SCHOOL OF OPERATIONS RESEARCH
AND INFORMATION ENGINEERING
COLLEGE OF ENGINEERING
CORNELL UNIVERSITY

ORIE 5550: TIME SERIES ANALYSIS

A Time Series Analysis on
Retail Trade and Food Service Sales and Consumer Sentiment Index

By:
Antong Su (as3657)
Charlotte Wang (xw476)
Vidita Gawade (vag39)

Professor: Ines Wilms
Teaching Assistant: Megan Gelsinger

1. Introduction

1.1 Background of Dataset and Research Question

Consumer sentiment is a statistical measurement and economic indicator of the overall health of the economy as determined by consumer opinion (Investopedia). When consumer sentiment is less positive, markets typically react bearishly and vice versa.

With the data released by University of Michigan, we are interested in studying the change in consumer sentiment from 1992 to 2017 and predicting consumer sentiment index in 2008 if possible. We intuitively think consumer spending will affect the consumer sentiment index. We are interested to answer the research question: for a given month and year, does monthly retail trade and food service sales affect monthly consumer sentiment index?

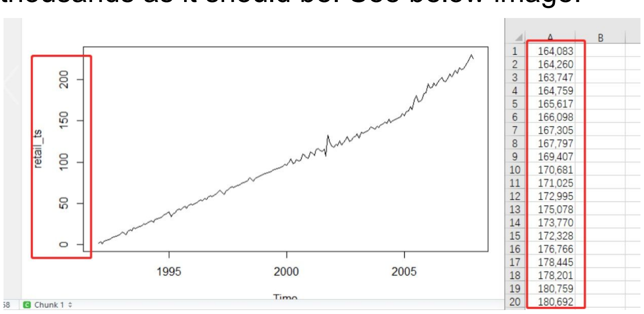
In this analysis, we worked with two datasets. The table summarizes both datasets.

	Dataset 1	Dataset 2
Name of Dataset	Consumer Sentiment Index	Retail trade and food service sales
Source	University of Michigan	U.S. Census Bureau
Size	Monthly data from January 1992 to December 2017	Monthly data from January 1992 to December 2017
Unit	CPI	Millions of dollars
Data Cleansing*	Analyzed data only from January 1992 to December 2007 to avoid recessionary period after December 2007	Got rid of commas, Analyzed data only from January 1992 to December 2007 to avoid recessionary period after December 2007

Our analysis was broken into two parts. The univariate time series analysis was performed only on consumer sentiment index. The multivariate time series analysis was performed on both aforementioned datasets to help answer our research question.

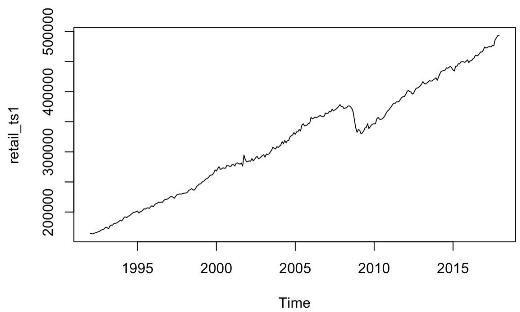
1.2. Data Cleansing

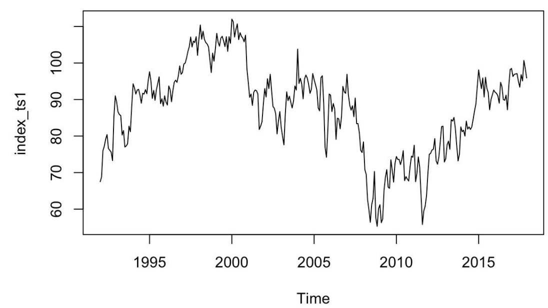
We first observed the entire time series dataset for both of our datasets to ensure that the data was read correctly (checked y-axis made sense) and to observe the general trend of both datasets. We observed that the values for the Retail Trade and Food Service sales were not being read correctly as the y-axis showed units in hundreds as opposed to hundred thousands as it should be. See below image:



To resolve this issue, we got rid of the commas in the csv file. For example, instead of reading in 100,000, we got rid of the comma, returning 100000 in the cell. This returned the correct plot with the correct sales values in hundred thousands.

We next observed both plots of dataset (see below image) and observed that in general CPI fluctuates over time and sales increases over time, except around the recessionary period (post 2007), we notice a dip in both datasets which is a pattern against the general trend of the datasets.



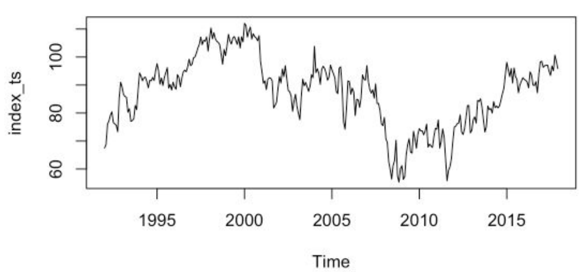


To resolve this, we decided to perform our multivariate analysis on both datasets from only January 1992 to December 2007.

2. Univariate Time Series Analysis

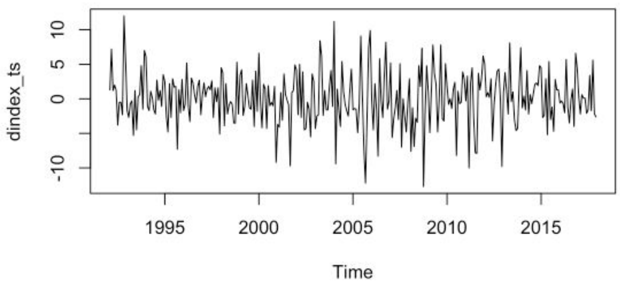
2.1. Stationarity Test

We first plotted the time series data and the plot showed no trends.



