Student: Arjun Panikar

Mentor: Neal Fultz

401K Fund Forecasting

springboard foundations of data science - Capstone Project Report

Contents

[Problem Definition 2](#_Toc480730586)

[Fund Information 3](#_Toc480730587)

[Data Set 10](#_Toc480730591)

[Analysis 11](#_Toc480730592)

[Linear Regression Analysis 13](#_Toc480730593)

[Time Series Modelling 15](#_Toc480730594)

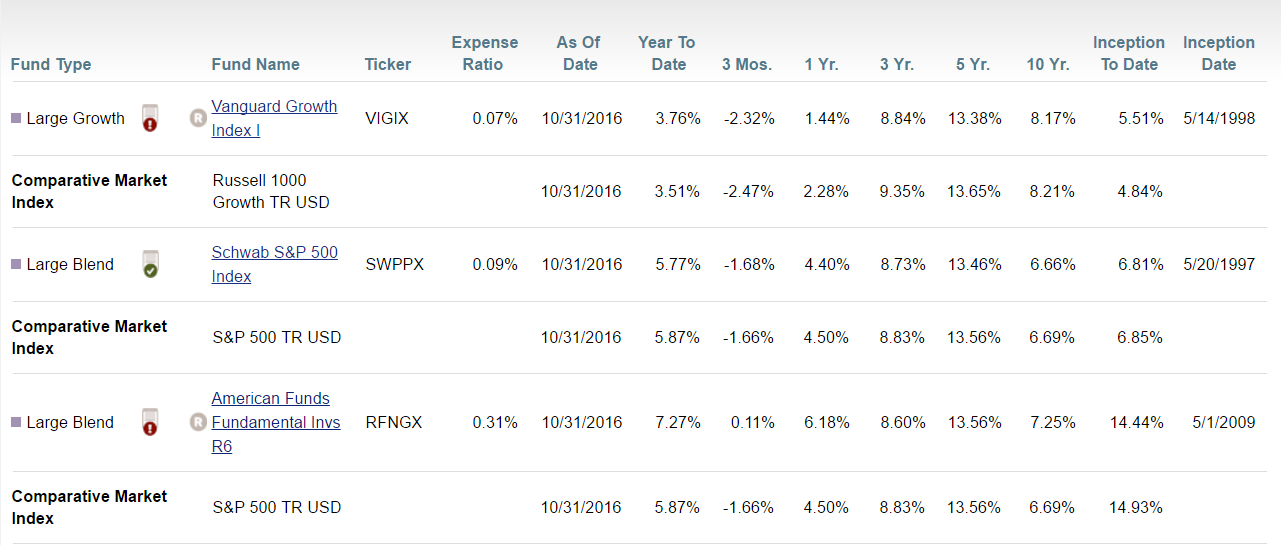
[Forecasting & Conclusion 15](#_Toc480730595)

[Appendix 16](#_Toc480730596)

