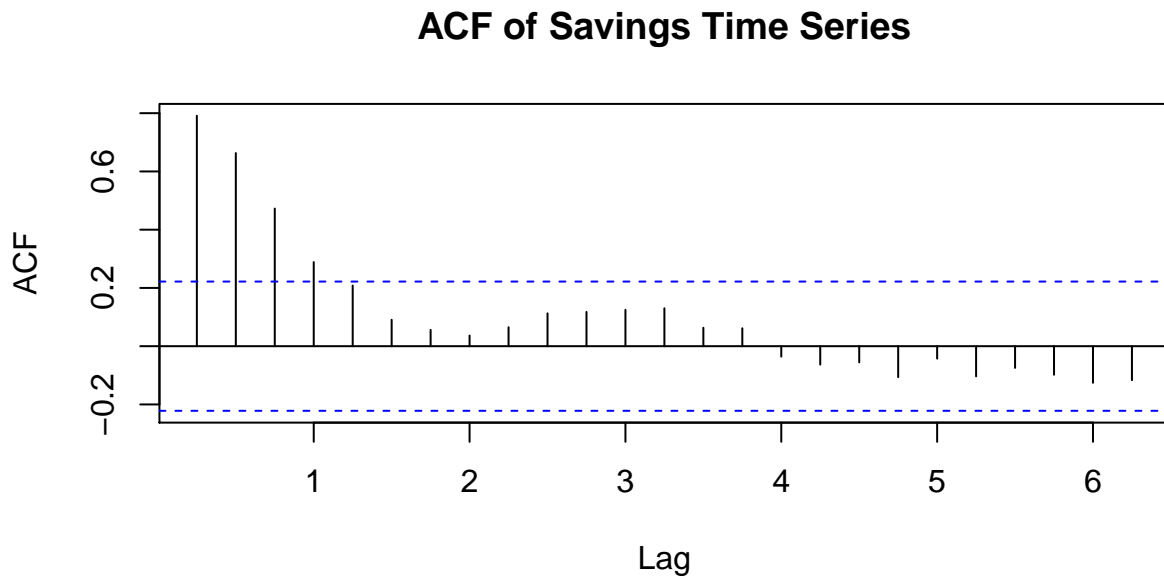


In the time series, we see a general upward trend in the data, which indicates that the time series may not be stationary and we may want to consider taking the first difference of the data. Also, while do not see any *obvious* seasonal trends, given that the data is divided into financial quarters we may want to consider the possibility of taking a seasonal difference to make our time series stationary. First, however, let's explore without taking the difference.

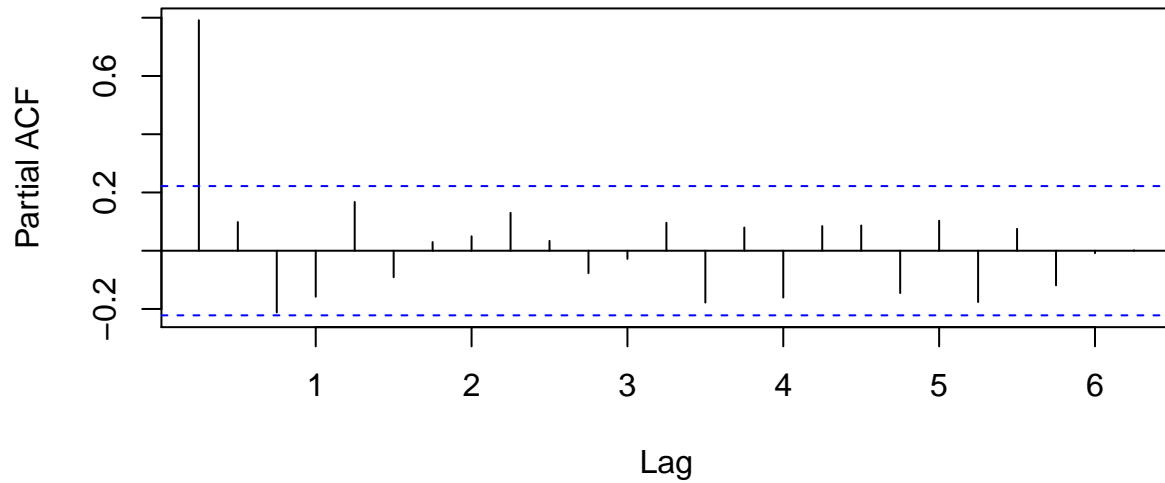
The first step in analyzing our time series is to consider the possible need for a transformation to stabilize the variance of the series over time. In order to do this, we use the function `BoxCox.AR` to determine the appropriate power transformation for time-series data.



The Boxcox output indicates that a transformation is not necessary in order to stabilize the variance since λ is about equal to 1. Therefore, we proceed by examining the acf and pacf of the series.