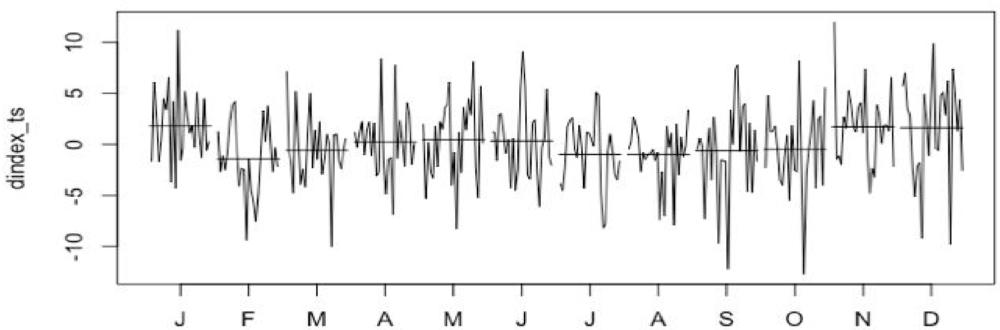
This implied that there is a possibility that the index dataset is stationary. We further tested the hypothesis using Augmented Dickey–Fuller unit root test, with trend specified as “drift”. As we can see in the following test result, p-value, 0.0792, is greater than significant alpha value of 0.05, which means we failed to reject the null hypothesis that the time series object has a unit root. In this case, the original time series dataset is not stationary.

We then tried the difference of original consumer sentiment index dataset and again observed no trend in the corresponding plot.



As a result, we performed unit root test with type specified as “drift” and received a p-value less than 2.2e-16 which helps us reject the null hypothesis that the time series object has a unit root. In this case, the new dataset is stationary.

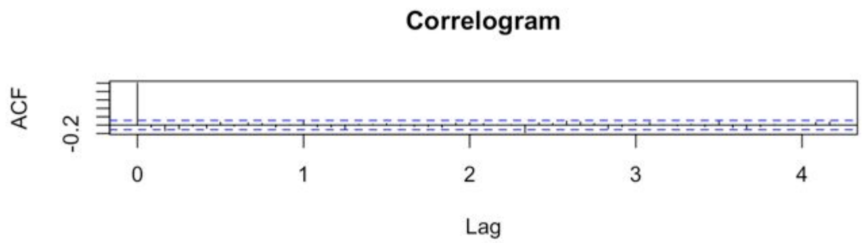
We further wanted to confirm if no seasonality exists. In the monthplot, variance in each month is significant, so we did not see any seasonality patterns.

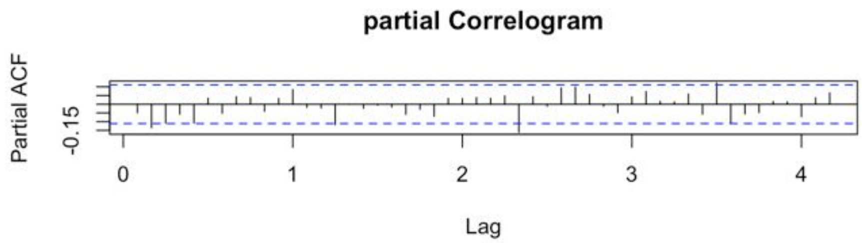


Knowing that the original index dataset is not stationary but the difference of index is, we may conclude that the original consumer sentiment index time series dataset is integrated of order 1.

2.2 Model Specification

With the new stationary difference-index time series, we specified appropriate ARMA models on the basis of correlogram and partial correlogram. From the correlogram, we can specify a parsimonious moving average model with an order of 2, because roughly no significant structures after lag of 2. From the partial correlogram, we can specify a parsimonious autoregressive model with an order of 2 because roughly no significant structures exist after lag of 2. Even though we could observe few structures present after lag of 2, they are not at the seasonal lags. We decided to treat those as noises.





We tried fitting an ARIMA(0,1,2) model for original index time series and observed that the highest term is significant.

```
Series: index_ts
ARIMA(0,1,2)

Coefficients:
          ma1      ma2
      -0.1001  -0.1649
s.e.   0.0571   0.0573

sigma^2 estimated as 14.72:  log likelihood=-858.48
AIC=1722.96   AICc=1723.04   BIC=1734.18
     ma1      ma2
1.755111 2.875922
```

To check the validity of the model, we plotted the correlogram of the residuals and observed that no significant structures were present in the plot. To further confirm the residuals series is white noise, we used Ljung-Box test. As a result, p-value of 0.4825 is greater than significance level of 0.05. We failed to reject the null hypothesis; thus, the residuals series is white noise and the moving average model is validated.

Similarly, we tried fitting an ARIMA(2,1,0) model for original index time series and observed that the highest term is significant.

```
Call:
arima(x = index_ts, order = c(2, 1, 0), seasonal = c(0, 0, 0))

Coefficients:
          ar1      ar2
      -0.0559  -0.1355
s.e.   0.0562   0.0564

sigma^2 estimated as 14.73:  log likelihood = -859.6,  aic = 1725.19
      ar1       ar2
0.9952852 2.4029132
```

To check the validity of the model, we plotted the correlogram of the residuals and observed that no significant structures were present in the plot. To further confirm the residuals series is white noise, we used Ljung-Box test. As a result, p-value of 0.353 is greater than significance level of 0.05. We failed to reject the null hypothesis; thus, the residuals series is white noise and the autoregressive model is validated.

2.3 Model Comparison

Now we have two forecasting models for consumer sentiment index time series: ARIMA(0,1,2) and ARIMA(2,1,0), we could compare them with each other based on both in-sample and out-of-sample forecast criteria. As for in-sample criteria, we chose Bayesian Information Criterion. This criterion was computed using forecast-errors, where the forecast was based on the model estimated using all available data. Moving average model has BIC of 1734.179 and autoregressive model has BIC of 1736.41. As a result, we may conclude that the moving average model with an order of 2 is better, given a smaller BIC, as the forecast result would be closer to the truth.