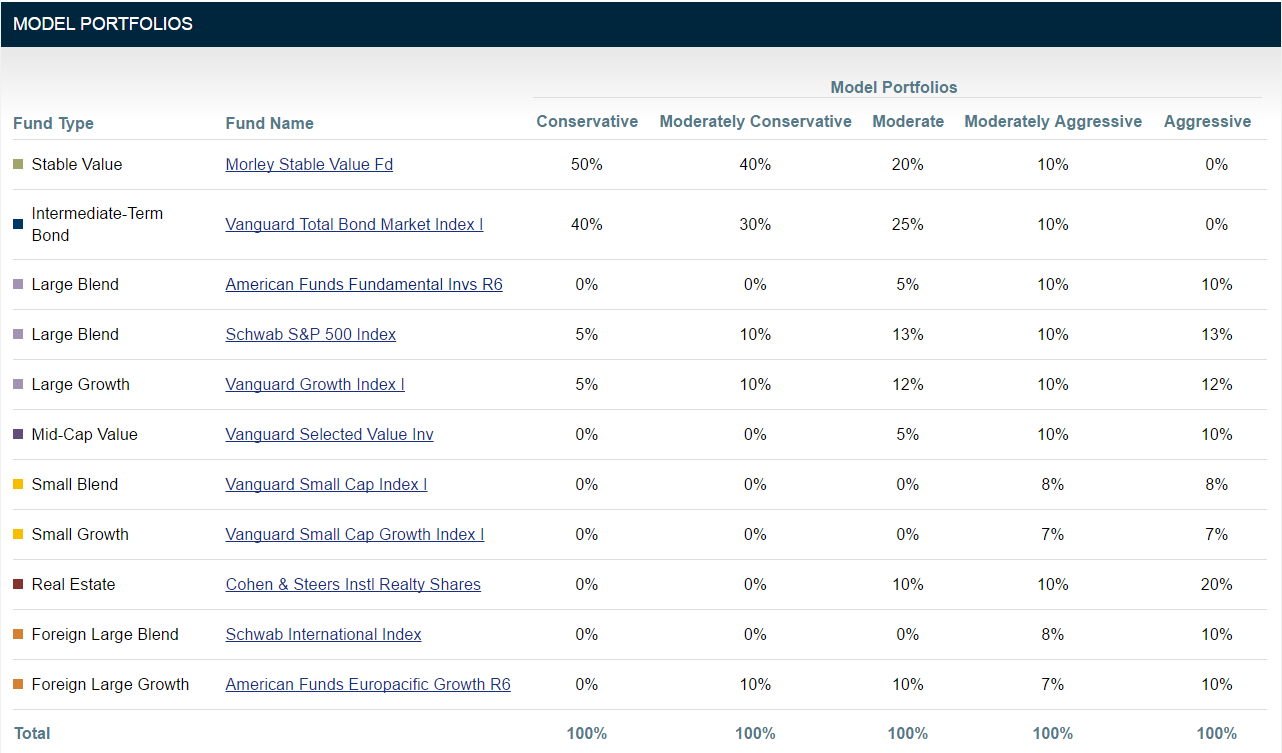
## Problem Definition

The 401K portfolio is an investment that any investor needs to be continuously monitoring to ensure that his/her hard earned money is being put to good use. Unfortunately, in many cases a lack of understanding on how to analyze fund data along with schedule constraints makes the management of this task cumbersome which results in missed opportunities and loss of money. In addition, the 401K management interface, does not have a robust suggestion feature that users can avail to make investment decisions. Investors are presented with one of two options:

* Make their own investment decisions using the fund information feature



* Use the model portfolio feature that looks like below to allocate funds by default



In both cases above, the crucial element missing is factoring in the actual performance of funds in the economy which is not ideal. The problem that this project is attempting to solve is to provide this information to an investor in a timely fashion, so that he/she can continuously visit their portfolio and make informed investment decisions.

## Fund Information

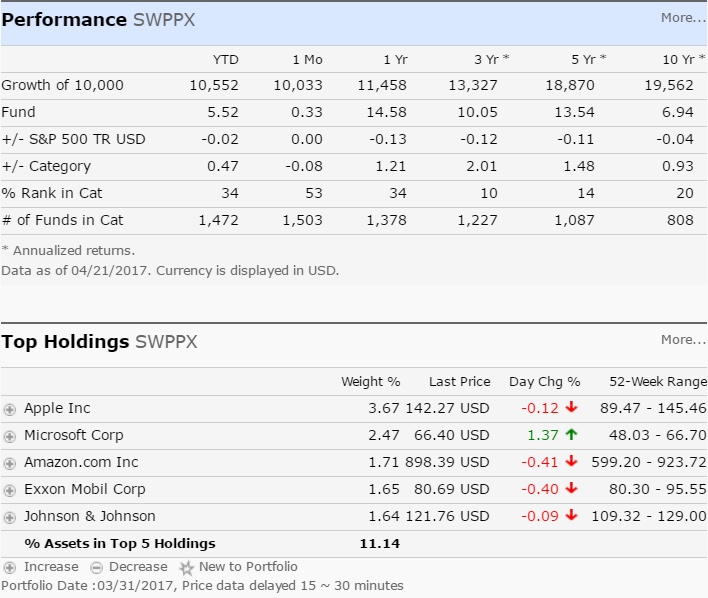
**Vanguard Growth Index Institutional (VIGIX)**

This fund invests in stocks of large U.S. companies in market sectors that tend to grow more quickly than the broad market. This low-cost index fund follows a buy-and-hold approach and invests in substantially all of the stocks represented in its broad benchmark. The fund’s primary risk, apart from general stock market volatility, comes from the fact that its focus on large-capitalization growth stocks may, at times, underperform the broader stock market.

# 

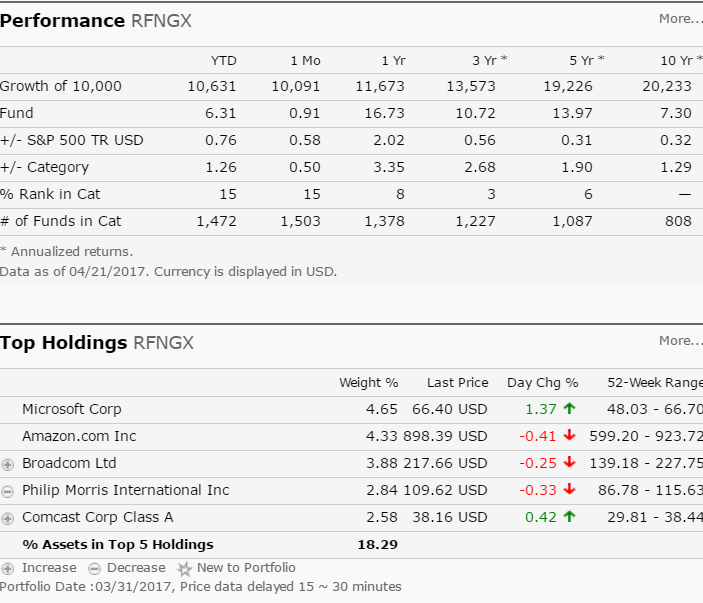
**Schwab S&P 500 Index (SWPPX)**

The Schwab S&P 500 Index tracks the Standard & Poor's 500 index, one of the most widely watched benchmarks for U.S. stocks. The index covers about 80 percent of the investable market capitalization of the U.S. equity market. The fund tracks the index very closely and generally. The only difference in the return of the fund and the index is the fund’s expense ratio.



