

Personal Savings Analysis

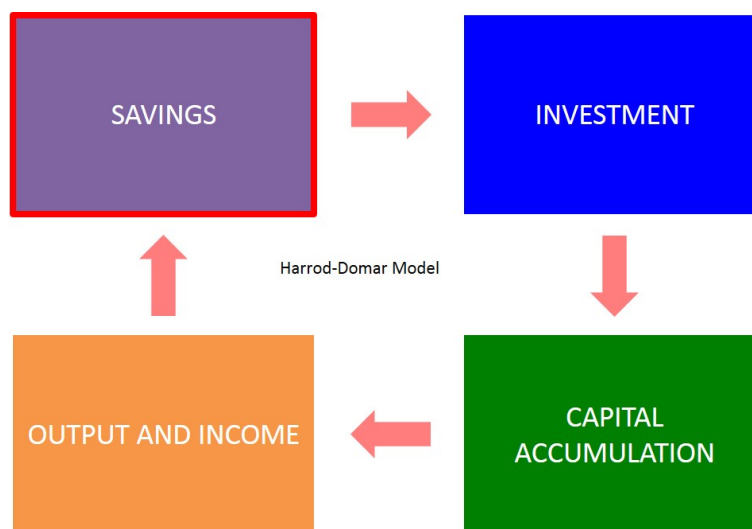
Andrew Brown, Melissa Hooke, Frances Hung, Mai Nguyen, Brenner Ryan

12/19/2018

Abstract: This paper explores personal savings data in the United States from 1955 - 1980; a span of 26 years. First, we examined the trend of our data by using various time series analysis tools such as model specification and model fitting. 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 and diagnostic analysis of the three models, we concluded that an $AR(1) \times AR(1)_4$ was the most appropriate model for the data set. With this finding, we were able to create a forecast of the U.S. personal savings rate. These analyses allowed us 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

The personal savings rate of consumers is one of many indicators of how a country's economy is performing. For example, savings drive long-run economic growth as they provide funds for investment in capital or projects, which then drives future economic growth. Typically, household savings are invested either directly (i.e. when purchasing equity) or indirectly, (i.e. putting them into a bank, which uses those funds for lending) (Carroll and Mowry). These investments lead to economic growth in free-market economies. Additionally, the Harrod-Domar Model of economic growth suggests that economic growth rates are driven by the level of savings. The model reasons that increased savings will lead to increased investment which result in higher capital stock, and thus, higher economic growth. An increase in savings rate would furthermore enable higher investment without a rise in net foreign borrowing. This allows a country to be less dependent on foreign capital and more insulated from possible international capital shocks that can lead to financial panic (Staff).



Maintaining a high savings rate is also crucial for personal prosperity and financial security in retirement. The Bureau of Economics Analysis found that households aged 40 and above, those in the bottom income quartile need to save about 21 percent more on average of their pre-tax income to ensure financial security in retirement (Staff). Failure to do so would result in households either working beyond official retirement age, accepting a lower standard of living in old age, or, in the worst case, running out of money altogether (Staff).

To understand the factors that influence personal savings rate, it is helpful to know how the savings rate is calculated. Personal savings can be understood as one minus the ratio of personal outlays (spending) to disposable income (personal income minus personal taxes). This calculation is expressed in the formula below.

$$\text{Personal Savings Rate (\%)} = 100 \times \left(1 - \frac{\text{Personal Outlays}}{\text{Personal Income} - \text{Personal Taxes}} \right)$$

Moreover, this shows that an increase in the personal savings rate can be associated with one of the following factors: increase in personal income, decrease in personal outlays, or decrease in personal taxes. (Carroll and Mowry)

The personal savings rate typically decreases when individuals spend more than they initially save, be it due to inflation, consumer habits, or reliance on other financial assets. Because the personal savings rate can only go so low, a low rate can be a sign of a looming recession. In contrast, after financial crises, consumers are

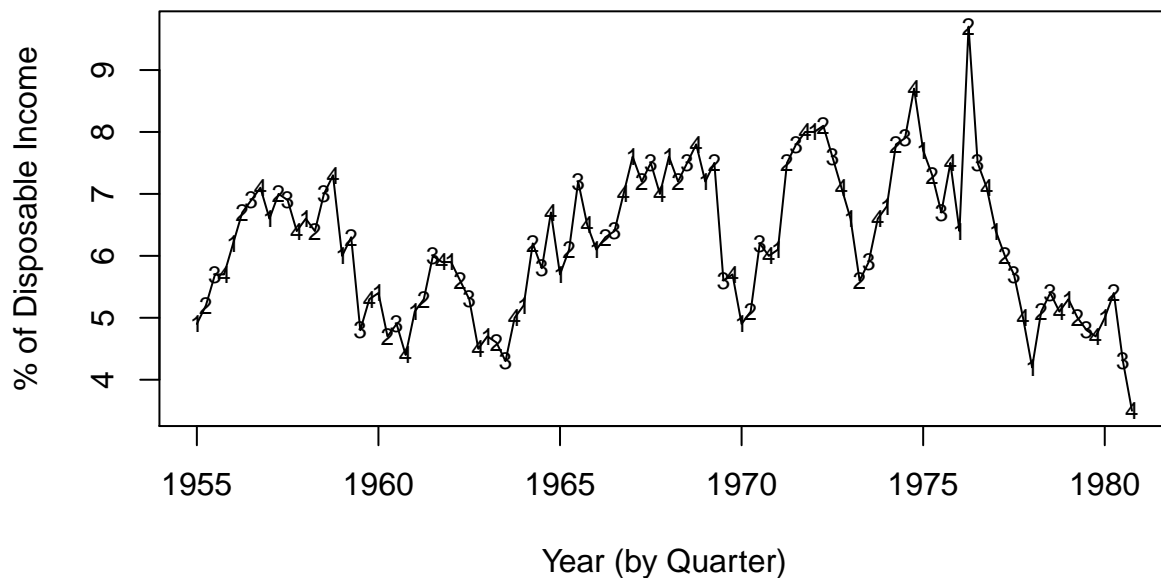
often more cautious and tend to raise the personal savings rate. For example, in 2008, personal savings rate rose from 1.4 percent to 2.6 percent and in 2009, it reached 4.3 percent, highest since 1998. (Carroll and Mowry) This increase can be explained by the 2008 recession, after which consumers became more cautious with their spending.

