THE PENNSYLVANIA STATE UNIVERSITY
SCHREYER HONORS COLLEGE

DEPARTMENT OF ECONOMICS

# MODELLING MAJOR ECONOMIC INDICATORS VIA MULTIVARIATE TIME SERIES ANALYSIS

XUANHAO ZHANG
SPRING 2017

A thesis
submitted in partial fulfillment
of the requirements
for baccalaureate degrees
in Mathematics, Economics and Statistics
with honors in Economics

Reviewed and approved* by the following:

Patrik Guggenberger
Professor of Economics
Thesis Supervisor

Russell Chuderewicz
Lecturer of Economics
Honors Adviser

* Signatures are on file in the Schreyer Honors College.

# ABSTRACT

Gross domestic product (GDP) is a measure of the market value of all goods and services produced by the country. It is one of the most important indicators to measure the performance of a nation's economy. The indicator has strong influence on the currency market and monetary policy of the central bank. Policy-makers rely on GDP growth to support and justify their decisions. A better forecasted result can help to formulate a more effective policy to keep the economy prosperous. Time series models have been increasingly prominent as forecasting tools in economics. This paper will focus on the predictability of quarterly real GDP growth in the United States. It aims to provide a reasonable model for the GDP growth based on other economic variables via multivariate time series analysis.

Key Word: GDP, Economics, Forecasting, Time Series Analysis, VARMA, ARMA

# TABLE OF CONTENTS

LIST OF FIGURES ..... iii

LIST OF TABLES ..... iv

ACKNOWLEDGEMENTS ..... v

Chapter 1 Introduction ..... 1

Chapter 2 Literature Review ..... 3

Chapter 3 Data and Background ..... 5

Chapter 4 Variable Analysis ..... 7

Chapter 5 Multivariate Time Series ..... 15

Chapter 6 Model and Forecasting ..... 22

Chapter 7 Conclusion ..... 32

BIBLIOGRAPHY ..... 34

# LIST OF FIGURES

Figure 1 Federal Funds Rate and U.S. GDP Growth ..... 8

Figure 2 Unemployment Rate and GDP Growth ..... 9

Figure 3 Canada GDP and U.S. GDP Growth ..... 11

Figure 4 Australia GDP Growth and U.S. GDP Growth ..... 14

Figure 5 VARMA (1,1) Simulation ..... 20

Figure 6 ACF and PACF of Series U.S. GDP Growth ..... 24

# LIST OF TABLES

Table 1 ADF Unit Roots Test ..... 23

Table 2 ARMA Coefficients ..... 25

Table 3 VARMA Models and AICs ..... 26

Table 4 AR (1) Coefficient Matrix ..... 27

Table 5 MA (1) Coefficient Matrix ..... 27

Table 6 MA (2) Coefficient Matrix ..... 28

Table 7 Quarterly U.S. GDP Growth Forecast in 2015 by ARMA Model ..... 29

Table 8 Quarterly U.S. GDP Growth Forecast in 2015 by VARMA Model ..... 30

Table 9 Quarterly U.S. GDP Growth Forecast in 2016 by ARMA Model ..... 30

Table 10 Quarterly U.S. GDP Growth Forecast in 2016 by VARMA Model ..... 31

# ACKNOWLEDGEMENTS

I would like to thank the following people for their tremendous support and making this thesis possible.

Dr. Patrik Guggenberger, my thesis supervisor. Over the past years, he has been a great mentor and has provided countless valuable advices. Without his guidance, I would not have completed this thesis.

Dr. Russell Chuderewicz, my honors adviser. He has been an inspiring role model and his lectures really lead me into economics. Without his influence, I would not have chosen to pursue my area of honors in Economics.

# Chapter 1
# Introduction

Given the significant influence of GDP growth, many forecasting methods have been implemented trying to capture the percentage change in real GDP. However, due to its complex and erratic nature, many methods fail to provide a meaningful forecasting. There are many models built including both parametric and non-parametric methods. Among all those models, time series models have become benchmarks, especially the ARIMA and VAR models.

Because GDP is a time sequential data, a univariate time series model can be easily constructed using ARMA or ARIMA. However, in order to utilize additional information and make better predictions, I will use a technique called multivariate time series analysis. VAR and VARMA models will be suitable to analyze multivariate time series. The model will have multiple variables including GDP by country and other variables have effect on GDP growth. All the data are time series, so I can construct a time series model with multiple variables. The goal of this paper is to provide a close estimate of the quarterly GDP growth in the United States from my model and compare the estimates with the data.

I will select some variables to construct a multivariate time series model by looking at their correlations with U.S. GDP growth. In order to select the most appropriate model, multiple models will be fitted. The model with the lowest AIC and BIC will be selected to make forecasting.

The rest of this paper will be constructed in the following order: Chapter 2 includes a literature review. Chapter 3 shows and explains the data used for building models. Chapter 4 includes variables and their analysis to build models. It will discuss the correlation among real GDP growth in the U.S., Federal Funds Rate, Unemployment Rate, and GDP growth for Canada and Australia. Chapter 5 introduces the methodology of multivariate time series. Chapter 6 is actual model fitting and forecasting. Chapter 7 provides the conclusion and summary of the obtained results.

# Chapter 2

# Literature Review

Since GDP growth is such an influential indicator, there is a large number of papers focusing on its predictability in many countries of large economy. This paper focuses on the predictability of quarterly GDP growth in the United States. However, due to the similarity of utilizing multivariate time series models, it is worth exploring papers using such models from other countries. One of this kinds of papers is written by Onwukwe and Nwafor (2014). It focuses on modelling major economic indicators in Nigeria, which includes currency, exchange rate, external reserve, price deflator, GDP and money supply. All of their data were obtained from the Central Bank of Nigeria. Because all six variables are quarterly time series from 1981 to 2010, multivariate time series models are employed to estimate the major economic indicators. To be specific, a vector autoregressive (VAR) model is implemented in the paper. The results show that lending rates and inflation have a negative impact on the output. It is quite intuitive, because when lending rates and inflation increase, the economy of the country is more likely to be in trouble. The results also suggest that exchange rate has significant influence on price level and money supply responds to positive shocks in price level. The paper truly develops a stable autoregressive model and the results explore the relationships among six. However, it fails to provide a concise comparison between forecasting values and the reality.

In the thesis of Holod (2000), the author also constructs a vector autoregressive model to estimate the relationships among CPI, money supply and exchange rate in Ukraine. The author

uses monthly data from 1995 to 1999 obtained from National Bank of Ukraine and UEPLAC. To avoid nonstationary of the time series, all three variables are taken their logarithms and to select time lags, AIC approach is implemented. The author's techniques are very simple and straightforward. It is worth trying to solve problems by keeping everything simple, and I will use similar techniques while building my model. The results do suggest causality in the model and a positive shock to the exchange rate will rise the price level, which agrees with the paper written by Onwukwe and Nwafor (2014). The effect of money supply is controversial and less significant in Holod's paper. The structure of this paper is very well organized, and each steps are clearly explained by the author. However, it provides insufficient information about the output of the model, which exploits very little of the VAR model. One of the main application of the VAR model is forecasting, which is missing in Holod's paper.