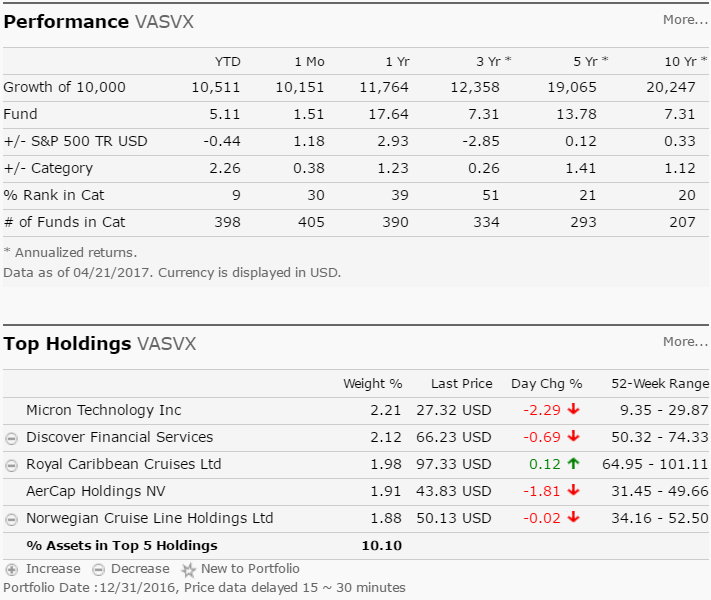
**American Funds Fundamental Investors (RFNGX)**

Of its large-blend funds, this is American Funds' most aggressive large-cap U.S. offering. With a diversified portfolio and low fees, the fund remains a solid option for investors seeking growth and diversification, according to Morningstar. As of April 07, 2017, the fund has assets totaling almost $84.95 billion invested in 211.00 different holdings. Its portfolio consists primarily of common stocks or securities convertible into common stocks, bonds, U.S. government securities and cash. Microsoft, Amazon, Comcast, Boeing and Apple make up some of the top assets in the fund.



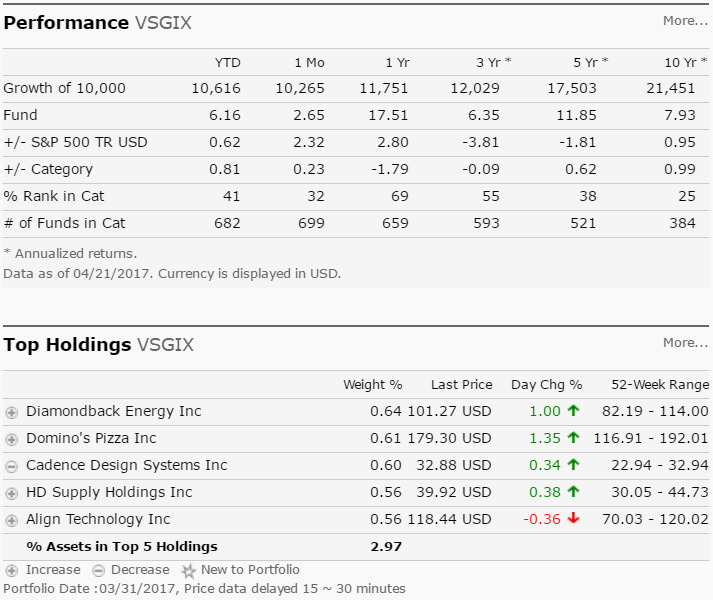
**Vanguard Selected Value Fund Investor Shares (VASVX)**

The Vanguard Selected Value Fund is designed to capture long-term gains by purchasing stock of currently undervalued companies. As of April 07, 2017, the fund has assets totaling almost $10.14 billion invested in 127.00 different holdings. Its portfolio consists primarily of mid-cap U.S. stocks trading at a bargain relative to future potential. The fund also holds a small number of non-U.S. stocks. It may invest up to 25 percent of assets outside the U.S. However, that non-U.S. stake currently represents less than 10 percent of total assets.



**Vanguard Small-Cap Growth Index Fund Institutional Shares (VSGIX)**

The Vanguard Small Cap Growth Index Fund has a very simple methodology: It tracks performance of the Center for Research in Security Prices' U.S. Small Cap Growth index. It may be an appropriate choice for investors who want to diversify from large-cap domestic stocks. As of April 07, 2017, the fund has assets totaling almost $18.25 billion invested in 683.00 different holdings. Its portfolio consists primarily of mid-cap and small-cap U.S. stocks. The median market capitalization of all holdings is $3.6 billion. That’s well below the $79.3 billion median market capitalization of the Vanguard 500 Index Fund. Top sectors represented in the fund are financials, industrials and technology.



**Vanguard Small-Cap Index Fund Institutional Shares (VSCIX)**

With very low expenses and a focus on smaller companies, the Vanguard Small Cap Index fund is a great long-term building block for investor's portfolios. As of April 07, 2017, the Vanguard Small Cap Index fund has assets totaling almost $74.74 billion invested in 1,439.00 different holdings. Its portfolio consists primarily of small-sized companies and tracks the CRSP US Small Cap index, an index designed to include firms with market capitalizations in the bottom 2-15 percent of the overall market. Some of its largest holdings are in the financial, consumer goods and industrial sector. The fund’s top holdings include Salix Pharmaceuticals, Foot Locker, Harman International Industries, Cooper Cos. and Gannett Co.