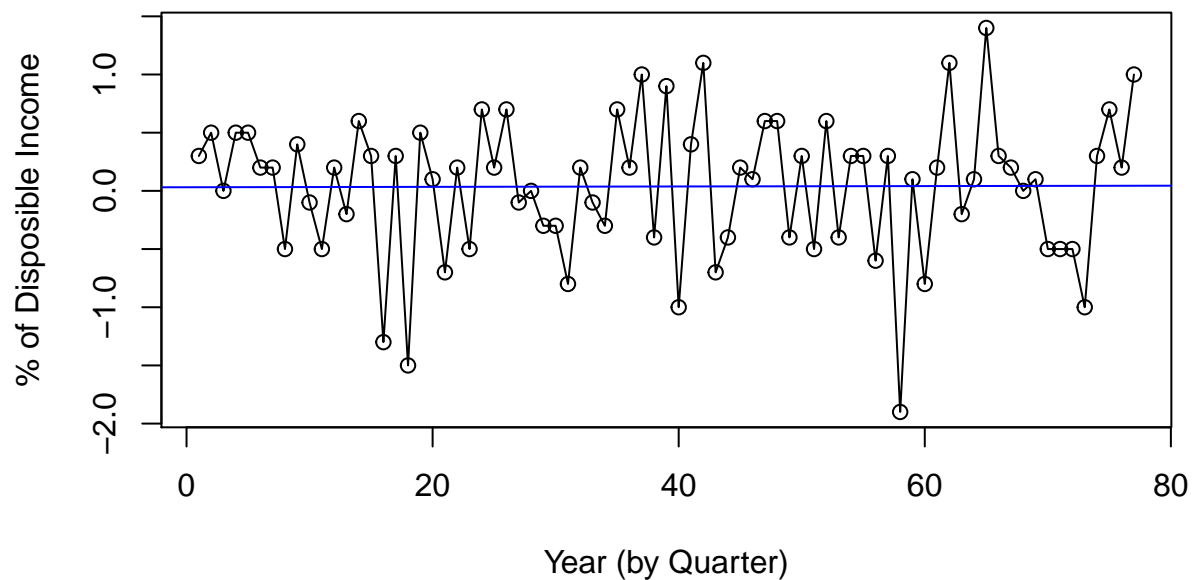
PACF of Savings Time Series



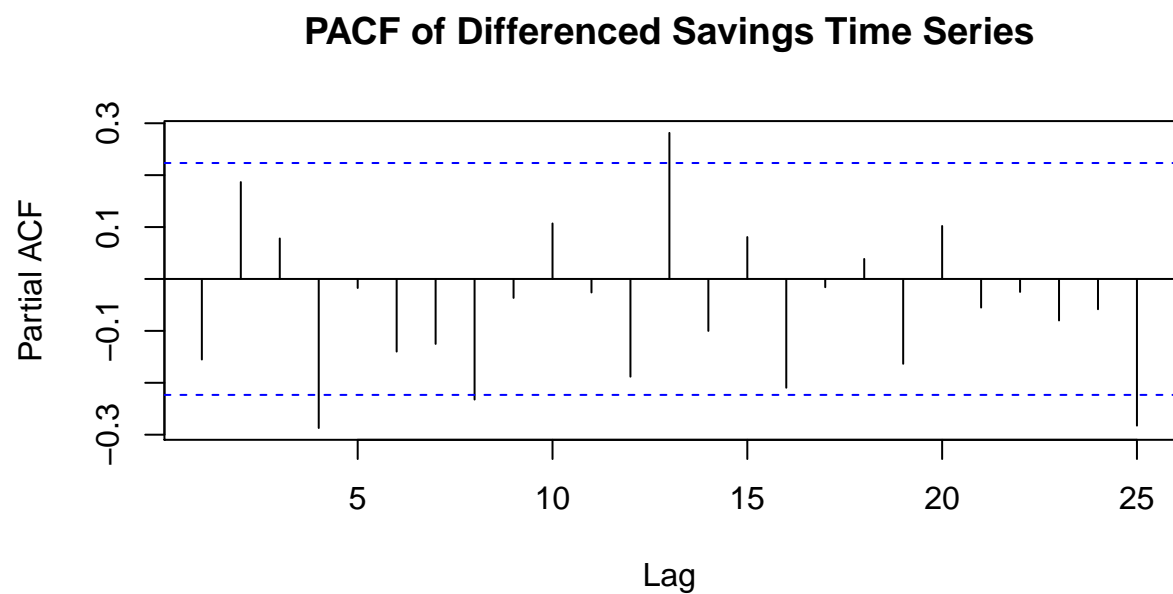
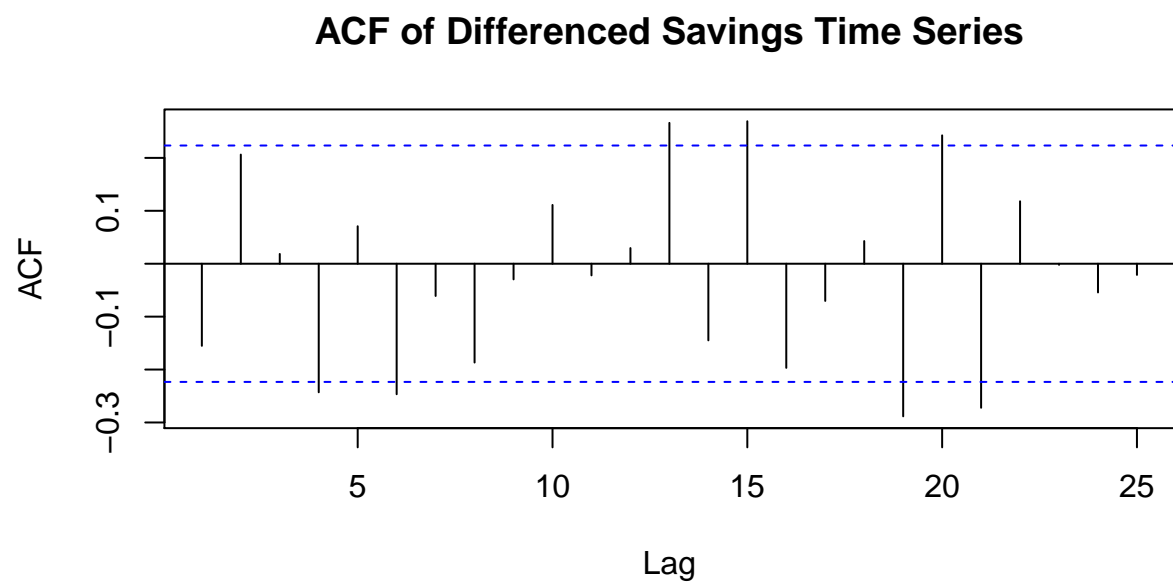
The PACF seems to indicate that an AR(1) process may be a good candidate model because the only non-zero sample partial autocorrelation is at lag $k = 1$. For lags $k \geq 1$, the partial autocorrelations appear to reduce to white-noise.