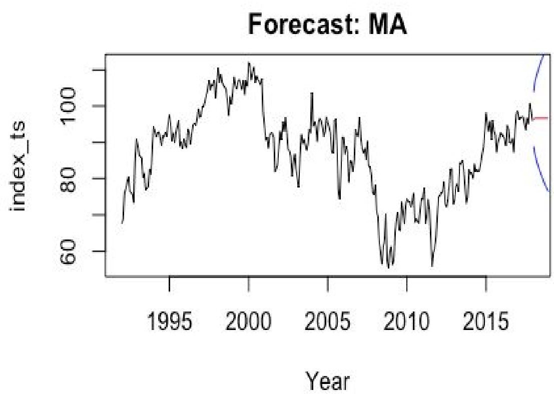
We used Diebold-Mariano Test to compare the out-of-sample forecast errors from two different models. We first tested mean absolute errors for two models and received p-value of 0.436, which is greater than the significant level of 0.05. As a result, we failed to reject the null hypothesis. Therefore, there is no evidence to prove that two errors from moving average model and autoregressive model are different. In this case, the forecast performances of moving average model and autoregressive model are equally well.

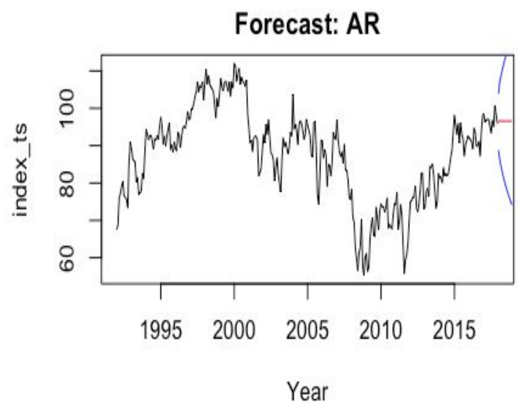
```
        Diebold-Mariano Test

data:  errorMA.herrorAR.h
DM = -0.78297, Forecast horizon = 1, Loss function power = 1, p-value = 0.436
alternative hypothesis: two.sided
```

To ensure the above conclusion is consistent for other error format, we performed Diebold-Mariano Test on mean squared errors as well for two models and received p-value of 0.791, which is greater than the significant level of 0.05. As a result, we failed to reject the null hypothesis. Therefore, there is no evidence to prove that two errors from moving average model and autoregressive model are different. In this case, the forecast performances of moving average model and autoregressive model are equally well.

```
        Diebold-Mariano Test

data:  errorMA.herrorAR.h
DM = -0.26596, Forecast horizon = 1, Loss function power = 2, p-value = 0.791
alternative hypothesis: two.sided
```

We then tried to forecast consumer sentiment index in 2018 with two different models. 95% prediction intervals for both models look similar and are widen over time.





3. Multivariate Analysis

3.1. Stationarity Test

Before we built the models using both variables, we made sure that both variables are stationary, and if they were not, apply transformations to the variables to make sure they did become stationary. The process to make the retail sales dataset stationary is summarized as follows. We forced looked at the respective plot of the dataset, and then applied a unit-root test to determine stationarity. This process repeated for each transformation we did on the dataset.

The plot for the retail sales dataset showed that it has an increasing trend; in general, the retail sales increases over time. This implied that the original retail sales time series dataset may not be stationary. But to make sure, we performed a unit-root test on the retail sales dataset. We used the CADF library to perform the unit-root test getting a p-value of 0.7668 which was greater than significant alpha value of 0.05. This meant that we could not reject the null hypothesis that the time series object is stochastic.

We next tried the log of retail sales dataset. Once again, the plot outputted on the log of the retail sales dataset showed that there was still trend. To make sure the object was no longer stochastic, we applied the unit-root test again. But, we still got a p-value greater than 0.05, this time of 0.1734 which meant that the object was still stochastic.

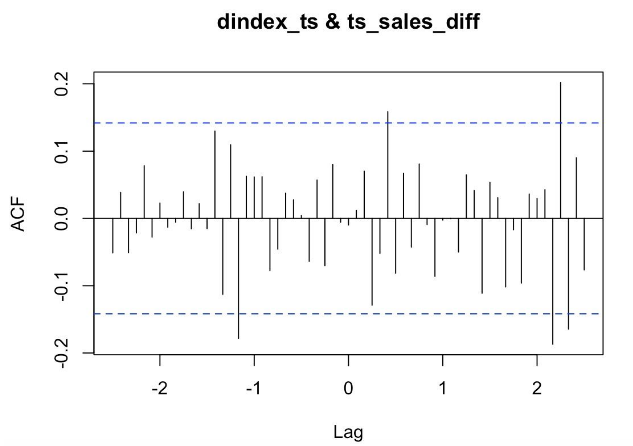
We next tried the differences of the retail sales dataset and observed no longer a trend in the plot but a drift. As a result, we performed a unit-root test with type= “drift” and received a p-value of 2.2e-16 which was less than significant alpha level of 0.05. Therefore, we were able to reject the null hypothesis that the differences of the retail sales dataset is not stationary. Instead, we were able accept the alternative hypothesis that differences of the retail sales time series is indeed stationary and is integrated of order 1.

In order to perform multivariate on both time series objects, we simply decided to also apply differences on the original consumer sentiment index dataset. We performed a unit-root test with type= “drift” and also received a p-value of 2.2e-16, which was less than 0.05. Therefore, we were able to reject null hypothesis that the differences of the consumer sentiment index is not stationary. Instead, we were able accept the alternative hypothesis that differences of consumer sentiment index time series is stationary and is integrated of order 1.