# 

# 

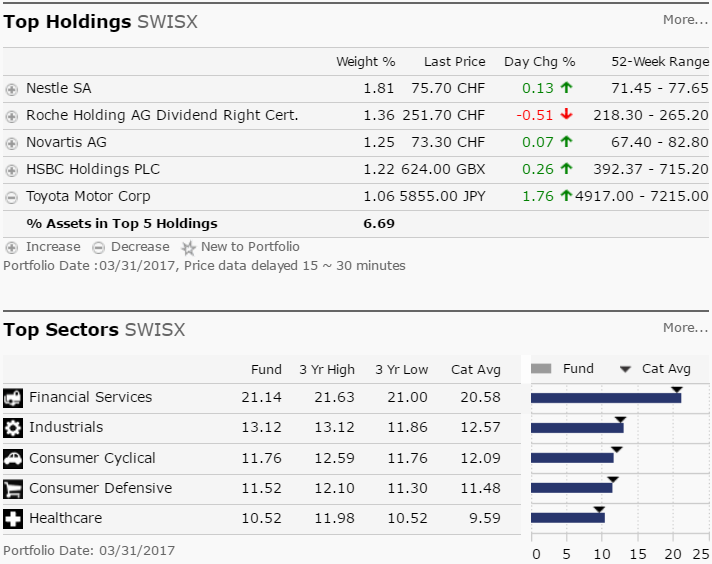
**Cohen & Steers Institutional Realty Shares (CSRIX)**

The investment seeks total return through investment in real estate securities. The fund invests at least 80%, and normally substantially all, of its total assets in common stocks and other equity securities issued by real estate companies. It may invest up to 20% of its total assets in securities of foreign issuers (including emerging market issuers) which meet the same criteria for investment as domestic companies, including investments in such companies in the form of American Depositary Receipts ("ADRs"), Global Depositary Receipts ("GDRs") and European Depositary Receipts ("EDRs"). The fund is non-diversified.

# 

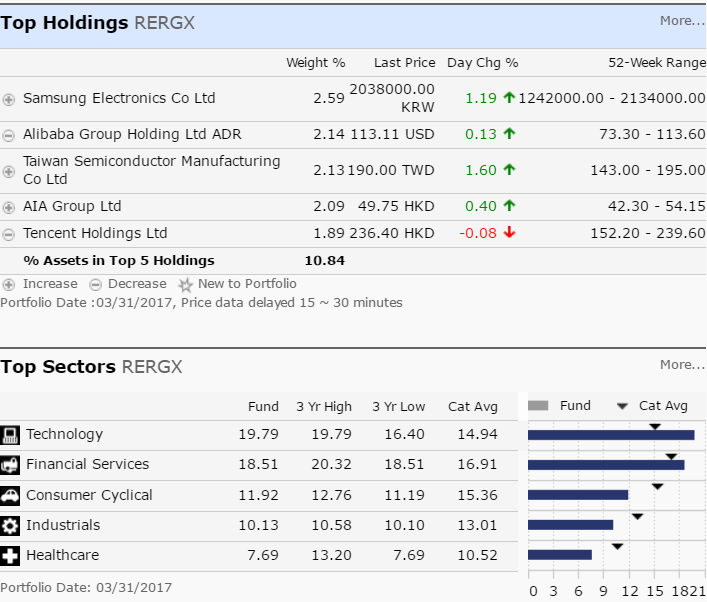
**Schwab International Index Fund (SWISX)**

The investment seeks to track the performance of a benchmark index that measures the total return of large, publicly traded non-U.S. companies from countries with developed equity markets outside of the United States. The fund generally invests at least 80% of its net assets in stocks that are included in the MSCI EAFE Index. It may invest in derivatives, principally futures contracts, and lend its securities to minimize the gap in performance that naturally exists between any index fund and its corresponding index. The fund may concentrate its investments in an industry or group of industries to the extent that its comparative index is also so concentrated.



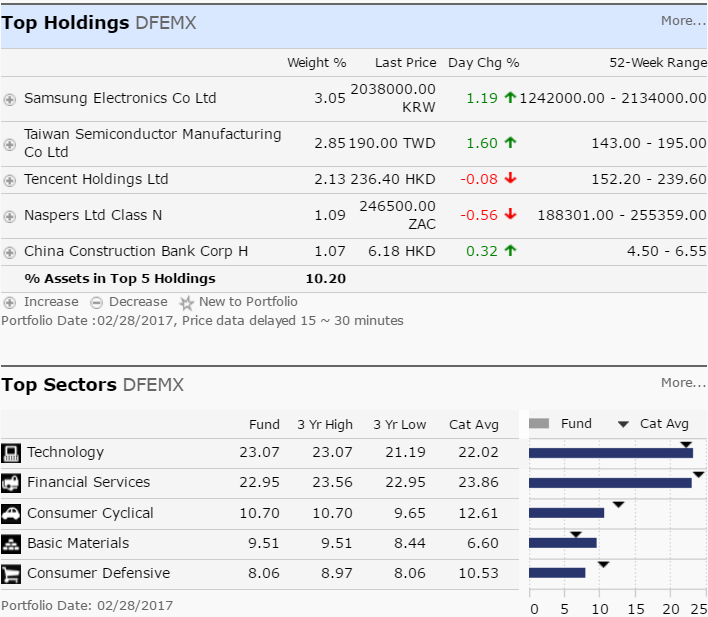
**American Funds EuroPacific Growth Fund Class R-6 (RERGX)**

The American Funds EuroPacific Growth Fund, the largest international fund, has a long history of investing in a diverse set of markets. As of April 07, 2017, the fund has assets totaling $132.31 billion. Its portfolio consists mostly of a blend of large-company growth and value stocks in Europe and the Pacific Basin. The fund had a banner year in 2009 but had a slow start in 2010. Emerging Markets make up about 20 percent of the portfolio as of the end of the first quarter. As the name suggests, the fund is well-diversified throughout Europe and Asia among developed and developing countries with more than 300 holdings. The fund has returned 13.10 percent over the past year and 2.17 percent over the past three years. Historically, the fund has a record of steady gains. Since its 1984 inception, the fund has weathered several overseas financial market storms. The fund has returned 6.64 percent over the past five years and 3.51 percent over the past decade.