Next, we consider the differenced time series, which is plotted below:

Differenced Time Series of Personal Savings in the US (1955–1974)



The plot of the first difference of our time series indicates that the upward trend in the data has been removed and the mean of the differenced series is about equal to zero.



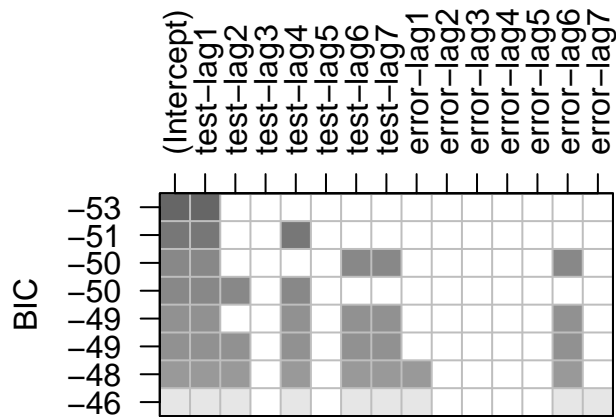
The ACF and PACF of the differenced series appear to suggest a seasonal trend in the differenced series, however, the period of this trend is unclear.

Model Specification

Based on the preliminary explorations of our original time series and the differenced series, the candidate models are have in mind are an AR(1) process or some type of seasonal model (period 4?) based on either the original or differenced series.

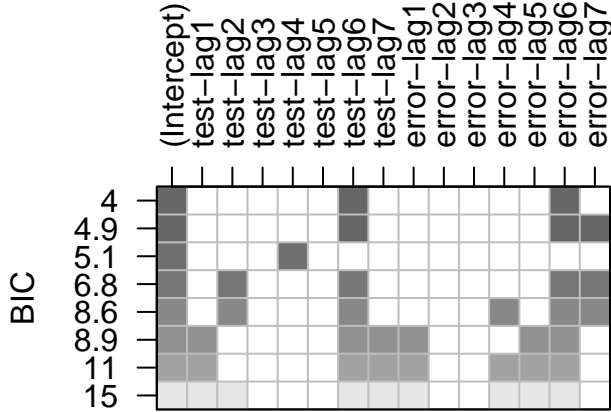
In order to validate these candidate models and determine the period of possible seasonal trends, we turn to the EACF and the best subsets methods.

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



While the EACF for the original (non-differenced) time series is inconclusive, the best subsets method indicates that, an AR(1) process or a multiplicative $AR(1) \times AR(1)_4$ with a seasonal period of 4 are good candidate models.

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



For the differenced series, a seasonal trend of 6 is suggested.

Model Fitting

Given the results in the previous section, we have decided to fit and compare 3 different models: an AR(1), a multiplicative $AR(1) \times AR(1)_4$, and an $ARIMA(0, 1, 0) \times ARIMA(1, 0, 1)_6$ model. The parameters for each model are given in the table below:

Model	Intercept	se	ar1	se	sar1	se	sma1	se	sigma^2	log likelihood	aic
AR(1)	6.28	0.35	.83	0.07	x	x	x	x	0.336	-68.71	141.43
Seasonal 4	6.27	0.35	0.86	0.06	-0.28	0.12	x	x	0.3198	-66.87	139.75
Seasonal 6	x	x	x	x	-0.18	0.38	-0.08	0.38	0.3402	-67.96	139.92

Thus the equations of our 3 models are:

1. AR(1): $Y_t - 6.28 = .825(Y_{t-1} - 6.28) + e_t$
2. $AR(1) \times AR(1)_4$: $(Y_t - 6.27)(1 - .861(B - 6.27))(1 + .228(B - 6.27)^4) = e_t$
3. $ARIMA(0, 1, 0) \times ARIMA(1, 0, 1)_6$: Y_t

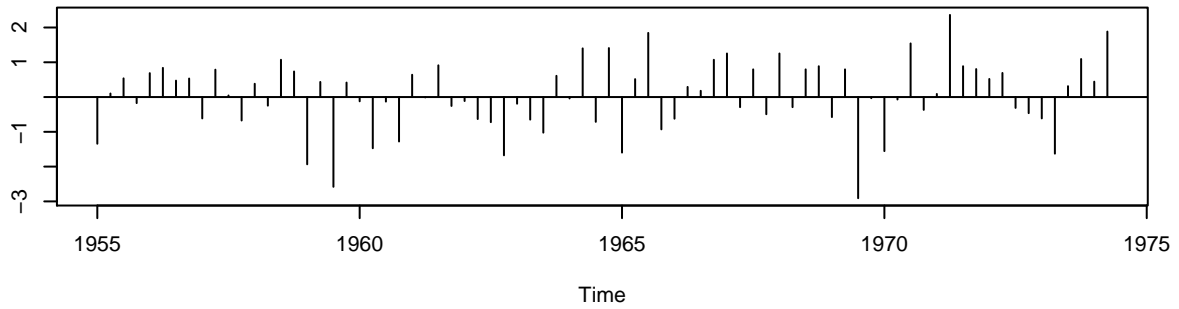
Diagnostics

Given the output from fitting the three models, we see that different error criterion point us to different model selections. While, the AR(1) model has the lowest BIC, but the seasonal model with a period of 4 has a lower standard error and AIC. Meanwhile the differenced seasonal model with a period of 6 has a similar AIC to the seasonal 4 model, but has the highest standard error. Thus, we turn to residual analysis to see if any of our models show abnormalities.