Given that saving rates are an indicator of a country's economic performance, it is important to explore personal saving rates data in order to gain insights on the performance of economy. In this project, we are most interested in exploring the U.S. personal savings rate. By analyzing past saving rates, we are able to form a time series which forecasts future rates. This understanding will enable us to take observe the projection of the personal savings rate and take precautionary measures to ensure economic growth or brace for recessions.

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:

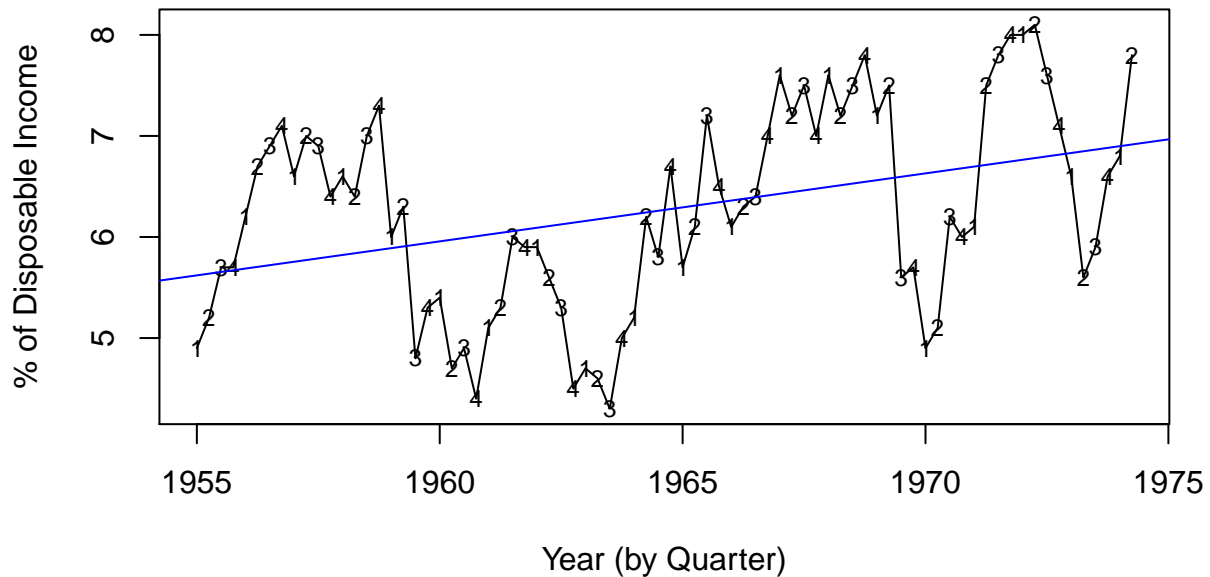
Time Series of Personal Savings in the US (1955–1980)



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:

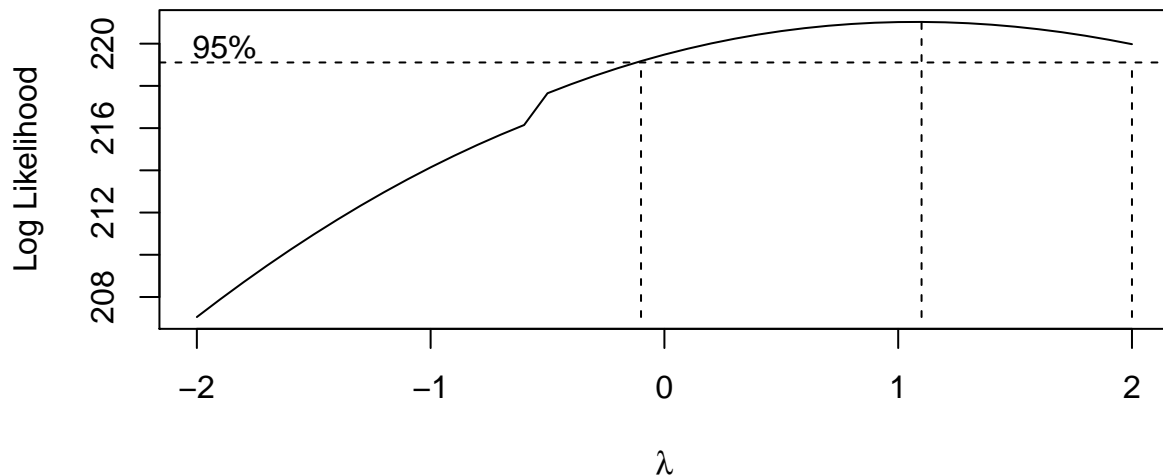
Time Series of Personal Savings in the US (1955–1974)



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 we 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.