Such papers involving multivariate time series analysis have derived many variations and contain almost every variable in macroeconomics. The majority of such papers focus more on the theory and interpretation of relations among variables. The economic variables selected are usually from one country. In my paper, I will focus more on the predictability of the model, and I also include other countries' economic indicators as variables to build my model. Both papers reviewed uses VAR model, which is the most popular in econometrics. However, I will include a moving average term to build a VARMA model. It will make the model slightly more complex, but it should capture more information from the data. With the difference in the models and variables, and a more focus on predictability, it remains interesting to model the situation in the United States.

# Chapter 3
# Data and Background

The purpose of GDP growth is to measure the aggregate production within a country and get a sense of how the overall economy behaves. Therefore, GDP growth is directly related to other economic variables in the country. With the recent upswing of globalization, other countries' economic variables are also important to consider.

The variable we want to predict, so called the response is GDP growth of the U.S., and the other four predictors are Federal Funds Rate, Unemployment Rate, GDP growth of Canada and GDP growth of Australia. The data for all five variables are seasonal adjusted quarterly time series data from Jan. 1st, 1962 to Oct. 1st, 2014. All of the data are retrieved from (FRED) Federal Reserve Electronic Data of Federal Reserve Bank of St. Louis.

Federal Funds Rate and Unemployment Rate, as key economic indicator, reflect the overall activity of the money market and labor market. Federal Funds Rate is a benchmark interest rate for banks to borrow from the Federal Reserve bank. When the economy is booming, the Fed will want to set the rate high to cool down the inflation and decrease money supply. When the economy is in recession, the Fed will want to set the rate low to increase money supply and encourage borrowing, so it will stimulate the economy. Therefore, Federal Funds Rate closely relating with macroeconomics should be include in the analysis. Unemployment Rate is a direct realization of the current economic situation. A booming economy will result in a low

unemployment rate and vice versa. With the simple deduction from the definition of the variable, Federal Funds Rate and Unemployment Rate should be negatively related to GDP growth of the U.S.

Taking into consideration of globalization and the increasingly international trade activities, variables from outside of the U.S. are also significant. Canada is the main trading partners with the U.S. among western countries. Therefore, it is crucial to have Canada GDP growth in the model. The GDP growth of Germany and Europe should also be included given their huge impact on the U.S. economy. However, the data for Europe only traced from 1995, and the data for Germany is from 1991. Comparing with other variables with data from 1962, the data for Germany and Europe have so much missing points that they cannot be included in the model. Australia is included because of the coherence and convenience of the data. Because the United States is the largest investor in Australia, they may have strong relationship. China as the largest trading partner and second largest economy should be considered, but due to insufficient data length, it is infeasible to include it into the model. From a globalization point of view, both the GDP growth in Canada and Australia should be positively related to the GDP growth in the U.S.

# Chapter 4

# Variable Analysis

## 4.1 Federal Funds Rate

As we already inferred from economic theory, Federal Funds Rate as a primary tool for monetary policy, should be negatively related to the GDP growth. To further evaluate the relationship between these two variables, the correlation is calculated using R software. As a result, the correlation between Federal Funds Rate and U.S. GDP growth is -0.06433916, which is fairly a small correlation. Although this value is not statistically significant from the hypothesis test, it does not contradict the theory. To have a better visual effect, the quarterly data of the two series are plotted from Jan. 1$^{st}$ 1962 to Oct. 1$^{st}$ 2014 in Figure 1. The red line represents Federal Funds Rate and the green line stands for U.S. GDP growth. As we can see, there is a negative relation at the start of 1960s, around 1980 and from 2000 to 2010. Therefore, it is still meaningful to include the Federal Funds Rate in the predictors.

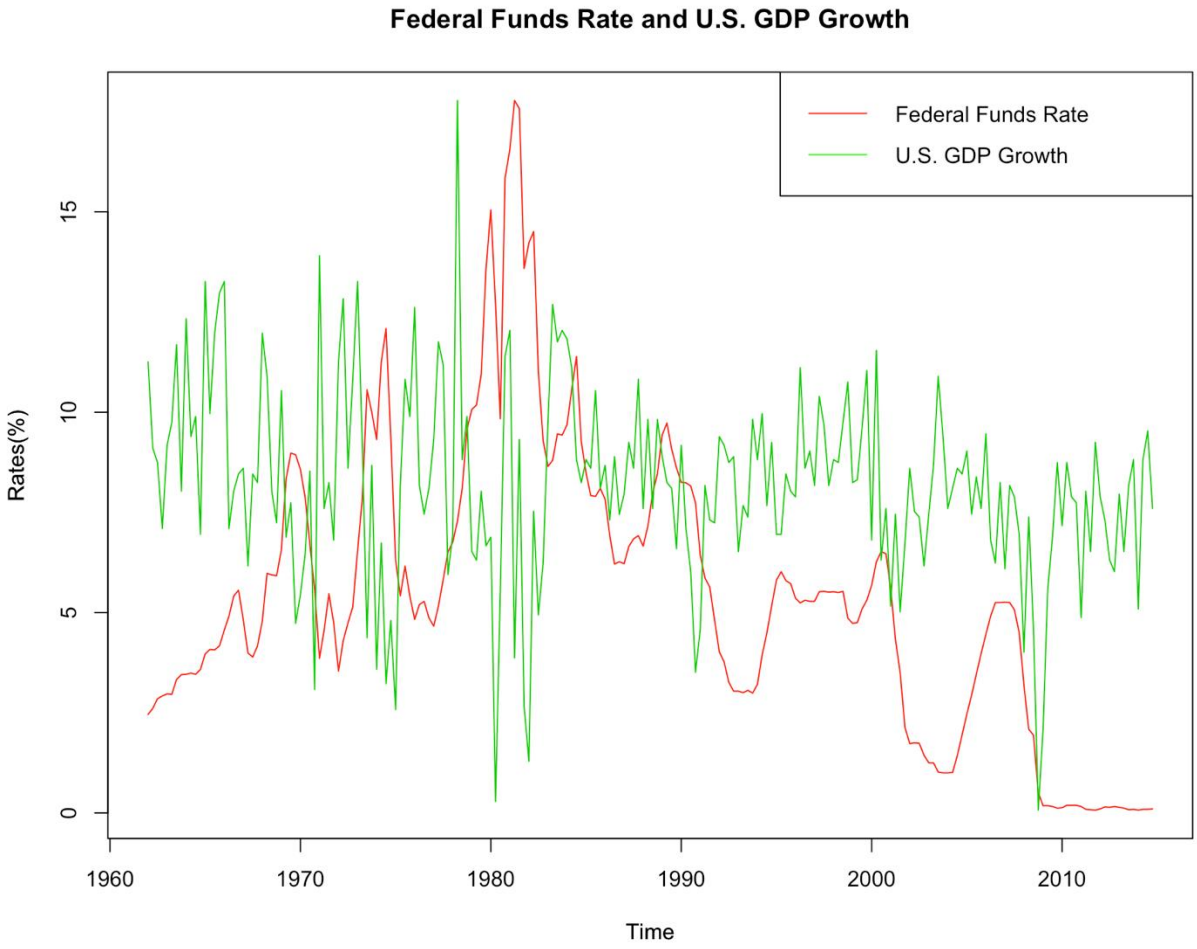



**Figure 1 Federal Funds Rate and U.S. GDP Growth**

## 4.2 U.S. Unemployment Rate

The Unemployment Rate, which describes the activity of the labor market, is one of the Fed's key target. It is also an important indicator of the overall wellness of the economy. When the economy is booming, more labors are demanded, so the rate will be low. It should be negatively related to the GDP growth.