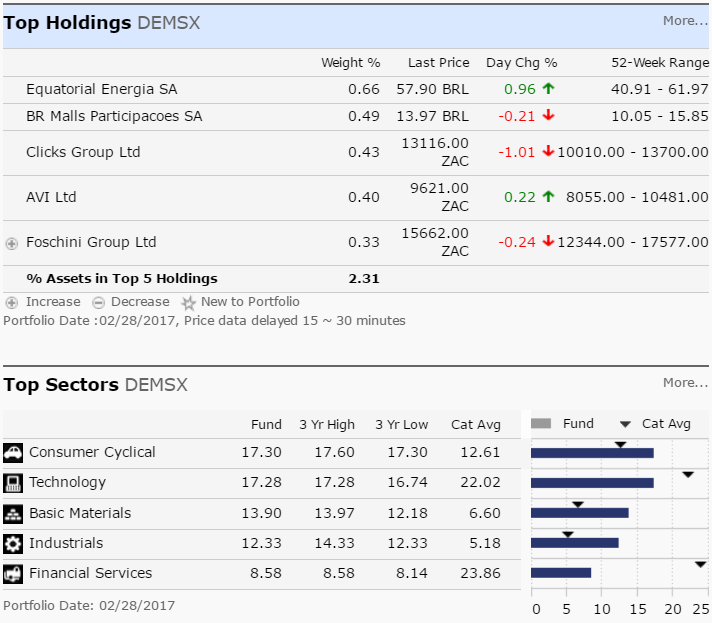
**DFA Emerging Markets Portfolio Institutional Class (DFEMX)**

DFA Emerging Markets I fund isn’t trying to beat the market. Like other funds offered by Dimensional Fund Advisors, this one seeks to instead provide investors with concentrated exposure to a particular area of the market—in this case large-cap emerging market stocks. As of April 07, 2017, the fund has assets totaling almost $5.58 billion invested in 1,099.00 different holdings. Its portfolio consists primarily of the largest companies in the emerging markets and is closely aligned with the MSCI Emerging Markets Index.



**DFA Emerging Markets Small Cap Portfolio Institutional Class (DEMSX)**

The investment seeks long-term capital appreciation. The Portfolio is a Feeder Portfolio and pursues its objective by investing substantially all of its assets in its corresponding master fund, the Emerging Markets Small Cap Series (the "Emerging Markets Small Cap Series") of the DFA Investment Trust Company (the "Trust"), which has the same investment objective and policies as the Portfolio. As a non-fundamental policy, under normal circumstances, the Emerging Markets Small Cap Series will invest at least 80% of its net assets in emerging market investments that are designated in the Prospectus as Approved Market securities of small companies.



**Loomis Sayles Bond Fund Institutional Class (LSBDX)**

The Loomis Sayles Bond fund is heavy on the corporate bond side but offers some flexibility within lower grade names and even equities. As of April 07, 2017, the fund has assets totaling almost $13.82 billion invested in 458.00 different holdings. Its portfolio consists primarily of fixed income securities but also below-investment-grade bonds and stocks. Nearly 70 percent of the fund is invested in the U.S. and nearly 12 percent invested in Canada.



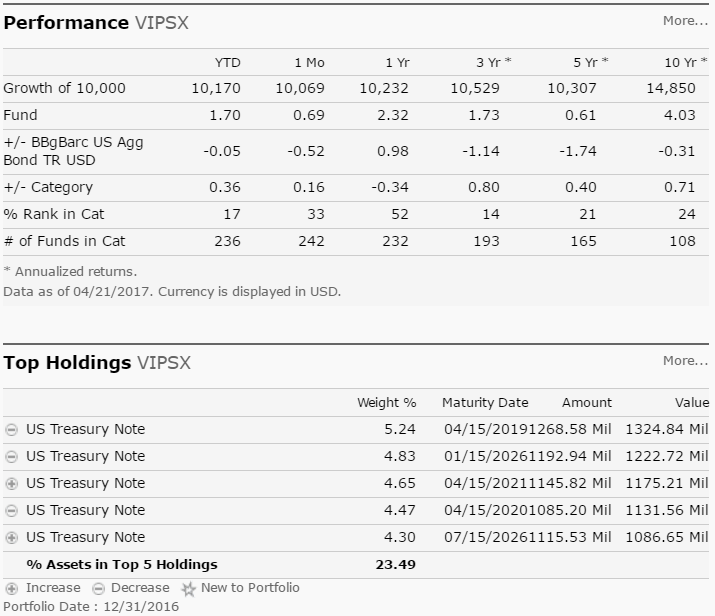
**Vanguard Total Bond Market Index Fund Institutional Shares (VBTIX)**

The Vanguard Total Bond Market Index fund holds higher quality investment-grade bonds, both corporate and government, with U.S. government debt making up themajority of assets in the portfolio. It's a broad-market bond fund and is used as a core fixed-income holding for a wide range of investors. Because the fund invests in all segments and maturities of the fixed income market, this is considered a core bond holding for investors.