3.2. Specify, Estimate, Validate some Dynamic Models (DL)

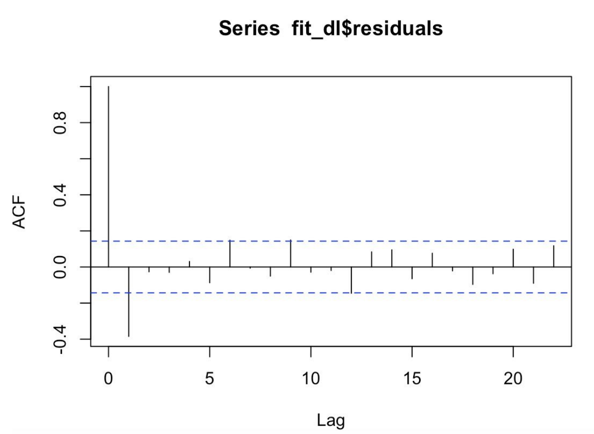
Predictive Power Analysis

Before we specified models for the two variables, we first looked at the cross correlogram of the variables to determine which variable had predictive power for the other. According to our research question, we hoped to observe that retail food and sales had predictive power for consumer sentiment index. As we can see by our cross correlogram, there are most significant lags (approximately 4) after instantaneous lag=0 than before (approximately 1). We also observe no significant lag at instantaneous correlation lag=0.



Distributed Lag Model Analysis

We applied a lag of 4 to our distributed lag model since that was observed in our cross correlogram. The residuals plot (see below) of our distributed lag model showed only one significant lag, but mostly no significant lags after lag=1.



This meant our model has no white-noise and would mean our model will most likely be not validated since it is preferred that we do see significant lags in our residuals plot for model validation. Indeed, after we performed the Box-Jenkins test, we got a p-value of 1.804e-05 which was less than significant alpha level of 0.05. Therefore, we reject the null hypothesis that our model white noise and accept the alternative hypothesis that our model is not white noise and cannot be valid.

<u>ADL Model Analysis</u>

We next tried an ADL model, first with lag of 4 as we did with our DL model. We got a very high p-value of 0.93 which validated our model. Since we aimed for parsimonious models, we decreased the lag to 1 and were able to still validate our model. Indeed, after we performed the Box-Jenkins test, we got a p-value of 0.5304 which was less than significant alpha level of 0.05. Therefore, we could not reject the null hypothesis that our model is indeed white noise and is indeed validated.

3.3. Investigate Granger Causality

Although we were able to get a validated ADL model, we were not able to confirm granger causality. The process for investigating granger causality occurred as follows. We built a linear model on just the consumer sentiment index (our response variable) without including our predictor variable of retail sales. After we performed an anova test comparing the ADL model with the linear model with only our response variable, we observed no significant difference between the models as we got a p-value of 0.878. Therefore, we could not reject the null hypothesis that our predictor variable granger causes our response variable. In other words, we cannot confirm that retail trade and food service sales granger causes consumer sentiment index.

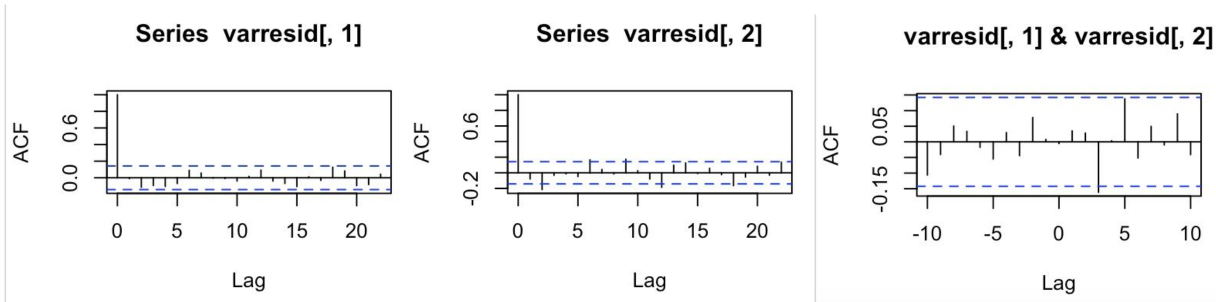
3.4. VAR Model Specification and Estimation and Validation

<u>Specification and Estimation of VAR Model</u>

Next, we specified and estimated a VAR model. The automatic lag selection resulted in the following: AIC selected 2 lags and BIC (SC) selected 1 lag. When we have a situation where both methods are selecting different lags, we will go with the lag that BIC selects since BIC penalizes higher order terms. As a result, we selected lag = 1.

<u>Validation Analysis of VAR Model</u>

In order to validate our model, we should not observe significant lags in our residuals, in other words have existence of white noise. So we observed the correlograms of the individual variables, as well as the cross correlogram of both variables. The graphs are shown:



From the CCF plot, we observed that there is no significant lag at instantaneous lag=0 and there is only 1 significant lag after instantaneous lag=0. From the ACF plot of consumer sentiment index (refer to varrresid[,1] in plot), we observed no significant lags, but from the ACF plot of the retail sales (refer to varesid[,2] in plot), we observed three significant lags. At this point, it was hard to tell if our model will be validated or not, but showed potential to be validated since our cross correlogram showed almost no significant lags for white noise.