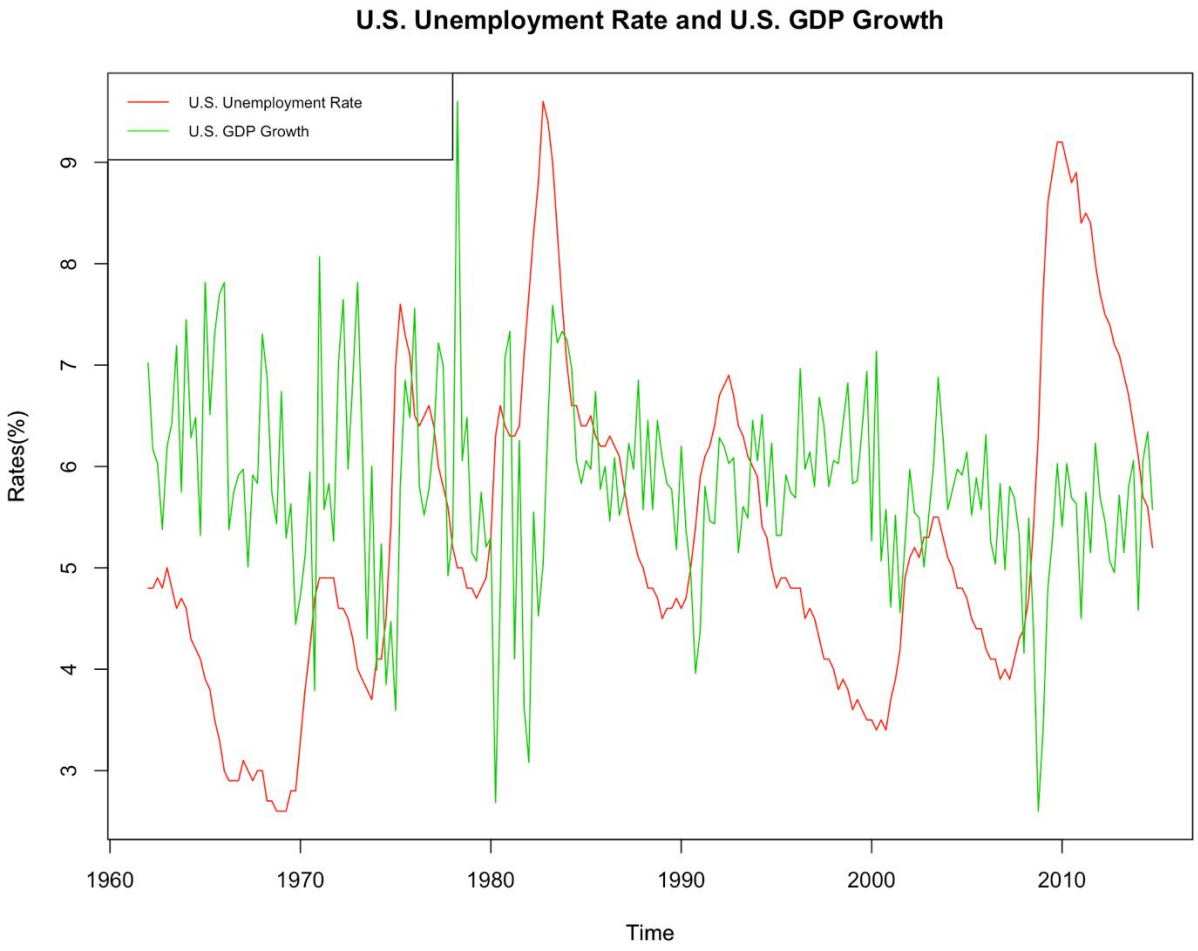



Figure 2 Unemployment Rate and GDP Growth

As Figure 2. suggests, the two variables diverge at the beginning of 1960s and there is a huge discrepancy around the year 2008. The Great Recession in 2008 really show the negative relationship between unemployment rate and GDP growth, as unemployment rate increase dramatically and GDP growth goes down to almost zero. We can see that this relationship magnifies when there is big economic shock. The correlation is -0.120725, which is still a weak relation, but it is better than the Federal Funds Rate. It also agrees with our assumption.