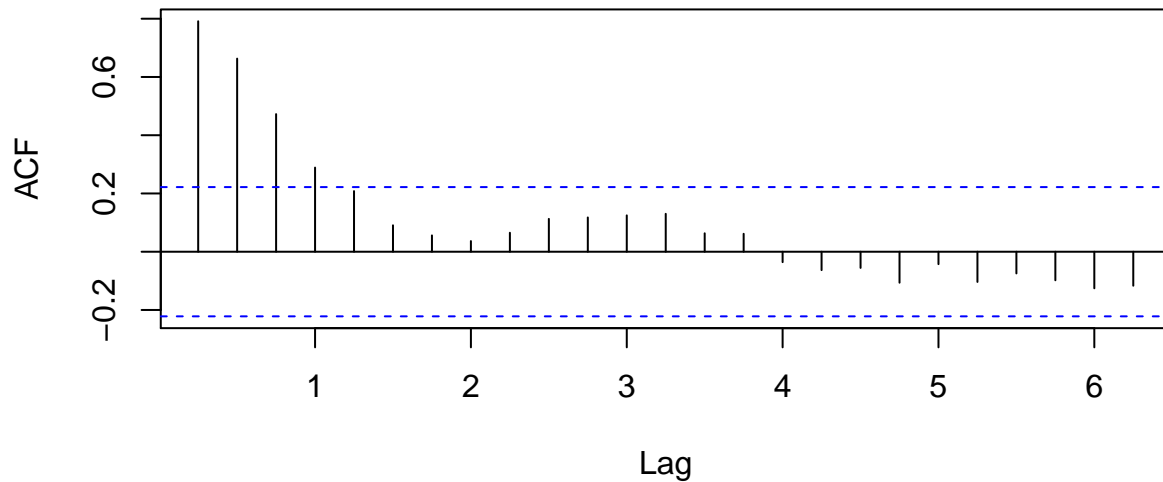
The first step in analyzing our time series is to consider the possible need for a transformation to stabilize the variance of the series over time. In order to do this, we use the function `BOXCOX.AR` to determine the appropriate power transformation for time-series data.

Box Cox Plot

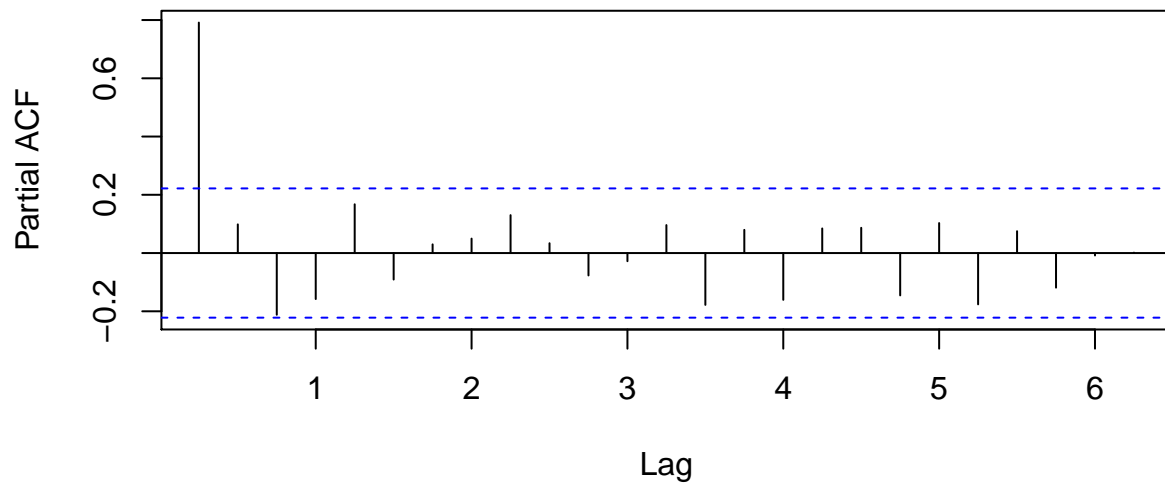


The Boxcox output indicates that a transformation is not necessary in order to stabilize the variance since λ is about equal to 1. Therefore, we proceed by examining the acf and pacf of the series.

ACF of Savings Time Series



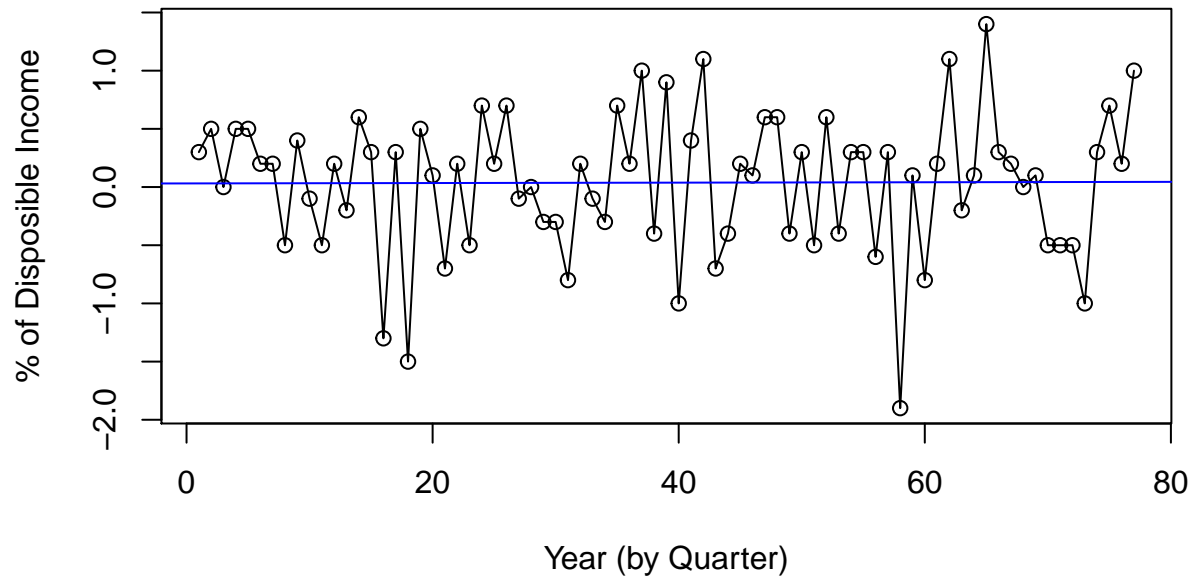
PACF of Savings Time Series



The ACF starts out large and then slowly decreases until it ends up within the white noise bounds. This kind of behavior is usually seen in ARMA and AR models. Looking at the PACF could give us more information about the nature of our data. 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.