From the contradicting observations above, it was important to validate the VAR model quantitatively using a multivariate noise test, rather than by just simply observing the plots. The multivariate quantitative noise test resulted in the following result:

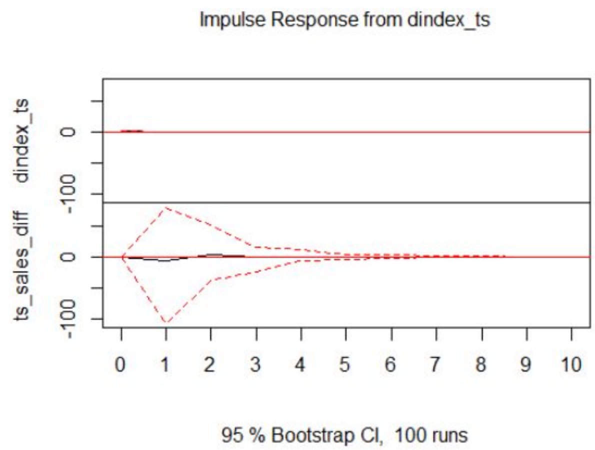
Ljung-Box Statistics:

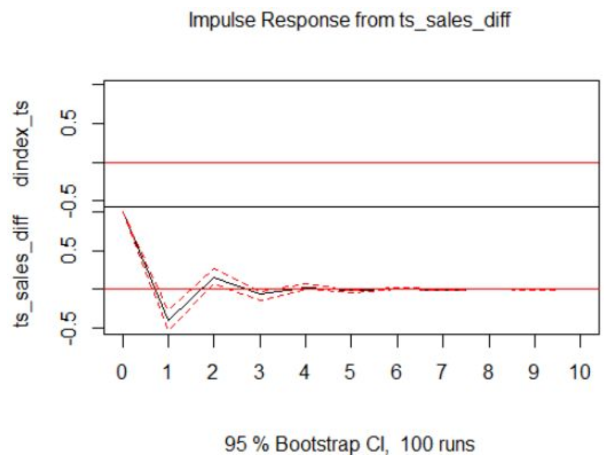
	m	Q(m)	df	p-value
[1,]	1.00	1.45	4.00	0.84
[2,]	2.00	14.07	8.00	0.08
[3,]	3.00	21.37	12.00	0.05
[4,]	4.00	23.67	16.00	0.10
[5,]	5.00	29.32	20.00	0.08
[6,]	6.00	36.83	24.00	0.05
[7,]	7.00	38.56	28.00	0.09
[8,]	8.00	39.12	32.00	0.18
[9,]	9.00	47.14	36.00	0.10
[10,]	10.00	50.18	40.00	0.13
[11,]	11.00	53.66	44.00	0.15
[12,]	12.00	64.67	48.00	0.05
[13,]	13.00	68.56	52.00	0.06

We were interested in the last row only which helped test our null hypothesis that all 13 coefficients are not significant and are equal to 0. Indeed, we observed a p-value of 0.06 which is just as approximate as 0.05. While we could technically accept this model, we would not be confident in this acceptance. This p-value means that we could be 94% confident that we can reject our null-hypothesis and accept the alternative hypothesis that at least one residual is non-zero. Thereby, this VAR model should be used with caution because of p-value of 0.06.

3.5. Impulse Response Functions Study

To interpret the effect of a one unit-shock in consumer sentiment index growth on retail sales growth and the effect of a one unit-shock in retail sales growth on retail sales growth itself, we plot the impulse-response functions:





From the plots above, we can see the effect of a one unit-shock in consumer sentiment index change on retail sales growth does not have significant lag. As for one unit-shock in retail sales growth on retail sales growth itself, there is negative significant effect on lag 1, positive significant effect on lag 2, and marginal negative effect on lag 3. There is no effect of impulse from consumer sentiment index change on consumer sentiment index change itself, no any effect of impulse from retail sales growth on retail sales growth itself.

4. Cointegration

4.1 VAR Selection

To determine the optimal lag length needed in the Johansen test equation, we performed automatic lag selection. After referring to the running result, we picked lag of 2 for Johansen test.

4.2 Johansen Trace Test

In our example, we used Johansen trace test to see whether there is cointegration between consumer sentiment index and retail sales:

```
######################
# Johansen-Procedure #
######################

Test type: trace statistic , without linear trend and constant in cointegration

Eigenvalues (lambda):
[1] 2.545515e-01 6.089855e-02 2.730640e-17

Values of teststatistic and critical values of test:

          test 10pct  5pct  1pct
r <= 1 | 11.94  7.52  9.24 12.97
r = 0  | 67.75 17.85 19.96 24.60
```

From the test result above, we can see the test statistics for r = 0 is 67.75, much greater than the critical value 19.96 under the 0.05 significance level, so we are going to reject r = 0. On the other hand, the test statistics for r <= 1 is 11.94, which is also greater than the critical value 7.52 under the 0.05 significant level; however, if we look at 0.01 significance level, we see the test statistics is less than the critical value 12.97. Therefore, there is one cointegration relation between consumer sentiment index and retail sales, but we are not very confident to this since it is only based on 0.01 significance level.

4.3 Estimation of VECM

After we validated the conitegration, we can find the vector error correcting model that fits the two time series.

```
$beta
                     ect1
retail_ts.l1        1.000
index_ts.l1      5433.962
constant       -95528.127
```

From the running result, we can summarize the cointegration equation below:

$$retail\ sales_t + 5433.692 \times consumer\ index_t - 95528.127 = 0$$

5. Conclusion

In this study, we analyzed two datasets using three time series techniques: univariate time series analysis, multivariate time series analysis and cointegration study.

For the univariate analysis, we built a moving-average model with order of 2 and an autoregressive model with order of 2 to forecast consumer sentiment index in 2018. We would prefer the moving-average model considering in-sample criterion; however, if looking at the out-of-sample criterion, the moving-average model shows no advantage over the autoregressive model.