## 4.3 Canada GDP Growth

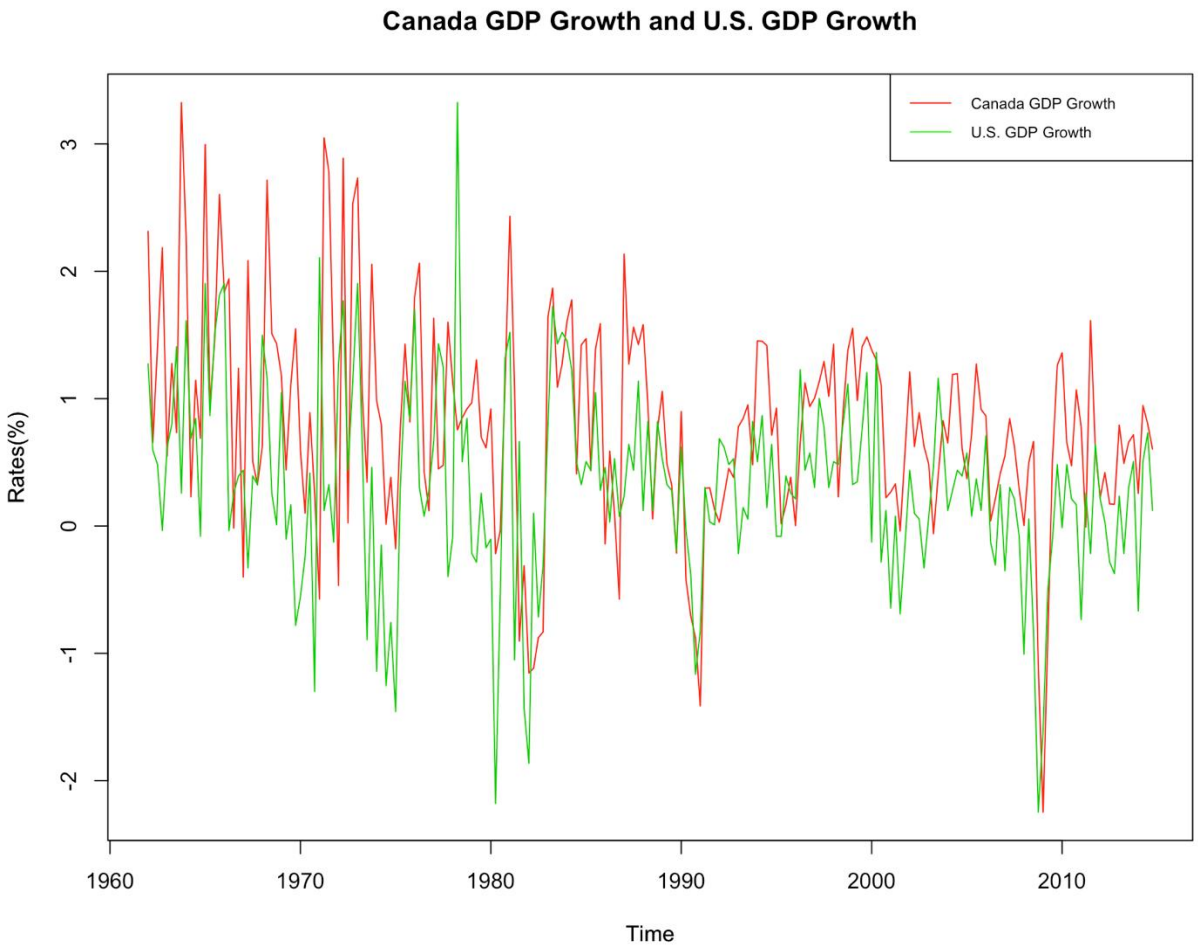



Figure 3 Canada GDP and U.S. GDP Growth

After analyzing the consequences of globalization and acknowledging the fact that Canada is the second largest trading partner with the United States, it is crucial to include the GDP growth of Canada as a predictor in our model. To evaluate and testify the correlation of these two variables, Figure 3. is plotted above. There are many overlaps between the two lines,

which suggest a positive relationship. The correlation is 0.4831004 from R output, and it is statistically significant. The result confirms the effect of the globalization and shows a strong positive correlation. Therefore, the GDP growth of Canada should definitely be one of the predictors in the model.

## 4.4 Australia GDP Growth

Following with the same concept of GDP growth of Canada, it is reasonable to further explore the situation in Australia. Although Australia is not a major trading partner with the United States, it is still a western style economy with good data disclosure. Figure 4. shows no clear correlation and the output from R is -0.1308293. It is not statistically significant and also a negative value, which contradict with our assumption. Thus, it is not necessary to include Australia GDP growth in the model.

To sum up this chapter, we evaluated four variables and their correlation with the U.S. GDP growth. Three of the four variables agree with our assumptions, so they will be the predictors of the model. Consequently, U.S. GDP growth, Federal Funds Rate, Unemployment Rate, and Canada GDP growth will be the four variables in the multivariate time series model.

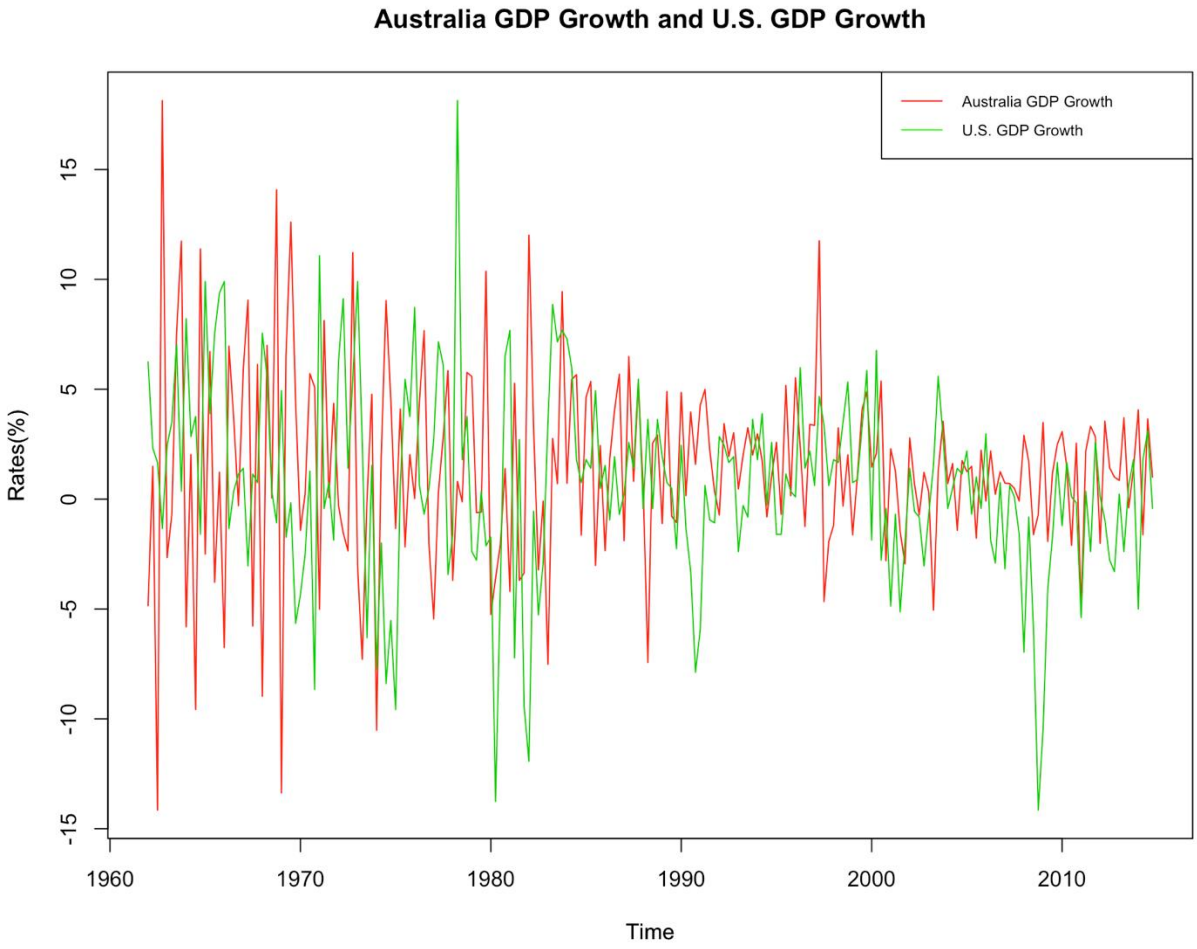



**Figure 4 Australia GDP Growth and U.S. GDP Growth**

# Chapter 5

# Multivariate Time Series

While univariate time series is to study a single time series, multivariate time series is used to analyze several time series. Multivariate time series is used to improve the accuracy of forecasts for individual time series by including additional information from multiple relevant time series. The notation $Z_t = (z_{1t}, \dots, z_{kt})'$ is expressed as the time series vector at time t. For example, $Z_{1t}$ is the first time series at time t and $Z_{2t}$ is the second time series.