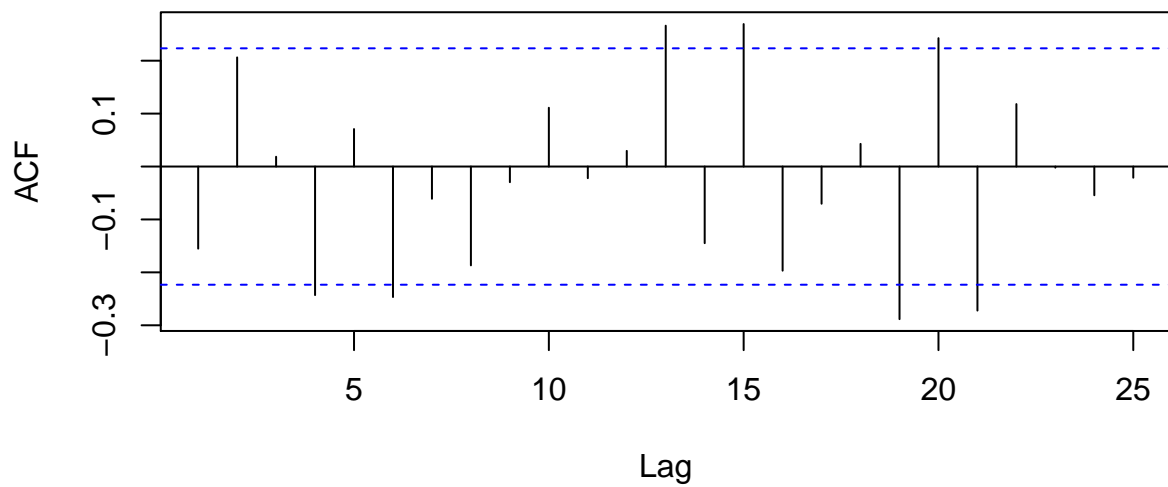
Next, we consider the differenced time series, which is plotted below:

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

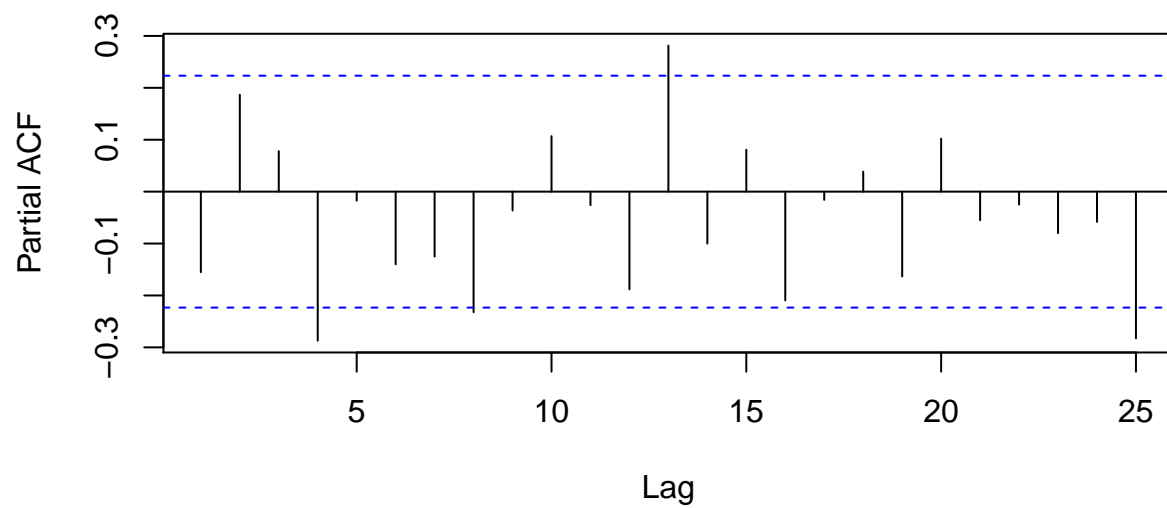


The plot of the first difference of our time series indicates that the upward trend in the data has been removed and the mean of the differenced series is about equal to zero.

ACF of Differenced Savings Time Series



PACF of Differenced Savings Time Series



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 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

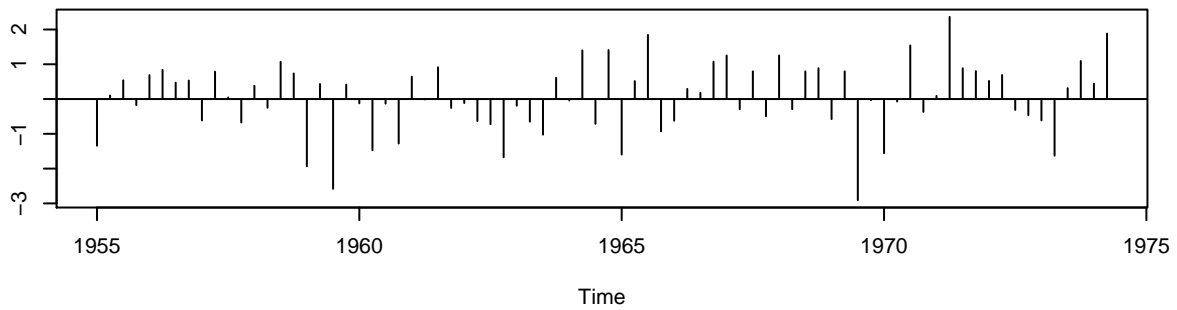
Diagnostics

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, 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.