**Vanguard Inflation-Protected Securities Fund Investor Shares (VIPSX)**

A strict approach to inflation protection and low fees is a strategy that keeps this fund working well for investors, though interest rate changes remain this fund’s foil.  
As of April 07, 2017, the fund has assets totaling almost $26.83 billion invested in 37.00 different holdings. The fund exclusively invests in TIPS (Treasury-inflation protected securities), and doesn't veer into more credit-sensitive areas such as mortgage debt or emerging-market debt like some of its peers. This, of course can be beneficial when the markets don't favor those riskier sectors, but the fund can underperform its peers when assets outside its narrow focus perform well. The fund has returned 1.32 percent over the past year and 1.94 percent over the past three years.



## Data Set

The day end prices for all funds can be obtained from Quandl. The data elements available through the API were recorded into data frames with following information.

* Date – Date of record
* Adjusted Close – Day end price of the 401K fund
* Ticker – Ticker symbol

Additionally, a few derived variables were added to add more meaning to the dataset. The approach was to create subsets for each ticker.

* Index – Increase/decrease from previous day
* Indicator – 1 if increase; 0 if decrease

Once above subsets were created, they were summarized/aggregated together into a single data frame for data analysis. The fields in the final frame were:

* Ticker
* Year
* Quarter
* Month
* No\_Growth\_Days – Count of days with positive growth
* No\_No\_Growth\_Days – Count of days with negative growth
* No\_Positive\_Runs – Highest continuous streak of positive growth
* No\_Negative\_Runs – Highest continuous streak of negative growth
* Month\_Begin\_Price – Ticker price at the beginning of the month
* Month\_End\_Price – Ticker price at the end of the month
* Range – Month end price – Month begin price
* Growth\_Ratio – Percentage of growth that was continuous
* No\_Growth\_Ratio – Percentage of negative growth that was continuous

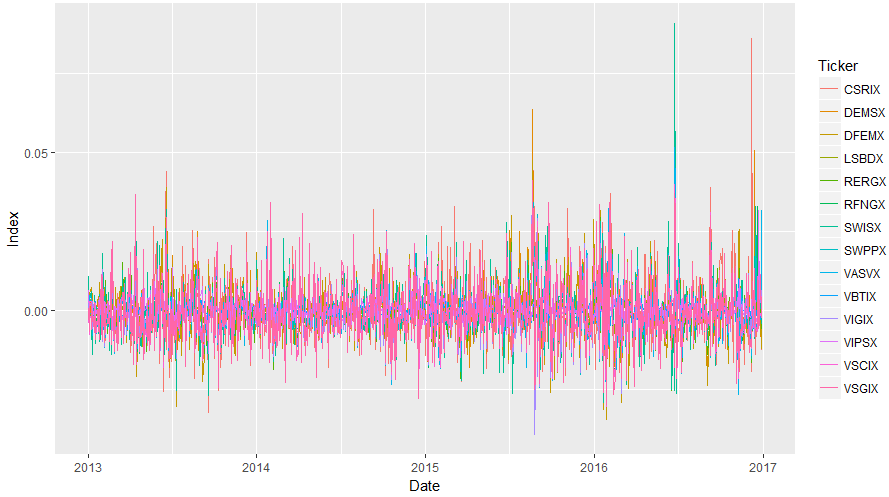
# Analysis

A summary table with the findings are as below with information on the recommended portfolio strategies and percentage allocations. Data range from 2013 – 2016. It is reliable to assume after the analysis that the suggested portfolios do result in good returns if invested over a longer duration. But the analysis was required to validate the data.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | % Months Positive Growth | Ticker Price Improvement | Conservative | Moderately Conservative | Moderate | Moderately Aggressive | Aggressive |
| RFNGX | 68.75 | 19.35 |  |  | 5% | 10% | 10% |
| VSCIX | 64.58 | 25.02 |  |  |  | 8% | 8% |
| VSGIX | 64.58 | 13.00 |  |  |  | 7% | 7% |
| SWPPX | 62.50 | 12.44 | 5% | 10% | 13% | 10% | 13% |
| VASVX | 60.42 | 10.23 |  |  | 5% | 10% | 10% |
| VIGIX | 58.33 | 19.38 | 5% | 10% | 12% | 10% | 12% |
| CSRIX | 58.33 | 7.50 |  |  | 10% | 10% | 20% |
| VBTIX | 56.25 | 0.66 | 40% | 30% | 25% | 10% |  |
| SWISX | 54.17 | 1.90 |  |  |  | 8% | 10% |
| LSBDX | 54.17 | 1.32 |  |  |  |  |  |
| RERGX | 50.00 | 6.00 |  | 10% | 10% | 7% | 10% |
| VIPSX | 50.00 | -0.52 |  |  |  |  |  |
| DFEMX | 45.83 | -2.31 |  |  |  |  |  |
| DEMSX | 41.67 | -1.02 |  |  |  |  |  |
| Morley\_Stable |  |  | 50% | 40% | 20% | 10% |  |

The generic trend on most funds has been an upward movement in the ticker price as shown below while the index change for the most part has a constant variance around zero mean.