## 5.1 Stationary Multivariate Time Series

Like the univariate case, the vector process $\{Z_t\}$ is stationary if the distribution of the random vectors $(Z_{t1}, Z_{t2}, \dots, Z_{tm})$ and $(Z_{t1+l}, Z_{t2+l}, \dots, Z_{tm+l})$ are the same for any times at $t_1, t_2, \dots, t_m$ and the lags $l = 0, \pm 1, \pm 2, \dots$ Therefore, it means that probability distribution remains the same when the vector process is shifted in time. If a multivariate time series is stationary, then each time series must have constant mean for all $t$, so the mean vector is constant through all time. The mathematical form can be represented as $E[z_t] = \mu$,

where $\mu = (\mu_1, \mu_2, \dots, \mu_k)$. Also, the vectors $Z_t$must have a constant covariance matrix for all $t$. In addition, we have cross covariance for multivariate time series and it is the covariance between $z_{i,t}$and $z_{j,t+l}$ denoted by $\gamma_{ij}(l) = cov[z_{it}, z_{j,t+l}]$ The corresponding cross correlations at lag $l$ are $\rho_{ij}(l) = corr[z_{it}, z_{j,t+l}] = \frac{\gamma_{ij}(l)}{\{\gamma_{ii}(0)\gamma_{jj}(0)\}^{1/2}}$ For a stationary process, the cross covariance only depends on the lag $l$.

## 5.2 Vector Autoregressive-Moving Average (ARMA)

$$(Z_t - \mu) - \sum_{j=1}^{p} \Phi_j (Z_{t-j} - \mu) = a_t - \sum_{j=1}^{q} \Theta_j a_{t-j}$$

$Z_t$ and $\mu$ are k-by-1 matrices, $\Phi_j$ and $\Theta_j$ are k-by-k matrices, and $a_t$ is vector white noise process. This equation can be intuitively explained as we want to model the residual as an ARMA model in vector form. $(Z_t - \mu)$ helps to get the residual which is done by subtract the mean function, and then the right-hand side becomes

$$\sum_{j=1}^{p} \Phi_j (Z_{t-j} - \mu) + a_t - \sum_{j=1}^{q} \Theta_j a_{t-j}$$

which is AR(p) plus MA(q) at their matrix form. Thus $Z_t$ is a Vector ARMA(p,q) process if it satisfies the equation above, regardless of whether $Z_t$ is stationary or not. For a stationary VARMA process, the AR model can also be expressed by an infinite MA process just like univariate case. Then the equation becomes $Z_t = \mu + \sum_{j=0}^{\infty} \Psi_j a_{t-j}$

## 5.3 Nonstationary Vector Autoregressive-Moving Average Models

In practice, a nonstationary behavior is quite common. Differencing can also be used to deal with multivariate nonstationary process just like the univariate case. Therefore, a vector ARIMA model is developed. It means that after differencing each series $z_{it}$ for $d_i$ times to reduce it to a stationary series, then the vector process becomes a stationary VARMA process. For example, a general vector ARMA can be written as $\Phi(B)Z_t = \Theta(B)a_t$ where $B$ is the backshift operator. Then the differenced model for a nonstationary series is $\Phi_1(B)D(B)Z_t = \Theta(B)a_t$ where $D(B) = diag[(1-B)^{d1}, \dots, (1-B)^{dk}]$ is a diagonal matrix, which is used to difference each time series. For the nonstationary component series with integrated order of one, if there exists a vector such that their linear combination is stationary, then they are said to be co-integrated. It means that each individual component $z_{it}$ has some common trend. For example, $z_{1t}$ and $z_{2t}$ are two nonstationary time series. If there exists a non-zero vector $(\alpha, \beta)$ such that $\alpha z_{1t} + \beta z_{2t}$ is stationary, then $z_{1t}$ and $z_{2t}$ are co-integrated.

## 5.4 VARMA Simulation

When simulating a univariate ARMA model, we can understand it by drawing the first value $y_1$ from a normal distribution with mean zero and variance $\sigma^2$, and then use the formula to get $y_2 \dots y_t$. The vector ARMA case is quite similar. The first vector $z_1$ can be draw from a multivariate normal distribution with covariance matrix $\Sigma$, then $z_2 \dots z_t$ are calculated by using the given coefficient matrix $\Phi$ and $\Theta$ from the formula.

The following simulation shown in Figure 5 is a VARMA (1,1) with coefficient matrix

$$\Phi = \begin{pmatrix} 0.2 & 0.3 \\ -0.6 & -0.8 \end{pmatrix}, \Theta = \begin{pmatrix} -0.5 & 1 \\ 2 & -0.5 \end{pmatrix} \text{ and } \Sigma = \begin{pmatrix} 1 & 0.6 \\ 0.6 & 2 \end{pmatrix}$$

The mathematical form is expressed in the following way:

$$\begin{pmatrix} x_t \\ y_t \end{pmatrix} = \begin{pmatrix} 0.2 & 0.3 \\ -0.6 & -0.8 \end{pmatrix} \begin{pmatrix} x_{t-1} \\ y_{t-1} \end{pmatrix} + \begin{pmatrix} -0.5 & 1 \\ 2 & -0.5 \end{pmatrix} \begin{pmatrix} w_{t-1}^x \\ w_{t-1}^y \end{pmatrix} + \begin{pmatrix} w_t^x \\ w_t^y \end{pmatrix}$$