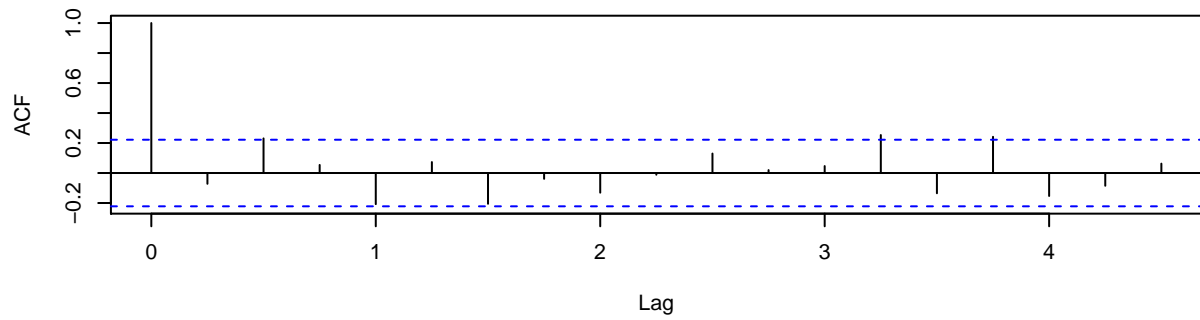
In our residual analysis we are looking for 3 things: residual nonnormality, residual dependence, and residual structure. The presence of any of these 3 things may indicate that our model has not sufficiently identified the structure of the data and is not an adequate model.

AR(1)