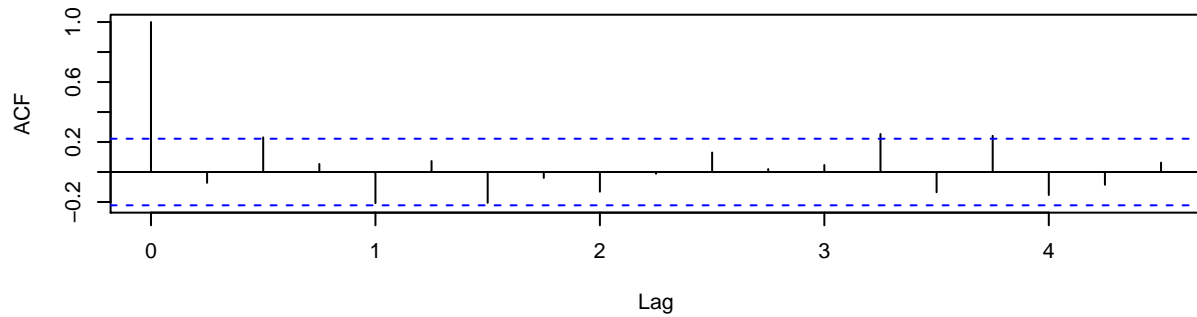
In our residual analysis we are looking for 3 things: residual nonnormality, residual dependence, and nonconstant variance among the residuals. The presence of any of these 3 things may indicate that our model has not sufficiently identified the structure of the data and is not an adequate model. In order to do so, first we look at the standardized residuals, the ACF of the residuals, and the Ljung-Box test for each of our three models. A plot of the standardized residuals can illustrate residual nonnormality in the model, the residual ACF looks for dependence among the residuals, and the Ljung-Box test checks if the specified model is appropriate for the data or not. For the Ljung-Box test, the alternative hypothesis is that the specified model is not appropriate for the data, so smaller p-values indicate a possibly inappropriate model.

AR(1)