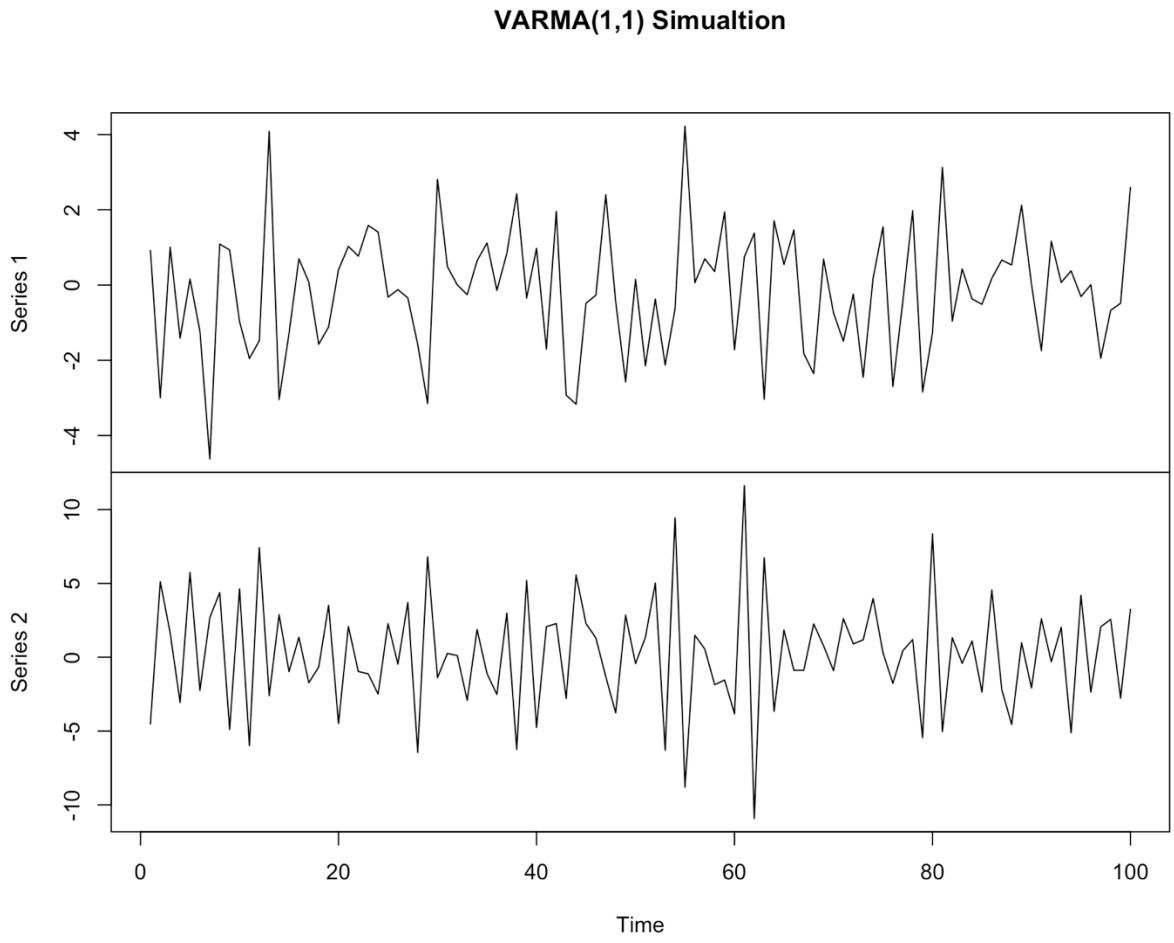



Figure 5 VARMA (1,1) Simulation

Because the variance for series 2 is larger than series 1, we can see that there is more variation in series 2. The coefficient of the AR component in the second series is negative, so we see that the value of series 2 is alternating around zero.

## 5.5 VARMA Model fitting

Suppose now that the simulated model is given, then I will attempt to fit a VARMA model to the simulated model. First, I fit a VARMA (1,0) model and the results are shown below.

AR ( 1 )-matrix

$$\begin{pmatrix} -0.00383 & 0.171 \\ -0.8423 & -0.983 \end{pmatrix}$$

Residuals covariance matrix

$$\begin{pmatrix} 2.048 & -0.63 \\ -0.63 & 3.3173 \end{pmatrix}$$

AIC = 1.93582 BIC = 2.040027

We can see the AR coefficients and residual covariance matrix are not very accurate with our simulated model, but AIC and BIC are not very large. Let us now try a VARMA (1,1) model

AR ( 1 )-matrix

$$\begin{pmatrix} 0.271 & 0.324 \\ -0.670 & -0.947 \end{pmatrix}$$

MA ( 1 )-matrix

$$\begin{pmatrix} 0.502 & 0.5276 \\ 0.531 & -0.0109 \end{pmatrix}$$

Residuals covariance matrix

$$\begin{pmatrix} 1.4059 & -0.66 \\ -0.66 & 3.2694 \end{pmatrix}$$

AIC = 1.585657 BIC = 1.79407

As we can see, the fitted VARMA (1,1) model is closer to our simulated model regarding all coefficient matrices. Both AIC and BIC are smaller than the fitted VARMA (1,0) model. Therefore, AIC and BIC can still be a useful model selection tool in the multivariate models.

# Chapter 6
# Model and Forecasting

## 6.1 Test for Stationarity

Before constructing any time series models, it is always necessary to check if the stationary assumption holds. If a model is built without assessing the stationarity, it is very likely to get a spurious result. The Augmented Dickey-Fuller (ADF) unit roots tests is implemented to check the stationarity of each variable. The alternative hypothesis of the test is when the variable is stationary, so a p-value smaller than 0.05 means stationarity for the tested series. Therefore, Table 1. shows the results of ADF unit roots test. The result shows that all variables are stationary except the unemployment rate. To avoid nonstationarity and keep the unemployment rate, the first order difference is taken to remove the trend. After taking the integration of order 1, the test shows a p-vaule of 0.01 for the differenced unemployment rate series.

**Table 1 ADF Unit Roots Test**

| Variable | P-Value |
|---|---|
| U.S. GDP Growth | 0.01 |
| Federal Funds Rate | 0.03279 |
| U.S. Unemployment Rate | 0.1164 |
| Canada GDP Growth | 0.01 |

Finally, after transforming the raw data into a prepared data frame, the next step will be model building and selection. As I mentioned earlier, I will build several Vector Autoregressive Moving Average (VARMA) models, and then select an ideal model based on AIC to do prediction. However, before utilizing the complex multivariate time series models, it is useful to construct an ARMA model first.

## 6.2 ARMA Model

Building a model by only using one variable and its previous values is not very meaningful, compared with multivariate cases, as discussed in Chapter 5. However, it is straightforward and provides a good comparison for the main model. Autocorrelation function (ACF) and Partial autocorrelation function (PACF) are plotted to select the appropriate ARMA model.