We were interested in exploring if retail food and trade sales affected consumer sentiment index using multivariate analysis. While we were able to build an autoregressive distributed lag model that was validated, we were not able to confirm that retail food and auto trade sales did indeed granger cause consumer sentiment index. Similarly, while we were able to validate our VAR model, our multivariate noise test said that we could only trust the model with 94% confidence. Our impulse function study showed that retail sales growth can impact retail sales growth itself, but consumer sentiment index growth is not significantly affected by retail sales.

By applying VAR Selection and Johansen trace test, we discovered one cointegration relation between consumer sentiment index and retail sales in 0.01 significant level. The final vector error correcting model summarizes the cointegration relationship between the two time series.

Therefore, if we are interested in using these the sales to help us predict consumer sentiment index, this study should not be used. However, we believe in the long-run, that the two variable move together. We learned that in the classroom setting, datasets can be selected so that it can work with the techniques learned in class. However, to apply a time series analysis using real-world data to answer a real-world question, data cleansing and different methodologies were required in addition to the techniques learned in class. For example, both datasets would need to have the same transformation for us to perform a multivariate analysis on it.

Appendix: Code of Analysis

````
---

title: "Time Series Analysis Project"

author: "Vidita Gawade, Charlotte Wang, Antong Su"

date: "4/12/2018"

output:

  word_document: default

  html_document: default

---
````

CLEAR ANY PREVIOUS SAVED VARIABLES and READ IN DATA FILES

````
```{r}

rm(list=ls())

CS <- read.csv(file="Consumer_Sentiment.csv", header=TRUE) # Consumer sentiment

RFS <- read.csv(file="Retail_Food_Sales.csv",header=FALSE) # Retail food sales

```
````

OBSERVE ORIGINAL TIME SERIES OBJECTS to see that data gets read in correctly and see general observation.

````
```{r}

index_ts1 <- ts(CS$Index, frequency = 12, start=c(1992,1))

retail_ts1 <- ts(RFS, frequency=12, start=c(1992,1));

ts.plot(index_ts1);ts.plot(retail_ts1)

```
````

READ IN NECESSARY LIBRARIES

```{r,warning=FALSE,message=FALSE}

library(CADFtest)

library(vars)

library(MTS)

library(urca)

```

UNIVARIATE ANALYSIS CHARLOTTE

```{r}

attach(CS)

index_ts <- ts(Index, frequency = 12, start=c(1992,1))

ts.plot(index_ts)

```

First explore time series data: we don't see any trends in the plot, and we will confirm this with the CADF test.

```{r}

library(CADFtest)

max.lag<-round(sqrt(length(index_ts)))

CADFtest(index_ts, type= "drift", criterion= "BIC", max.lag.y=max.lag)

```

The p-value for the stationarity test on the index_ts object is 0.0792, therefore we cannot reject the null hypothesis that the object is not stationary.

```{r}
dindex_ts <- diff(index_ts)

ts.plot(dindex_ts)
```

We don't see any trends in the plot, and we will confirm this with the CADF test.

```{r, warning=FALSE, message=FALSE}
library(CADFtest)

max.lag<-round(sqrt(length(dindex_ts)))

CADFtest(dindex_ts,type="drift",max.lag.y=max.lag,criterion="BIC")
```

The p-value for the stationarity test on the dindex_ts object is less than 0.05, therefore we reject reject the null hypothesis that the object is not stationary. Now we have a stationary time series. index_ts is integrated of order 1.

```{r}
monthplot(dindex_ts)
```

No seasonality is shown in the plot.

```{r}
```

acf(dindex_ts,lag.max=50) #MA(2)

pacf(dindex_ts,lag.max=50) #AR(2)

```

We can specify a moving average model with an order of 2 from acf. Because roughly no significant peaks after 2.

We can specify a autoregressive model with an order of 2 from pacf. Because roughly no significant peaks after 2.

```{r, warning=FALSE,message=FALSE}

library(forecast)

fit_ma <- Arima(index_ts, order = c(0,1,2))

fit_ma

abs(fit_ma$coef/sqrt(diag(fit_ma$var.coef)))

```

ma2 is significant, we could go with MA(2)

```{r}

ts.plot(fit_ma$residuals)

acf(fit_ma$residuals)

Box.test(fit_ma$residuals, lag = max.lag, type = "Ljung-Box")

```

no significant structures in residual acf plot, and p-value=0.4825>0.05, we cannot reject the null hypothesis. Residuals are white noise, and the model is validated.
```

```{r}

fit_ar <- Arima(index_ts, order = c(2,1,0), seasonal = c(0,0,0))

fit_ar

abs(fit_ar$coef/sqrt(diag(fit_ar$var.coef)))

```

ar2 is significant, we could go with AR(2)

```{r}

ts.plot(fit_ar$residuals)

acf(fit_ar$residuals)

Box.test(fit_ar$residuals, lag = max.lag, type = "Ljung-Box")


```

no significant structures in residual acf plot, and p-value=0.353>0.05, we cannot reject the null hypothesis. Residuals are white noise, and the model is validated.

```{r}

BIC(fit_ma)

BIC(fit_ar)

```

BIC(fit_ma) = 1734.179

BIC(fit_ar) = 1736.41

MA(2) model is better, closer to the truth, given a smaller BIC

```{r}
```

````
forecast_MA <- predict(fit_ma, n.ahead = 12) #how many lags ahead we want to forecast
names(forecast_MA)
expected <- forecast_MA$pred
```


```{r}
lower <- forecast_MA$pred - qnorm(0.975)*forecast_MA$se
upper <- forecast_MA$pred + qnorm(0.975)*forecast_MA$se
cbind(lower, expected, upper)
```


```{r}
plot.ts(index_ts, xlim=c(1992, 2018), main = "Forecast: MA", xlab = "Year")
lines(expected, col = "red")
lines(lower, col = "blue")
lines(upper, col = "blue")
```


```{r}
forecast_AR <- predict(fit_ar, n.ahead = 12) #how many lags ahead we want to forecast
names(forecast_AR)
expected <- forecast_AR$pred
```
````
````

````
```{r}
lower <- forecast_AR$pred - qnorm(0.975)*forecast_AR$se
upper <- forecast_AR$pred + qnorm(0.975)*forecast_AR$se
cbind(lower, expected, upper)
```