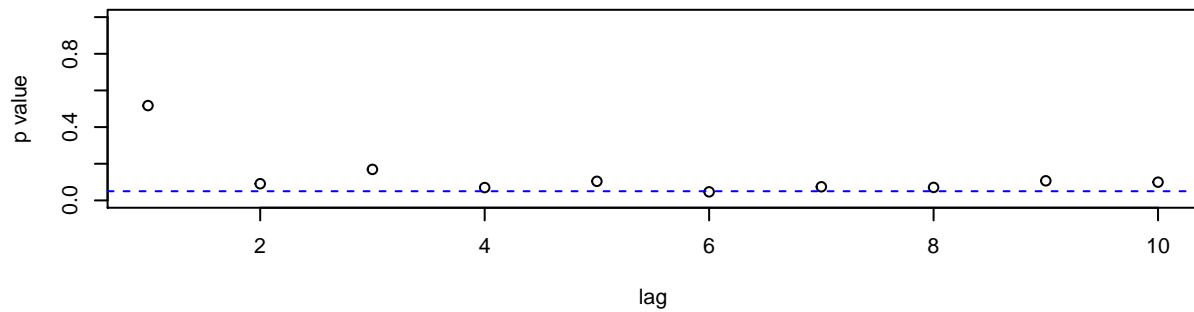
Standardized Residuals



ACF of Residuals



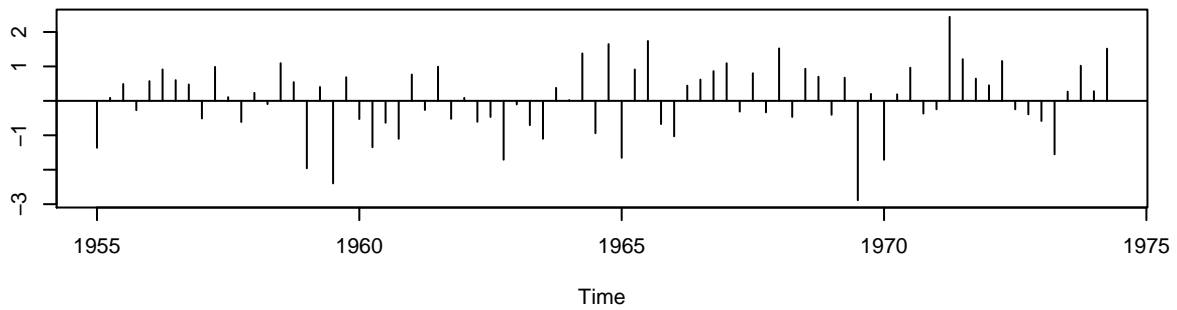
p values for Ljung-Box statistic



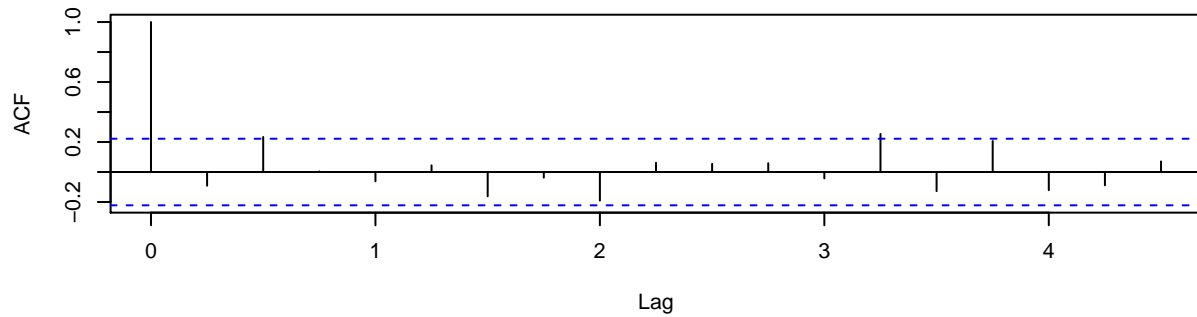
The residuals for the AR(1) model do not indicate any nonnormality or heteroskedasticity, but the ACF shows that some lags do approach the error bounds, and the Ljung-Box statistic is borderline significant for higher lags. This indicates possible dependence among residuals.