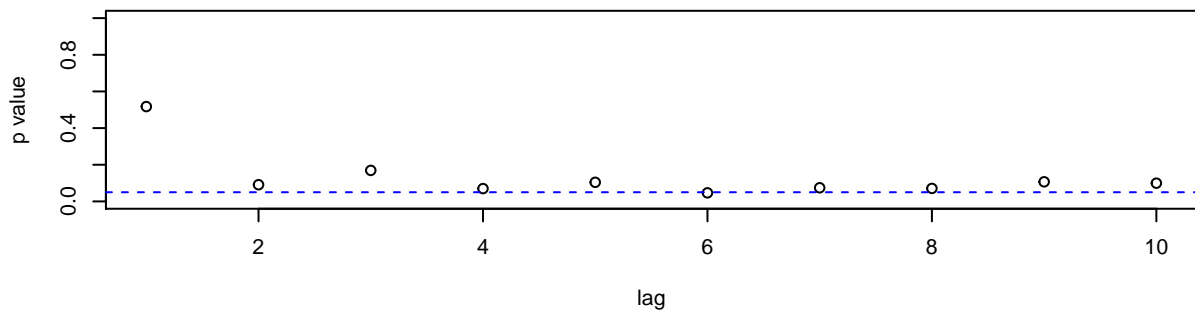
Standardized Residuals



ACF of Residuals



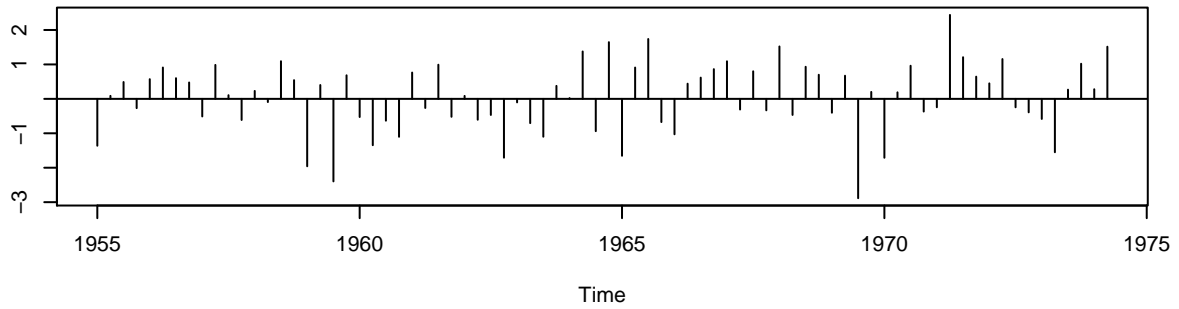
p values for Ljung-Box statistic



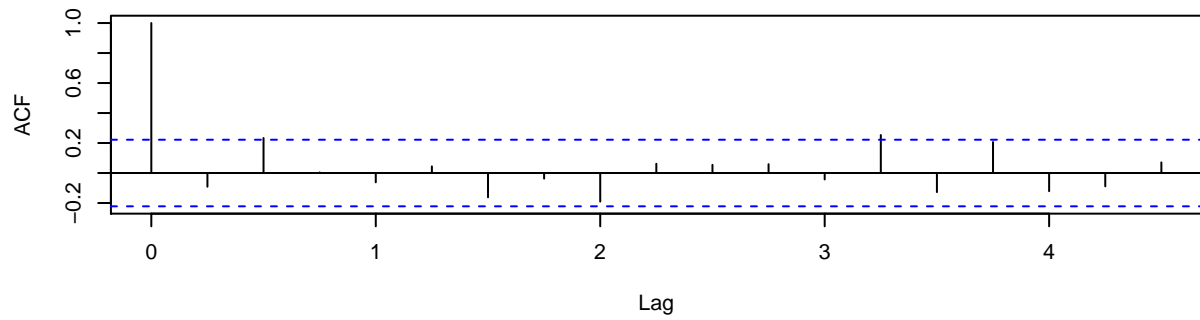
The residuals for the AR(1) model resemble a white-noise process, the ACF shows no correlation patterns between residuals, and the Ljung-Box statistic is borderline significant for higher lags. This indicates possible dependence among residuals.

AR(1) and Seasonal AR(1) with Period 4

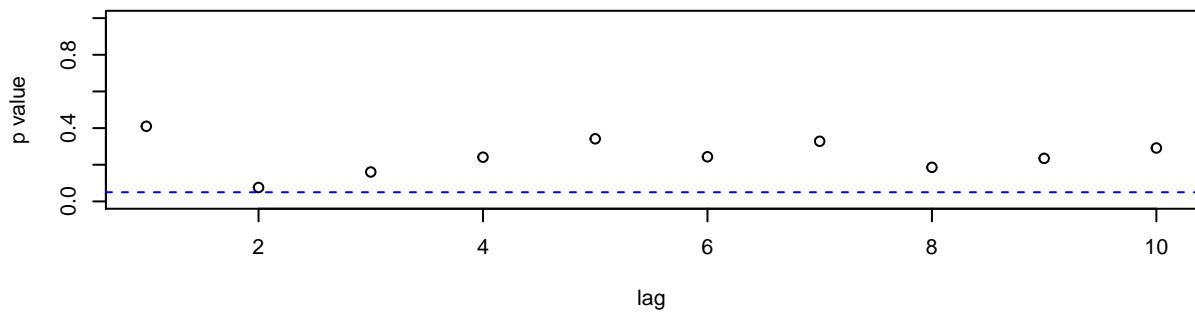
Standardized Residuals



ACF of Residuals



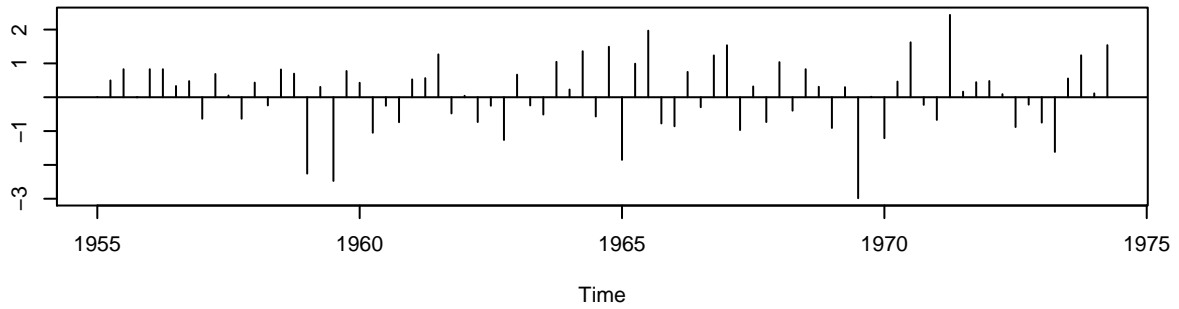
p values for Ljung-Box statistic



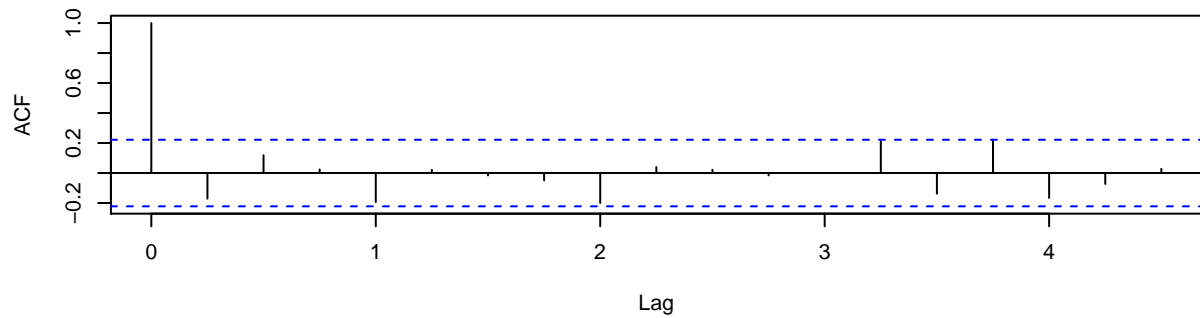
The residuals for the $AR(1) \times AR(1)_4$ model resemble a white-noise process, the ACF shows no correlation patterns between residuals, and the Ljung-Box statistic is not significant for all tested lags.