```{r}
plot.ts(index_ts, xlim=c(1992, 2018), main = "Forecast: AR", xlab = "Year")
lines(expected, col = "red")
lines(lower, col = "blue")
lines(upper, col = "blue")
```


```{r}
y <- index_ts
S <- round(0.75*length(y)) # using 75% of data as training set
h <- 1 # forecasting window
errorMA.h <- c() # Initialization (empty vector)
for (i in S:(length(y)-h))  # Expanding Window Forecast (MA)
{
  mymodel.sub <- Arima(y[1:i], order = c(0,1,2))
  predict.h <- predict(mymodel.sub, n.ahead = h)$pred[h]
  errorMA.h <- c(errorMA.h, y[i+h] - predict.h)
}
````
````

errorAR.h <- c()

for (i in S:(length(y)-h)) # Expanding Window Forecast (AR)

{

  mymodel.sub <- Arima(y[1:i], order = c(2,1,0))

  predict.h <- predict(mymodel.sub, n.ahead = h)$pred[h]

  errorAR.h <- c(errorAR.h, y[i+h] - predict.h)

}

#cbind(errorMA.h, errorAR.h)

```

We first look at mean absolute error

```{r}

MAE1 <- mean(abs(errorMA.h))

MAE1

MAE2 <- mean(abs(errorAR.h))

MAE2

dm.test(errorMA.h, errorAR.h, h = h, power = 1)

```

p-value = 0.436 > 0.05, we fail to reject the null hypothesis, therefore there is no evidence to prove that two errors from MA(2) model and AR(2) model are different.
```

We could also look at mean squared error

```{r}

MSE1 <- mean(errorMA.h^2)

MSE1

MSE2 <- mean(errorAR.h^2)

MSE2


dm.test(errorMA.h, errorAR.h, h = h, power = 2)

```

p-value = 0.791 > 0.05, we fail to reject the null hypothesis, again there is no evidence to prove that two errors from MA(2) model and AR(2) model are different.

MULTIVARIATE ANALYSIS

CUT TIME PERIOD FOR BOTH TIME SERIES to avoid recession

```{r}

index_ts <- ts(CS$Index, frequency = 12,start=c(1992,1),end=c(2007,12))

retail_ts <- ts(RFS, frequency=12, start=c(1992,1),end=c(2007,12));

ts.plot(index_ts);ts.plot(retail_ts)


#below plotting both time series in same does not work since they have different

#magnitude y values

#ts.plot(index_ts,retail_ts, col=c("black", "red"))

#legend("topright", legend = c("index", "retail"), col = c("black", "red"), lty = 1)
```

```

CHECK FOR STATIONARITY

The plot for the retail sales dataset shows that it has a trend. In general, the retail sales increases over time. We will perform a unit root test on the retail sales time series to determine if it is stationary or not based on the p value we get. If the p value of the time series retail sales is > 0.05 then we can say that the original time series is not stationary and we must do some changes to make it stationary.

```{r}

max.lag<-round(sqrt(length(retail_ts)))

CADFtest(retail_ts, type= "trend", criterion= "BIC", max.lag.y=max.lag)

```

Since original retail time series is not stationary (p value is 0.7668 which is greater than 0.05 in CADF test), we cannot reject the null hypothesis that the time series object is stationary.

Since the retail food sales object retail_ts is not stationary, we will try the log of the retail_ts and see if that will give a stationary object.

Log of retail food sales:

```{r}

ts_sales_log <- log(retail_ts)

ts.plot(ts_sales_log)

```

There is still a trend on the log(retail_ts), but we will confirm this with the CADF test.
```

```{r}
max.lag<-round(sqrt(length(ts_sales_log)))

CADFtest(ts_sales_log, type= "trend", criterion= "BIC", max.lag.y=max.lag)
```

The p-value for the stationarity test on the log(retail_ts) object is 0.1734, therefore we cannot reject the null hypothesis that the object is not stationary.

We will try going in differences on the original retail_ts object.

```{r}
ts_sales_diff <- diff(retail_ts)

ts.plot(ts_sales_diff)
```

Our plot no longer shows a trend, which is good. We will confirm this with a CADF test

```{r}
max.lag<-round(sqrt(length(ts_sales_diff)))

CADFtest(ts_sales_diff, type= "drift", criterion= "BIC", max.lag.y=max.lag)
```

The original time series dataset in differences shows no trend. Furthermore, the unit root test results in a p-value of less than 2.2e-16 which is less than significant level 0.05. Therefore we can reject our null hypotehsis and state that the retail sales object is stationary in differences.

We will now perform unit root test on consumer sentiment index - index_ts to make sure that this dataset is also stationary. The plot of the consumer sentiment dataset - index_ts showed no trend, so we apply drift in the test.

```{r}
```

max.lag <- round(sqrt(length(index_ts)))

CADFtest(index_ts, type = "drift", criterion = "BIC", max.lag.y = max.lag)

```

Since our p value is < 0.05, our time series for consumer sentiment is not statioanry, go in differences as next part will show.

```{r}

dindex_ts <- diff(index_ts)

ts.plot(dindex_ts)

max.lag<-round(sqrt(length(dindex_ts)))

CADFtest(dindex_ts,type="drift",max.lag.y=max.lag,criterion="BIC")

```

Make the cross-correlogram of consumer sentiment index and retail food sales. Is there a significant instantaneous correlation?

Which time series seems to contain predictive power for the other time series?