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

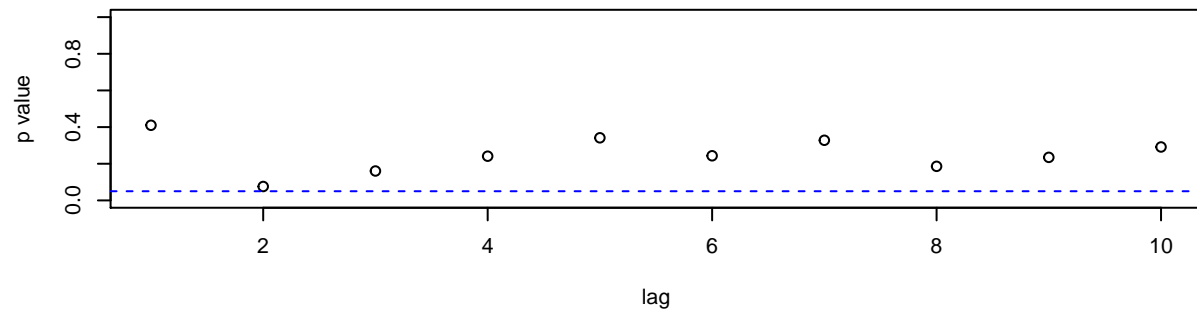
Standardized Residuals



ACF of Residuals

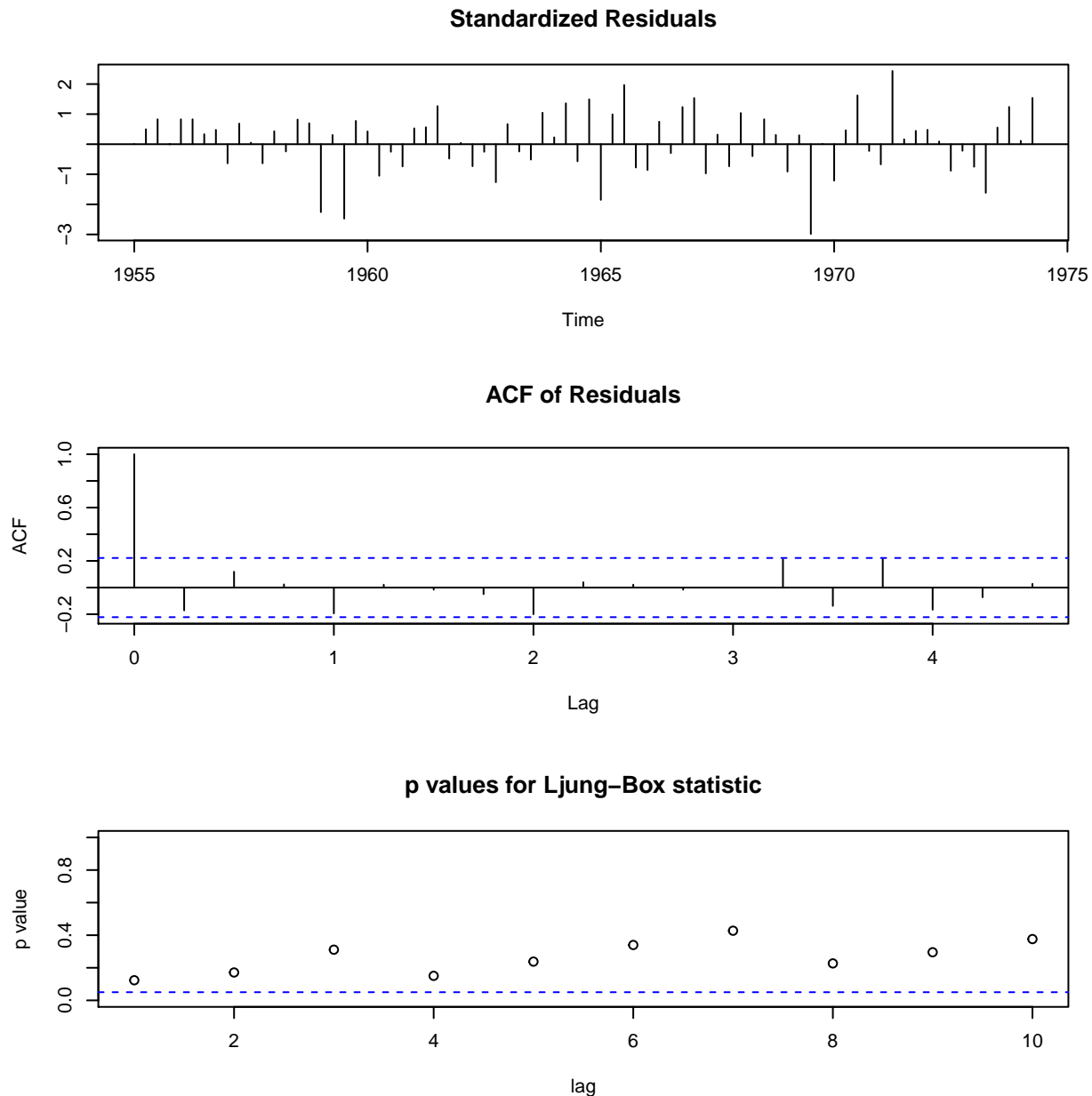


p values for Ljung-Box statistic



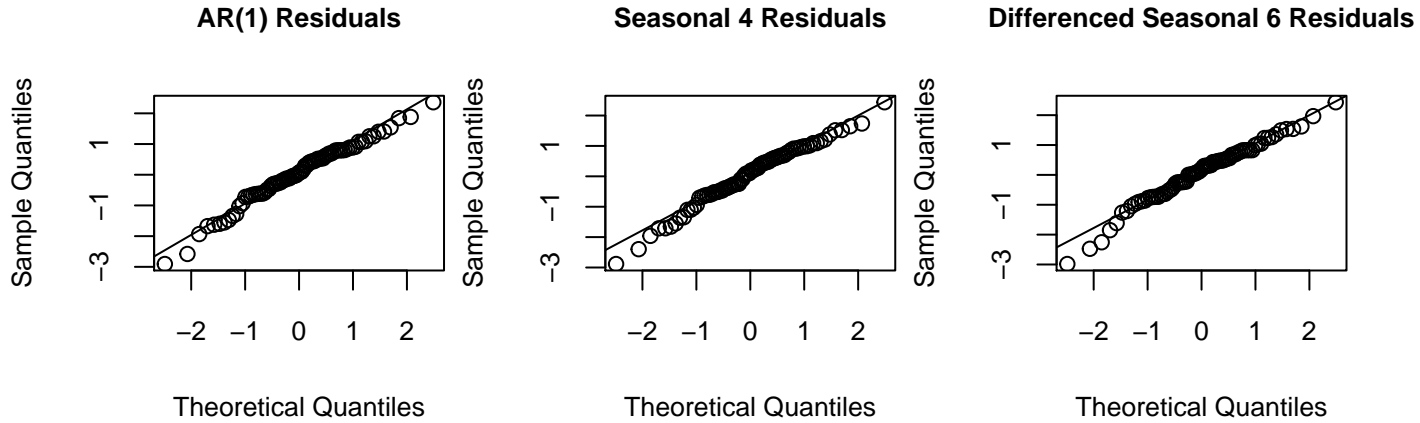
The residuals for the $AR(1) \times AR(1)_4$ model also do not indicate any nonnormality or heteroskedasticity, all bounds of the residual ACF are within the error bounds, pointing to a lack of correlation between residuals, and the Ljung-Box statistic is not significant for all tested lags.

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



The residuals for the $ARIMA(0, 1, 0) \times ARIMA(1, 0, 1)_6$ model again do not indicate any nonnormality or heteroskedasticity, all bounds of the residual ACF are also within the error bounds, pointing to a lack of correlation between residuals, and the Ljung-Box statistic is not significant for all tested lags.

Now, we will look at the Q-Q plots and Shapiro-Wilks tests for each of our three candidate models. Q-Q plots graph the quantiles of two probability distributions on each axis - in our case, a normal distribution on the x-axis, and the distribution of our residuals on the y-axis. If the residuals are normally distributed, the plot should form a straight line along the diagonal. The farther the plot strays from the diagonal, the less normal our residuals are. Similarly, the Shapiro-Wilks test tests for normality among standardized residuals. The alternative hypothesis of the Shapiro-Wilks test is that the standardized residuals of a certain model are not normally distributed, so smaller significant p-values indicate nonnormality among residuals.



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 nonnormality.

Finally, we will turn to runs tests. A run is a sequence of increasing or decreasing values, so a large difference in expected vs. observed runs may point towards residual dependence. Runs tests test for independence among standardized residuals of a certain model. The alternative hypothesis of the runs test is that the standardized residuals of a certain model are not independent, so smaller p-values indicate dependence among residuals.