ARIMA(0,1,0) and Seasonal ARIMA(1,0,1) with Period 6

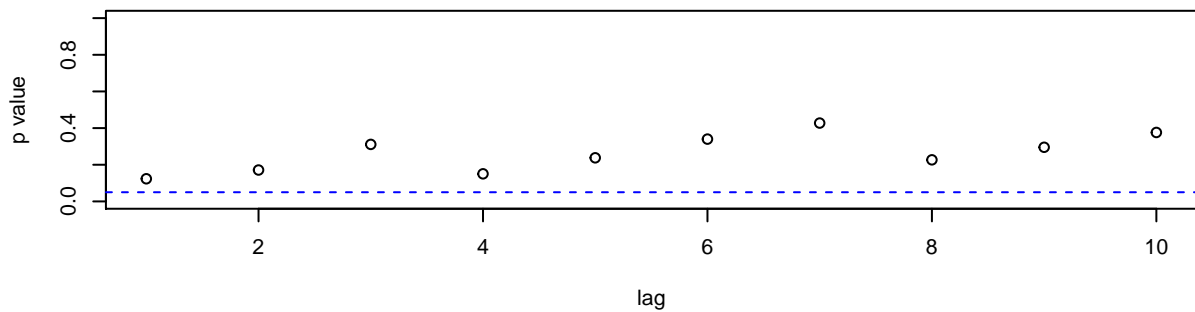
Standardized Residuals



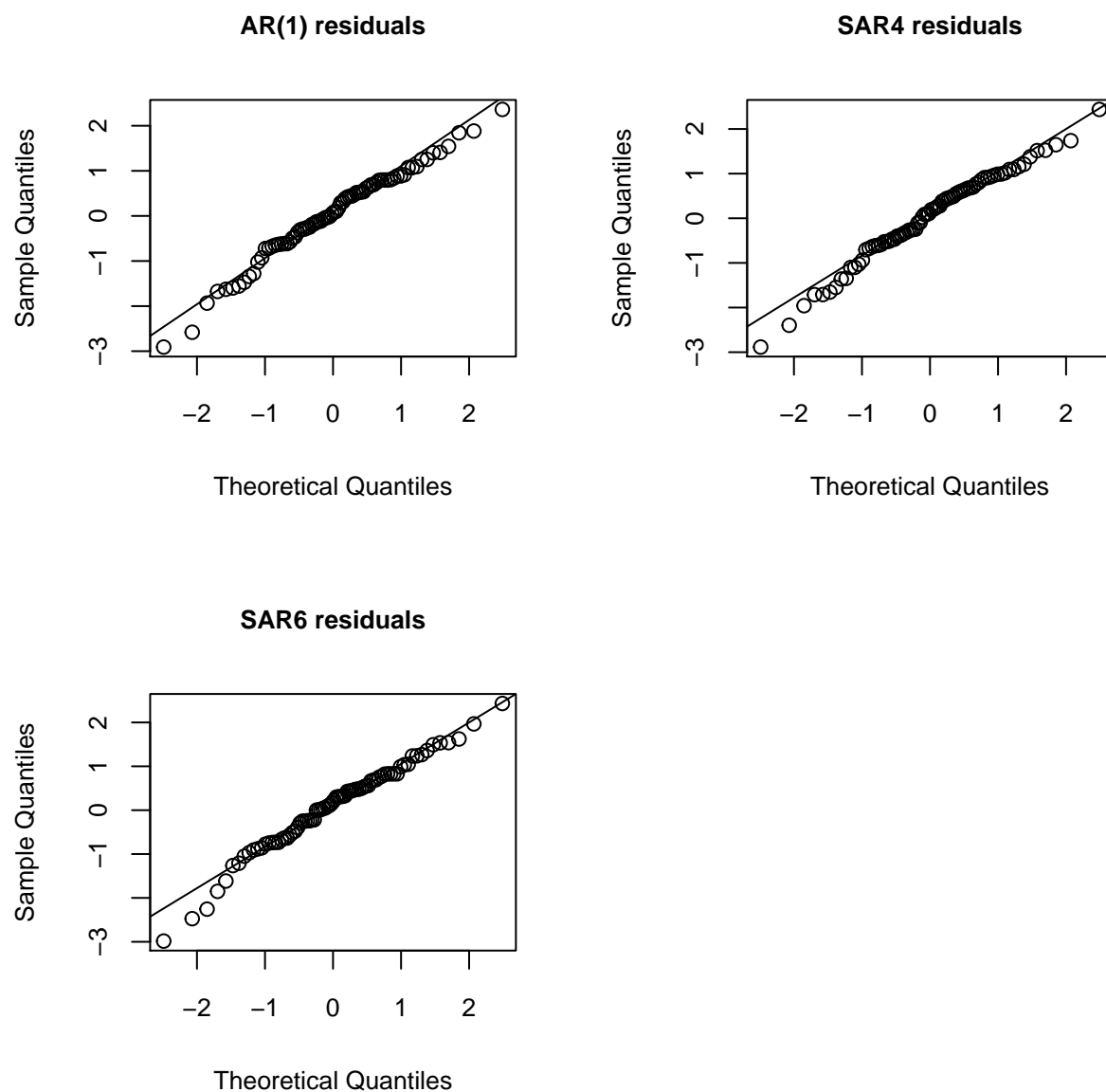
ACF of Residuals



p values for Ljung-Box statistic



The residuals for the $ARIMA(0,1,0) \times ARIMA(1,0,1)_6$ model resemble a white-noise process, the ACF shows no correlation patterns between residuals, and the Ljung-Box statistic is not significant for all tested lags.



Shapiro-Wilks Test

Model	AR(1)	Seasonal 4	Seasonal 6
W	0.981	0.983	0.979
p-value	0.295	0.369	0.237

Based on their Q-Q plots and their Shapiro-Wilk tests, none of the models show evidence of residual non-normality.