Special notes from the analysis:

Tickers VASVX, VSCIX and VSGIX showed positive results in the month of February for all years

Tickers DEMSX, LSBDX, SWISX showed positive results in the month of February for all years

Tickers RFNGX, SWPPX, VIGIX, VSCIX, VSGIX showed positive results in the month of May for all years

Tickers VIPSX showed positive results in the month of July for all years

Tickers VSCIX, VSGIX showed positive results in the month of November for all years

# Linear Regression Analysis

The approach to better understand data was to conduct a linear regression and the results were as follows:

Call:

lm(formula = Month\_End\_Price ~ No\_Growth\_Days + No\_No\_Growth\_Days +

No\_Positive\_Runs + No\_Negative\_Runs + Ticker + as.factor(Month),

data = summary\_funds\_df, subset = Year == 2013 | Year ==

2014 | Year == 2015 | Year == 2016)

Residuals:

Min 1Q Median 3Q Max

-12.179 -1.153 0.181 1.802 9.890

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 40.514731 3.919584 10.336 < 2e-16 \*\*\*

No\_Growth\_Days 0.008633 0.199753 0.043 0.965541

No\_No\_Growth\_Days -0.095325 0.194167 -0.491 0.623635

No\_Positive\_Runs -0.081839 0.119711 -0.684 0.494448

No\_Negative\_Runs -0.034348 0.126505 -0.272 0.786078

TickerDEMSX -21.816533 0.716826 -30.435 < 2e-16 \*\*\*

TickerDFEMX -17.004421 0.711973 -23.884 < 2e-16 \*\*\*

TickerLSBDX -27.668558 0.718996 -38.482 < 2e-16 \*\*\*

TickerRERGX 4.305813 0.707481 6.086 1.99e-09 \*\*\*

TickerRFNGX 5.534063 0.706293 7.835 1.94e-14 \*\*\*

TickerSWISX -23.238023 0.708030 -32.821 < 2e-16 \*\*\*

TickerSWPPX -11.575618 0.707363 -16.364 < 2e-16 \*\*\*

TickerVASVX -15.571401 0.707061 -22.023 < 2e-16 \*\*\*

TickerVBTIX -30.687745 0.715969 -42.862 < 2e-16 \*\*\*

TickerVIGIX 8.984572 0.706822 12.711 < 2e-16 \*\*\*

TickerVIPSX -27.846904 0.714582 -38.969 < 2e-16 \*\*\*

TickerVSCIX 11.072119 0.707595 15.648 < 2e-16 \*\*\*

TickerVSGIX -7.411023 0.707886 -10.469 < 2e-16 \*\*\*

as.factor(Month)2 0.720255 0.685955 1.050 0.294111

as.factor(Month)3 1.483569 0.685866 2.163 0.030904 \*

as.factor(Month)4 1.751903 0.687796 2.547 0.011093 \*

as.factor(Month)5 1.904342 0.669730 2.843 0.004605 \*\*

as.factor(Month)6 1.929973 0.685350 2.816 0.005011 \*\*

as.factor(Month)7 2.427354 0.701631 3.460 0.000577 \*\*\*

as.factor(Month)8 1.954140 0.716502 2.727 0.006559 \*\*

as.factor(Month)9 1.833611 0.664794 2.758 0.005978 \*\*

as.factor(Month)10 2.562745 0.756838 3.386 0.000752 \*\*\*

as.factor(Month)11 2.786380 0.658866 4.229 2.69e-05 \*\*\*

as.factor(Month)12 2.673894 0.699507 3.823 0.000145 \*\*\*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 3.46 on 643 degrees of freedom

Multiple R-squared: 0.9458, Adjusted R-squared: 0.9434

F-statistic: 400.7 on 28 and 643 DF, p-value: < 2.2e-16

While testing for assumptions of linear regression, it was found that while the data fit well into the model, the premise for using a multiple linear regression would not be accurate as explained by diagnostics(Appendix section 1) owing to the non-linear character of the data set.

# Time Series Modelling

Since the data is a time series, the next step was to perform a version of time series modelling that can provide a reliable forecast for the tickers. To check to see if an ARIMA model or one of its variants can be applied, we test for autocorrelation and partial autocorrelation on the ticker price. From the diagnostics (Appendix section 2), it was found that the ACF is exponentially decaying and there is a significant spike at lag p in PACF, but none beyond that in most of the models. Therefore, the data may follow an ARIMA(p, d, 0) model.

In using the Auto.Arima function to automatically detect the appropriate model, the following results were obtained.