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

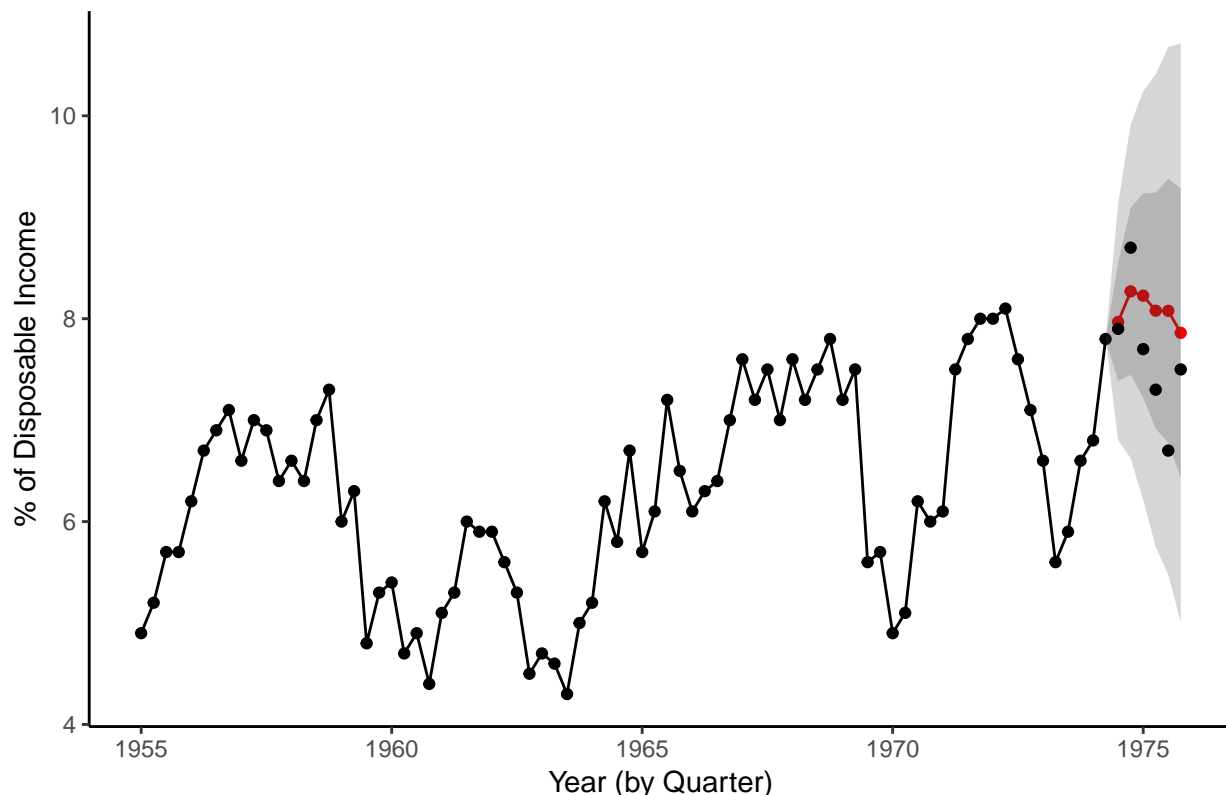
Our runs tests support the Ljung-Box tests, which indicated that there is no sufficient evidence to reject residual independence.

Forecasting

In our previous section, we concluded that the best model for our series is an $ARIMA(0,1,0) \times ARIMA(1,0,1)_6$. We now want to predict the values of the series at future times. This forecast will enable us to gain insights into the direction of which the savings rate is headed as well as the future health of the economy.

We forecasted our model with a built-in forecast ARIMA function in R. The red line signifies the forecasted values with a confidence bound of one standard error in darker grey and of two standard errors in lighter grey. We observed that the forecasted values increase at first and then slowly begin to decay. Additionally, the forecasted values generally maintain the previous seasonal pattern and its upward trend. Our forecasted values stay within one standard deviation from our actual values, which shows that our forecasts are a good prediction of our data.

Time Series of Personal Savings with Forecasts (1955–1975)



Our analysis concludes that the savings rate will decrease in the near future time points as predicted by the model. However, our analysis also supports a general upward trend of the savings rate. This suggests that the economy is heading towards a temporary cyclical trend of choosing short-term consumption over long-term investment, but generally, savings rates are growing. A continuation of this temporary downward trend can indicate a looming economic stagnation. We hope that our analysis will help indicate when the savings rate becomes abnormally low or high compared to our forecasts, which may be a sign of impending economic trouble.

Discussion

The goal of the present analysis on the personal savings data was to construct a time series model which would adequately capture the savings trends across the data and accurately predict future data points in the years to come. In our analysis, we considered 3 different ARIMA and SARIMA models: a basic AR(1) process, a seasonal model with a period of 4, and a differenced seasonal model with a period of 6. Through our model diagnostics, we settled on the differenced seasonal 6 model as most appropriate model because removing the trend from the time series appeared to sufficiently reduce the series to a white noise process with normal and independent residuals.

One major challenge that we experienced in choosing our final model was the debate between the seasonal 4 model and the differenced seasonal 6 model. The fact that the data was broken down by financial quarter made the appearance of the seasonal 4 trend enticing. However, we decided that the upward trend in the data was too strong to ignore in the final model of the time series. Differencing the time series as we did in the final chosen model adequately preserved the upward trend in personal savings over time.

We acknowledge that our analysis has a few shortcomings. The first one is that we cut off the last 5 years of our data set in order to obtain data with a general upward trend. As a consequence of this, our model does not include nor predict those 5 years of drastically tumbling savings rates. Moreover, our analysis is not generalizable to today's savings rate because this data was captured 40 years ago. Though our time series provided insights into the past health of the economy, we would also be interested in seeing a similar time series of more current data. If the same seasonal and upward trend is prevalent for today's data, it would be supportive of economic commonalities which drive savings rate changes back from the 70's through today.

Bibliography

“Personal Savings as % of Disposable Income 1955-1979.” DataMarket, <https://datamarket.com/data/set/22vy/personal-savings-as-of-disposable-income-1955-1979#!ds=22vy&display=line>

Sullivan, Bob. “Once Again, Americans Are Not Saving Enough.” MarketWatch, MarketWatch, 28 Aug. 2018, www.marketwatch.com/story/once-again-americans-are-not-saving-enough-2018-08-28.

Carroll, Daniel, and Beth Mowry. “Personal Savings Up, National Savings Down.” Economic Trends (07482922), vol. Apr2010, P8, 2010.

Pettinger, Tejvan. “Would an Increase in Savings Help the Economy?” Economics Help, www.economicshelp.org/blog/7102/economy/an-increase-in-savings-help-the-economy/

“Understanding the Economic Benefits of Increased Saving | Commentary.” Roll Call, 26 June 2014, www.rollcall.com/news/understanding_the_economic_benefits_of_increased_saving_commentary-234283-1.html.

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

# 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)

```