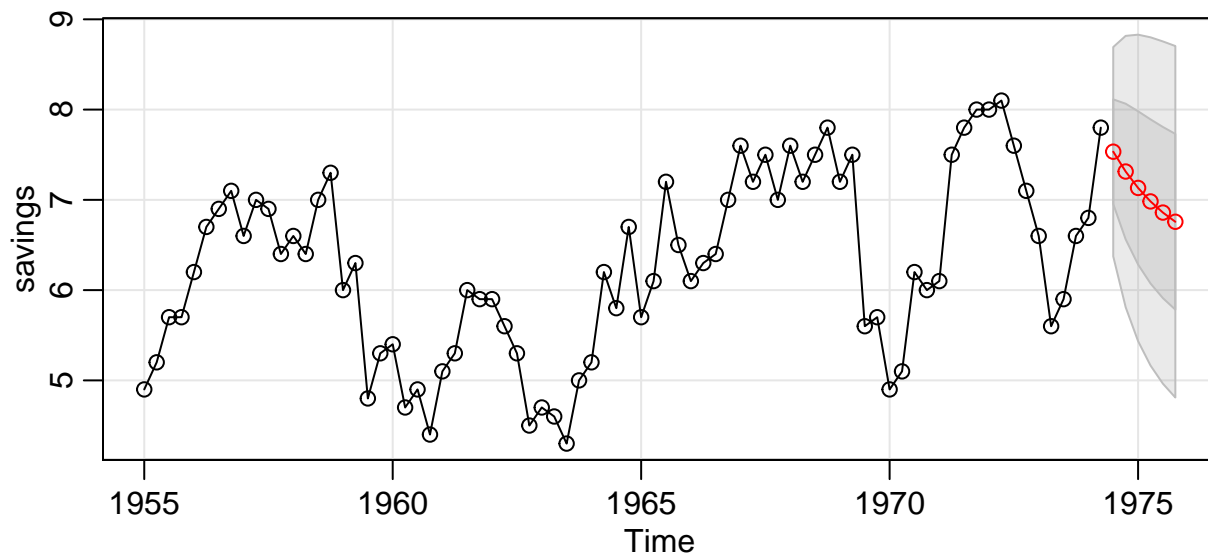
Runs Test

Model	AR(1)	Seasonal 4	Seasonal 6
p-value	0.73	0.37	0.18
Observed runs	42	44	45
Expected runs	30	40	39
n1	38	35	32
n2	40	43	46
k	0	0	0

Our runs tests support the Ljung-Box tests, which indicated that there is no sufficient evidence to reject residual independence.

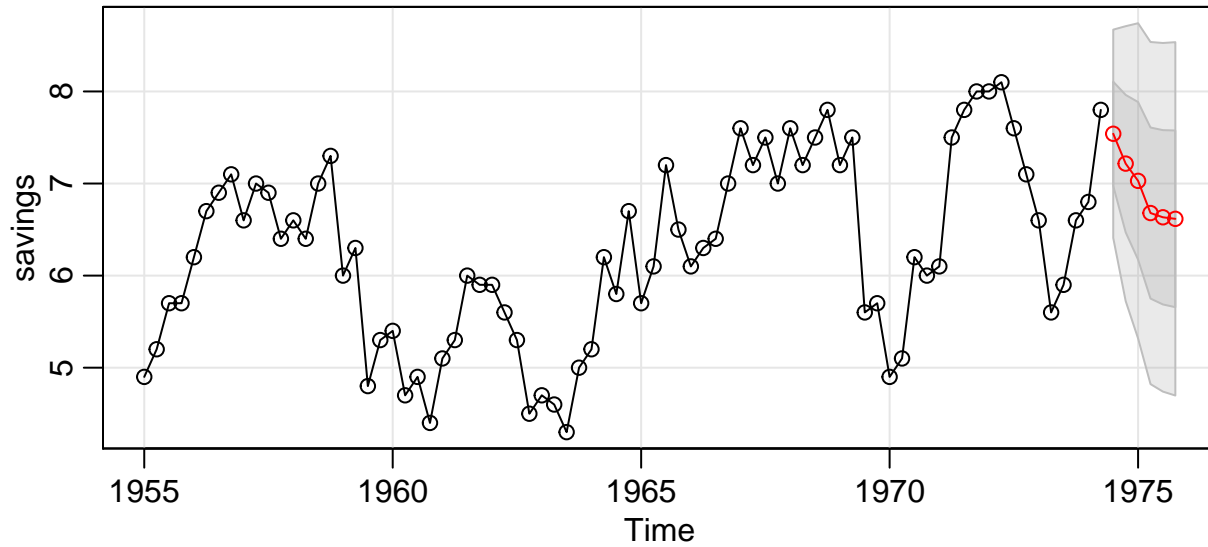
Forecasting

```
## $pred
##      Qtr1      Qtr2      Qtr3      Qtr4
## 1974                7.533497 7.313623
## 1975 7.132218 6.982552 6.859072 6.757196
##
## $se
##      Qtr1      Qtr2      Qtr3      Qtr4
## 1974                0.5796561 0.7514738
## 1975 0.8487602 0.9090457 0.9478906 0.9734455
```