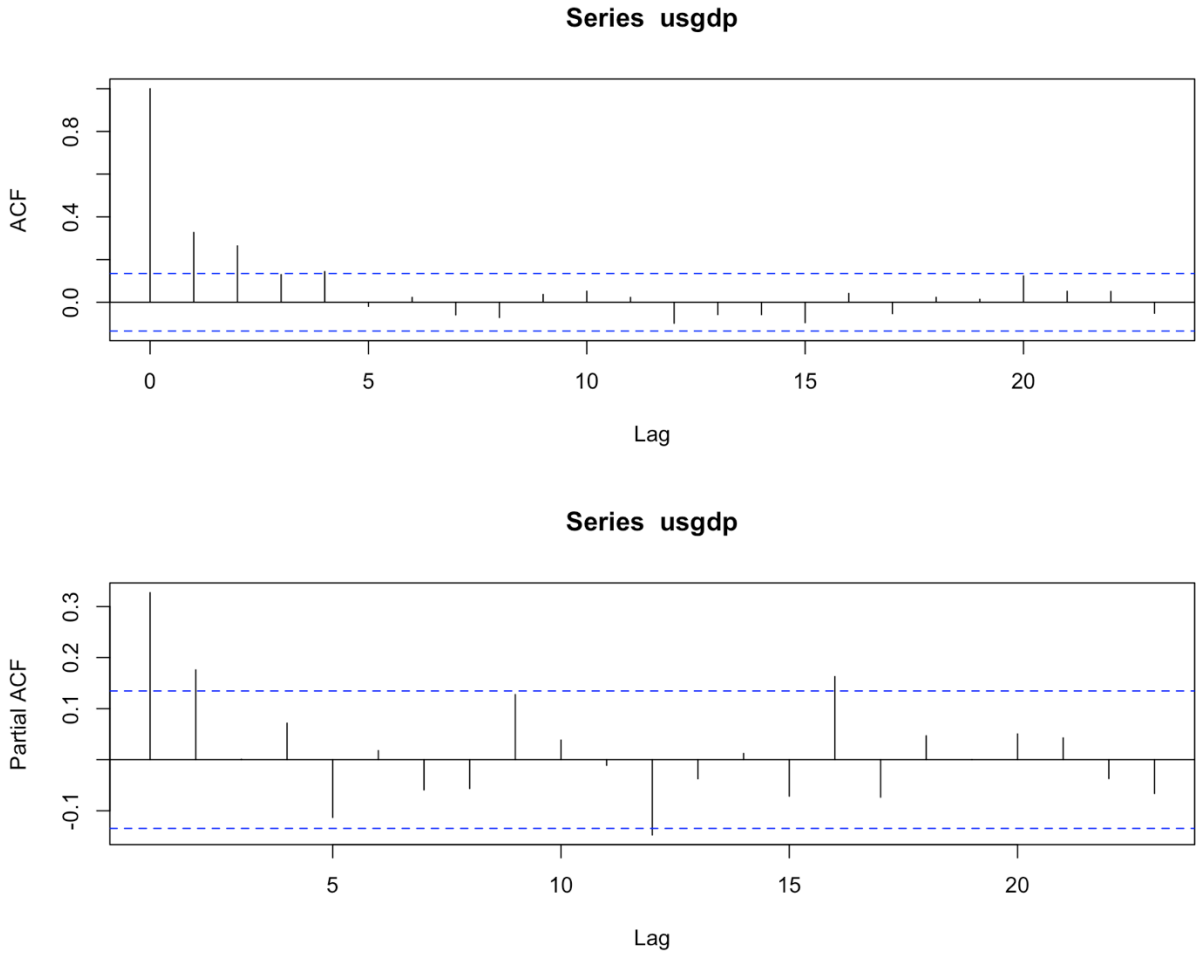



Figure 6 ACF and PACF of Series U.S. GDP Growth

As Figure 6. suggests, there is a clear cut at lag 2 in the PACF plot, so it is reasonable to build a AR (2) model. Using R to fit the data into the model, we get the following results.

**Table 2 ARMA Coefficients**

| ARMA (2,0,0) | AR 1 | AR 2 | Intercept |
|---|---|---|---|
| Coefficients | 0.2702 | 0.1760 | 3.1333 |
| Standard Error | 0.0676 | 0.0676 | 0.3891 |

All of the coefficients are statistically significant, so the ARMA (2,0,0) model or equivalently AR (2) model fits fairly well to the data set.

## 6.3 VARMA Model

The concepts of multivariate time series, especially VARMA models, have already been discussed in Chapter 5. Thanks to the computing power, it is not hard to run several models in R with MTS package. There are many ways to evaluate a model. In this paper, I use AIC as criteria to select models because of its convenience and well-acceptance. Table 3. are the results of VARMA models with different AR(p) and MA(q) lags.

**Table 3 VARMA Models and AICs**

| VARMA | AIC |
|---|---|
| p=0, q=1 | 1.236511 |
| p=1, q=0 | -1.809754 |
| p=1, q=1 | -1.981371 |
| p=1, q=2 | -2.10176 |
| p=2, q=1 | -2.088561 |
| p=2, q=2 | -2.045572 |

Although it is not the best way to find the lags of AR and MA components of VARMA models, it provides a clear comparison of these models based on AIC. From the above results, a VARMA (1, 2) model is the best and the AR lag is 1, MA lag is 2. The coefficient matrices are shown in the following tables.

**Table 4 AR (1) Coefficient Matrix**

| | | | |
|---|---|---|---|
| 0.3378 | 0.09719 | 1.5959 | 0.464 |
| -0.0330 | 0.97839 | 0.2895 | -0.564 |
| 0.1105 | -0.01778 | 0.6578 | 0.904 |
| -0.0126 | 0.00324 | 0.0465 | 0.651 |

**Table 5 MA (1) Coefficient Matrix**

| | | | |
|---|---|---|---|
| 0.31376 | 0.2049 | 1.4686 | 6.8505 |
| -0.05434 | -0.1691 | 0.1896 | 0.2456 |
| 0.02734 | -0.0569 | 0.6102 | 2.0076 |
| 0.00922 | -0.0754 | 0.0405 | 0.0965 |

**Table 6 MA (2) Coefficient Matrix**

| | | | |
|---|---|---|---|
| -0.08301 | 1.5772 | -0.44319 | 0.906 |
| -0.00645 | 0.2015 | 0.03021 | 0.750 |
| -0.02643 | 0.1291 | 0.04772 | -0.154 |
| 0.00537 | -0.0156 | -0.00505 | -0.071 |

The model is expressed in Matrix form below:

U=U.S. GDP; F=FED FUNDS RATE; C=CANADA GDP; UN=UNEMPLOYMENT RATE

$$\begin{pmatrix} U \\ F \\ C \\ UN \end{pmatrix} = \begin{pmatrix} 0.3378 & 0.09719 & 1.5959 & 0.464 \\ -0.0330 & 0.97839 & 0.2895 & -0.564 \\ 0.1105 & -0.01778 & 0.6578 & 0.904 \\ -0.0126 & 0.00324 & 0.0465 & 0.651 \end{pmatrix} \begin{pmatrix} U_{t-1} \\ F_{t-1} \\ C_{t-1} \\ UN_{t-1} \end{pmatrix}$$

$$+ \begin{pmatrix} 0.31376 & 0.2049 & 1.4686 & 6.8505 \\ -0.05434 & -0.1691 & 0.1896 & 0.2456 \\ 0.02734 & -0.0569 & 0.6102 & 2.0076 \\ 0.00922 & -0.0754 & 0.0405 & 0.0965 \end{pmatrix} \begin{pmatrix} w_{t-1}^{U} \\ w_{t-1}^{F} \\ w_{t-1}^{C} \\ w_{t-1}^{UN} \end{pmatrix}$$

$$+ \begin{pmatrix} -0.08301 & 1.5772 & -0.44319 & 0.906 \\ -0.00645 & 0.2015 & 0.03021 & 0.750 \\ -0.02643 & 0.1291 & 0.04772 & -0.154 \\ 0.00537 & -0.0156 & -0.00505 & -0.071 \end{pmatrix} \begin{pmatrix} w_{t-2}^{U} \\ w_{t-2}^{F} \\ w_{t-2}^{C} \\ w_{t-2}^{UN} \end{pmatrix}$$

## 6.4 Predictability of U.S. GDP Growth

After selecting the appropriate models, the next step is to make forecasting with the fitted model. Making a prediction is never an easy task. However, it still provides some information and could help us better understand the future. In this paper, I make a forecast of four steps ahead about the quarterly GDP growth in the United States. Therefore, I will have all four quarters in 2015 predicted from the previous model. Then I can make a comparison of the actual data released. The results are in following tables.