```{r}

ccf(x = dindex_ts, y =ts_sales_diff,lag.max=30)

```

TRY DISTRIBUTED LAG MODEL

```{r}

lag <- 4 # based on cross-correlogram

ts_sales_diff_data <- embed(ts_sales_diff, dimension = lag + 1)

dindex_ts_data <- embed(dindex_ts, dimension = lag + 1)

#update model we are testing

fit_dl <- lm(ts_sales_diff_data[, 1] ~ dindex_ts_data) #yt is in first column [,1] [row,col]

acf(fit_dl$residuals)

Box.test(fit_dl$residuals, lag = max.lag, type = "Ljung-Box")

```

Residual plots show no significant langs and box test result shows a p value < 0.05 for the residuals of the distributed lag model. Therefore, we cannot validate the distributed lag model.

TRY ADL MODEL

```{r}

lag <- 1 # based on cross-correlogram   #lag=2 is validated, then try lag=1

ts_sales_diff_data <- embed(ts_sales_diff, dimension = lag + 1)

dindex_data <- embed(dindex_ts, dimension = lag + 1)

#fit_adl1 <- lm(ts_sales_diff_data[, 1] ~ ts_sales_diff_data[, -1] + dindex_data[,-1]) #yt is in first column [,1] [row,col]

fit_adl1 <- lm(dindex_data[, 1] ~ ts_sales_diff_data[, -1] + dindex_data[,-1])

acf(fit_adl1$residuals)

Box.test(fit_adl1$residuals, lag = round(sqrt(length(fit_adl1$residuals))), type = "Ljung-Box")

```

Residual plot shows no significant lags and box test results in p value > 0.05 for the residuals of the autoregressive distributed lag model resulting in valid model.
```

INVESTIGATE GRANGER CAUSALITY

```{r}

#fit_adl2_nox <- lm(ts_sales_diff_data[, 1] ~ ts_sales_diff_data[, -1])

fit_adl2_nox <- lm(dindex_data[, 1] ~ dindex_data[, -1])

anova(fit_adl1, fit_adl2_nox)


```

####Based on p value > 0.05 we cannot reject null hypotehsis. NO GC

Specify and estimate a VAR model + look at the impulse response functions

```{r}

mydata <- cbind(dindex_ts,ts_sales_diff)

VARselect(mydata) # BIC (SC): 1

```

For automatic lag selection, AIC selects 2 lags and BIC (SC) selects 1 lag. When we have a situation where both methods are selecting different lags, we will go with the lag that BIC selects since BIC penalizes higher order terms, in this case we select 1.

```{r}

varfit <- vars:::VAR(mydata, p = 1)

varresid <- resid(varfit)

par(mfrow=c(2,2))

acf(varresid[,1])
```

acf(varresid[,2])

ccf(varresid[,1],varresid[,2],lag.max=10)

#we wnat to validate model so look acf. For 2nd time series we have one sig lag, for ccf we have one significant lag.

```

From the CCF plot, we can see there is only one significant lag. For ACF plot of consumer sentiment, there is no singnificant lag, but for ACF plot of retailer sales, there are several significant lags. From the observation above, we need to validate VAR(1) further quantitatively.

multivariate white noise test to assess the validity of this model quantitatively.

```{r}

library(MTS)

mq(varresid, lag = floor(sqrt(dim(varresid)[1])))

#we are interested in last line

#null hypothesis R1...R13 = 0. Fail to reject first  are = 0. we validate this model.

```

From the quantitative test, VAR(1) is validated since p value 0.06 is greater than 0.05.

look at the impulse response functions

```{r}

irf_var <- irf(varfit, ortho = FALSE, boot = TRUE)

plot(irf_var)

#below comments are from discussion, to replace with project data comments

#k*k --> 2*2 = 4 plots. we want the bands to be above and below the 0 line in this case

#it is not happening. #borderline on 1 and 4 and none elsewhere.

#cons - 1 unit shock in log diff consumption

#effect of shocks in industrial production: a little bit more going on. IP 1 unit shock

#sig imact at lag = 1 since band below 0 a neg effect, pos effect at lag=4 and lag=8.

```

For the effect of one unit-shock in consumer sentiment on itself, there is no significant lag. For the retailer sales on itself, there is negative significance on lag 1.

Coitegration

Validate that the retail sales is I(1)

```{r}

CADFtest(diff(retail_ts), type = "drift", criterion = "BIC", max.lag.y = max.lag) # Reject H0 -> log(retail sales) is I(1)

```

Validate that the consumer index is I(1)

```{r}

CADFtest(diff(index_ts), type = "drift", criterion = "BIC", max.lag.y = max.lag) # Reject H0 -> log(consumer index) is I(1)

```

```{r}
data=cbind(retail_ts,index_ts)

#fit_ci <- lm(retail_ts~index_ts)

fit_ci <- lm(index_ts~retail_ts)

res_fit_ci <- fit_ci$residuals

max.lag <- round(sqrt(length(res_fit_ci)))

library(CADFtest)

CADFtest(res_fit_ci, type = "drift", criterion = "BIC", max.lag.y = max.lag)
```

ADF(0) = -2.6533 is greater than -3.41. Accept null, no cointegration

No cointegration in either direction

```{r}
VARselect(data, type = "const") # BIC/SC selects p = 2

```

From BIC, we get p = 2

```{r}
library(urca)

trace_test <- ca.jo(data, type = "trace", K = 2, ecdet = "const", spec = "transitory")

```
summary(trace_test)
```

test 10pct 5pct 1pct

r <= 1 | 11.94 7.52 9.24 12.97

r = 0 | 67.75 17.85 19.96 24.60

67.75 > 17.85 and 11.94 > 9.24, we reject r <= 1 and reject r = 0

No coitegration