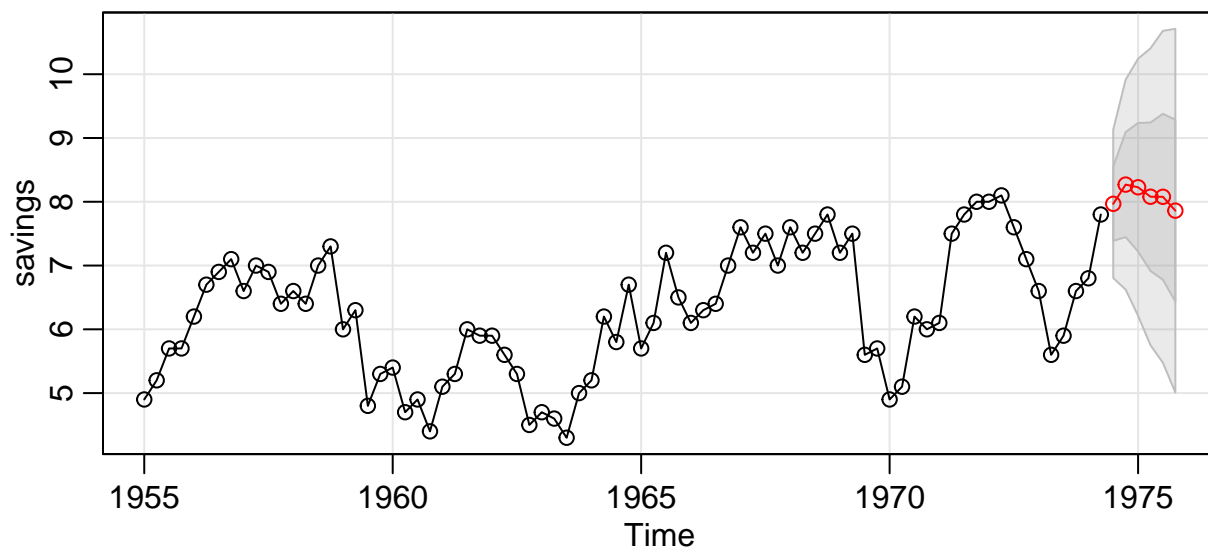


```
## $pred
##      Qtr1      Qtr2      Qtr3      Qtr4
## 1974                7.540401 7.216360
## 1975 7.029013 6.679309 6.633252 6.616476
##
## $se
```

```
##          Qtr1      Qtr2      Qtr3      Qtr4
## 1974                0.5654719 0.7461892
## 1975 0.8558728 0.9288601 0.9465329 0.9594238
```



```
## $pred
##          Qtr1      Qtr2      Qtr3      Qtr4
## 1974                7.967863 8.269178
## 1975 8.227219 8.078950 8.077971 7.860737
##
## $se
##          Qtr1      Qtr2      Qtr3      Qtr4
## 1974                0.5819161 0.8229537
## 1975 1.0079083 1.1638322 1.3012040 1.4253975
```



Appendix: R Code

```
# load the time series
savings = read.csv("savings.csv",header=TRUE, nrows=104)
savings = savings %>% select(2)
savings.entire = savings[1:84,]
savings<-ts(savings, start=c(1955),frequency=4)

# plot the original time series
plot(savings, xlab="Year (by Quarter)",
      ylab= "% of Disposable Income",
      main= "Time Series of Personal Savings in the US (1955-1980)")
points(y=savings,x=as.vector(time(savings)),pch=as.vector(season(savings)), cex=.75)

# set aside points to validate our forecasts
savings.test = savings[79:84,]

# remove the last 5 years because of the huge dip due to recession
savings = savings[1:78,]
savings<-ts(savings, start=c(1955),frequency=4)

# plot the shortened time series with labels for quarters
plot(savings, xlab="Year (by Quarter)",
      ylab= "% of Disposable Income",
      main= "Time Series of Personal Savings in the US (1955-1980)")
points(y=savings,x=as.vector(time(savings)),pch=as.vector(season(savings)), cex=.75)
abline(lm(savings~time(savings)), col='blue')

# should we do a transformation?
boxcox = BoxCox.ar(savings)
boxcox
boxcox$mle

# plot the acf and pacf of the original series
acf(savings, lag.max = 25)
pacf(savings, lag.max = 25)

# calculate the differenced time series
diffs = (savings-zlag(savings))[2:78]

# plot the differenced time series
plot(diffs, xlab="Year (by Quarter)",
      ylab= "% of Disposable Income",
      main= "Differenced Time Series of Personal Savings in the US (1955-1980)", type="o")
abline(lm(diffs~time(diffs)), col="blue")

# plot the acf and pacf of the differenced time series
acf(diffs, lag.max = 25)
pacf(diffs, lag.max = 25)
```

```
# use the eacf and best subsets to find a candidate model
eacf(savings)
sub = armasubsets(y=savings,nar=7,nma=7, y.name='test', ar.method='ols')
plot(sub)
```

```
# plot the eacf and best subsets for the differenced series
eacf(diffs)
sub = armasubsets(y=diffs,nar=7,nma=7, y.name='test', ar.method='ols')
plot(sub)
```

```
# fit an AR(1) process
AR1model = arima(savings, order = c(1, 0, 0), seasonal = list(order = c(0, 0, 0)), method=c('ML'))
AR1model
```

```
# run some general diagnostics on the models
tsdiag(AR1model)
tsdiag(SAR4model)
tsdiag(SAR6model)
```

```
# test the residuals for normality
shapiro.test(ARresids)
shapiro.test(SAR4resids)
shapiro.test(SAR6resids)
```

```
# test the residuals for independence
runs(ARresids)
runs(SAR4resids)
runs(SAR6resids)
```