**Table 7 Quarterly U.S. GDP Growth Forecast in 2015 by ARMA Model**

| Date | Predicted Value | Standard Error | Actual Value |
|---|---|---|---|
| 1st Quarter 2015 | 3.236648 | 3.153641 | 2.0 |
| 2nd Quarter 2015 | 3.014593 | 3.266729 | 2.6 |
| 3rd Quarter 2015 | 3.119422 | 3.359768 | 2.0 |
| 4th Quarter 2015 | 3.108670 | 3.379224 | 0.9 |

**Table 8 Quarterly U.S. GDP Growth Forecast in 2015 by VARMA Model**

| Date | Predicted Value | Standard Error | Actual Value |
|---|---|---|---|
| 1st Quarter 2015 | 3.617 | 2.929 | 2.0 |
| 2nd Quarter 2015 | 2.372 | 3.452 | 2.6 |
| 3rd Quarter 2015 | 1.856 | 3.585 | 2.0 |
| 4th Quarter 2015 | 1.516 | 3.653 | 0.9 |

By looking at the two sets of outputs directly, it is very obvious that VARMA model does a better forecasting than the ARMA model in the last three quarters in 2015. However, although the ARMA model has a better prediction in the first quarter than the VARMA model, the predictions are not accurate in both cases. It is interesting that the VARMA model has better results for the second and third quarters than the first. For ARMA models, the results get worse when the dates get further in the future. In order to have a better comparison, another four predictions are made from both models.

**Table 9 Quarterly U.S. GDP Growth Forecast in 2016 by ARMA Model**

| Date | Predicted Value | Standard Error | Actual Value |
|---|---|---|---|
| 1st Quarter 2016 | 3.124212 | 3.387456 | 0.8 |
| 2nd Quarter 2016 | 3.126519 | 3.389854 | 1.4 |
| 3rd Quarter 2016 | 3.129877 | 3.390705 | 3.5 |
| 4th Quarter 2016 | 3.131191 | 3.390978 | 1.9 |

**Table 10 Quarterly U.S. GDP Growth Forecast in 2016 by VARMA Model**

| Date | Predicted Value | Standard Error | Actual Value |
|---|---|---|---|
| 1st Quarter 2016 | 1.2887 | 3.693 | 0.8 |
| 2nd Quarter 2016 | 1.1356 | 3.718 | 1.4 |
| 3rd Quarter 2016 | 1.0318 | 3.736 | 3.5 |
| 4th Quarter 2016 | 0.9607 | 3.751 | 1.9 |

When doing prediction in a longer future time span, the results are less meaningful and less accurate. We can see from Table 9 that ARMA model fails to provide any meaningful predictions. All four predications are around 3.12 and far from the actual value. The VARMA model provides a better result than ARMA model, but it also becomes less accurate than the predictions in 2015.

# Chapter 7
# Conclusion

As the famous British statistician George E.P. Box wrote in his book, “Essentially, all models are wrong, but some are useful.” There are no perfect models, and we should not trust our model too much.

In this paper, two methods were applied to predict the U.S. quarterly GDP growth. For the ARMA approach, a AR (2) model is the best fit based on the data set. For the VARMA approach, a VARMA (1,2) is the optimal model based on AIC. It is constructed with four variables including U.S. GDP Growth, Federal Funds Rate, U.S. Unemployment Rate and Canada GDP Growth.

By comparing the performance of the two models, the VARMA model seems to be much better than the ARMA model, which is consistent with the assumption that multivariate time series is better at forecasting by utilizing additional information from other variables. The ARMA model gets worse as the period of prediction gets longer. However, for the VARMA model, the predictions of the second and third quarters were actually better than the first and fourth quarters in 2015. In the last three quarters in 2015, there is a decreasing trend for U.S. GDP growth. The VARMA model captured the trend, but the ARMA model fails.

Lastly, the overall performances of the models in this paper meet the expectation. The main strength of multivariate time series models is addressed and the forecasting is reasonable. However, there are still many adjustments and calibrations that can be done to improve the model, which is beyond the scope of this paper. There maybe other variables more appropriate

which are not included due to availability of the data. For example, Germany and Europe GDP growth should be included in the model. There maybe political factors that shock the GDP growth, which are not in the model. All of these remain to be further explored.

# BIBLIOGRAPHY

Board of Governors of the Federal Reserve System (US), Effective Federal Funds Rate [FEDFUNDS], retrieved from FRED, Federal Reserve Bank of St. Louis; https://fred.stlouisfed.org/series/FEDFUNDS, January 10, 2017.

Box, George E. P, and Jenkins, Gwilym M. 2008. *Time Series Analysis: Forecasting and Control.* Wiley.

Diebold, Francis X. 2001. *Elements of Forecasting.* South-Western

Directorate, OECD Glossary of Statistical Terms - Gross domestic product (GDP) Definition. Retrieved September 30, 2016, from http://stats.oecd.org/glossary/detail.asp?ID=1163

Dmyto, Holod. 2000. The relationship between price level, money supply and exchange rate in Ukraine. A thesis submitted to the National University of keiv-mohyla Academic.

Hamilton, James D. 1994. *Time Series Analysis*. Princeton University Press

Onwukwe, C. E., & Nwafor, G. O. (2014). A Multivariate Time Series Modeling of Major Economic Indicators in Nigeria. *American Journal of Applied Mathematics and Statistics 2.6 (2014): 376-385.*

Organization for Economic Co-operation and Development, Gross Domestic Product by Expenditure in Constant Prices: Total Gross Domestic Product for Canada© [NAEXKP01CAQ657S], retrieved from FRED, Federal Reserve Bank of St. Louis; https://fred.stlouisfed.org/series/NAEXKP01CAQ657S, January 11, 2017.

Organization for Economic Co-operation and Development, Gross Domestic Product by Expenditure in Constant Prices: Exports of Goods and Services for Australia© [NAEXKP06AUQ657S], retrieved from FRED, Federal Reserve Bank of St. Louis; https://fred.stlouisfed.org/series/NAEXKP06AUQ657S, January 10, 2017.

Ruey S. Tsay (2015). MTS: All-Purpose Toolkit for Analyzing Multivariate Time Series (MTS) and Estimating Multivariate Volatility Models. R package version 0.33. https://CRAN.R-project.org/package=MTS

U.S. Bureau of Economic Analysis, Real Gross Domestic Product [A191RL1Q225SBEA], retrieved from FRED, Federal Reserve Bank of St. Louis; https://fred.stlouisfed.org/series/A191RL1Q225SBEA, January 10, 2017.

U.S. Bureau of Labor Statistics, Unemployment Rate: 20 years and over [LNS14000024], retrieved from FRED, Federal Reserve Bank of St. Louis; https://fred.stlouisfed.org/series/LNS14000024, January 10, 2017.

# ACADEMIC VITA

**Academic Vita of Xuanhao Zhang**
Email xwz5161@psu.edu

Education: Pennsylvania State University, University Park

Major(s) and Minor(s): Mathematics, Economics and Statistics
Honors: Economics

Thesis Title: Modelling Major Economic Indicators via Multivariate Time Series

Thesis Supervisor: Patrik Guggenberger

Work Experience

Date: Jan. 2016 – May. 2016

Title: Research Assistant

Description: Selected to participate in Penn State's Research Experience for Undergraduates program funded by Bates White committing an average of 9 hours per week to supervised research with a faculty member

Institution/Company: Penn State Economic Department

Supervisor's Name: David Shapiro

Language Proficiency: Mandarin, English