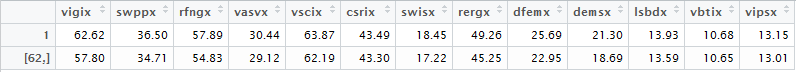
|  |  |  |
| --- | --- | --- |
| Ticker | (p, d, q) Model Selected | Model Formula |
| VIGIX | (0, 1, 0) | auto.arima(vigix\_df[,2],seasonal=FALSE) |
| SWPPX | (1, 1, 1) | auto.arima(swppx\_df[,2],seasonal=FALSE) |
| RFNGX | (0, 1, 0) | auto.arima(rfngx\_df[,2],seasonal=FALSE) |
| VASVX | (1, 1, 0) | auto.arima(vasvx\_df[,2],seasonal=FALSE) |
| VSGIX | (0, 1, 0) | auto.arima(vsgix\_df[,2],seasonal=FALSE) |
| VSCIX | (0, 1, 0) | auto.arima(vscix\_df[,2],seasonal=FALSE) |
| CSRIX | (0, 1, 0) | auto.arima(csrix\_df[,2],seasonal=FALSE) |
| SWISX | (0, 1, 0) | auto.arima(swisx\_df[,2],seasonal=FALSE) |
| RERGX | (0, 1, 1) | auto.arima(rergx\_df[,2],seasonal=FALSE) |
| DFEMX | (0, 1, 1) | auto.arima(dfemx\_df[,2],seasonal=FALSE) |
| DEMSX | (0, 1, 1) | auto.arima(demsx\_df[,2],seasonal=FALSE) |
| LSBDX | (2, 1, 0) | auto.arima(lsbdx\_df[,2],seasonal=FALSE) |
| VBTIX | (0, 1, 0) | auto.arima(vbtix\_df[,2],seasonal=FALSE) |
| VIPSX | (2, 1, 2) | auto.arima(vipsx\_df[,2],seasonal=FALSE) |

# Forecasting & Conclusion

In analyzing residual diagnostics (Appendix section 3), the residuals for the most part seem to be uncorrelated to each other and seem to have a constant variance and a mean close to zero. With the assumption that the models are accurate, a separate data frame was created with a forecast of 90 data points to account for the testing data set from Jan 2017 – Mar 2017. The data frame was summarized to take into factor the first and last data points and a range variable was defined that describes the positive or negative movement of the ticker within the forecasted data frame for each ticker within the 90 day period. The result looked as below for each ticker.



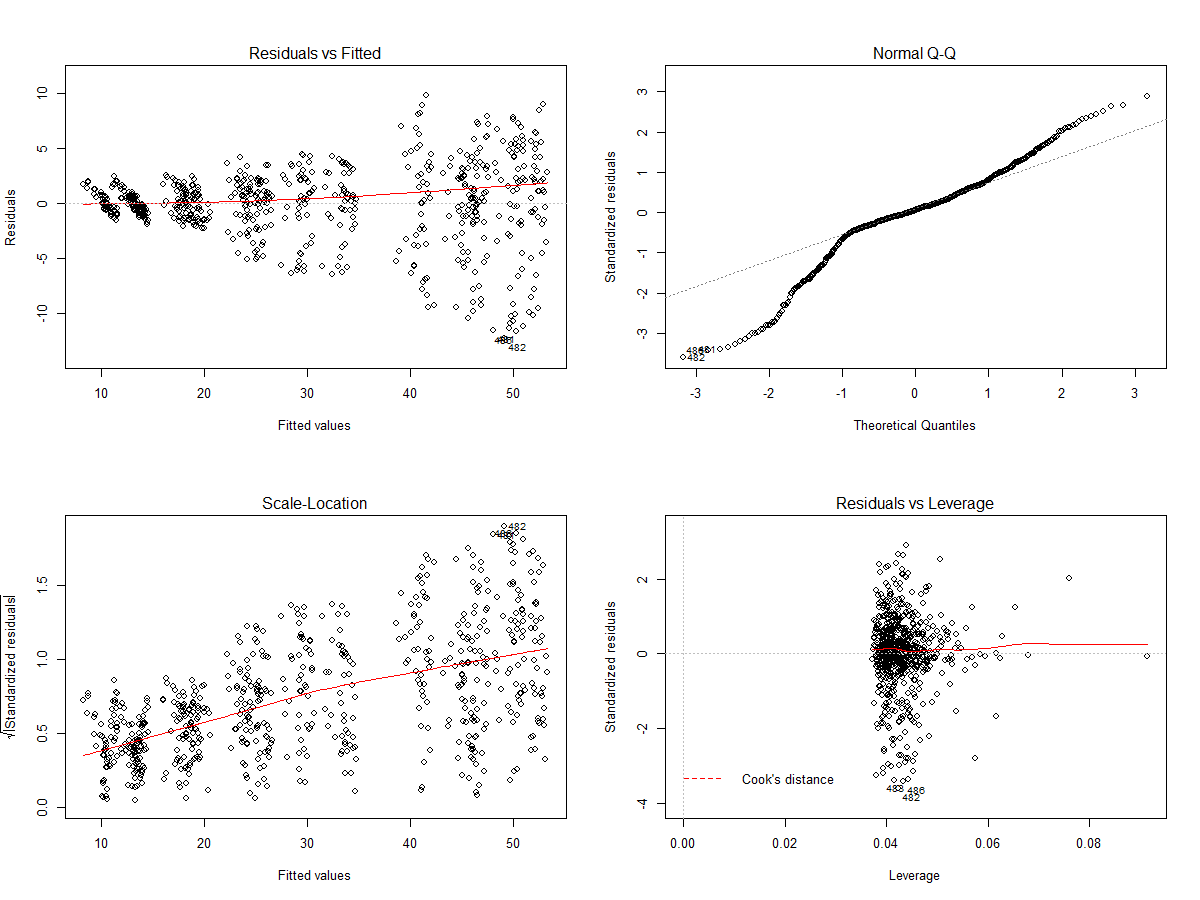
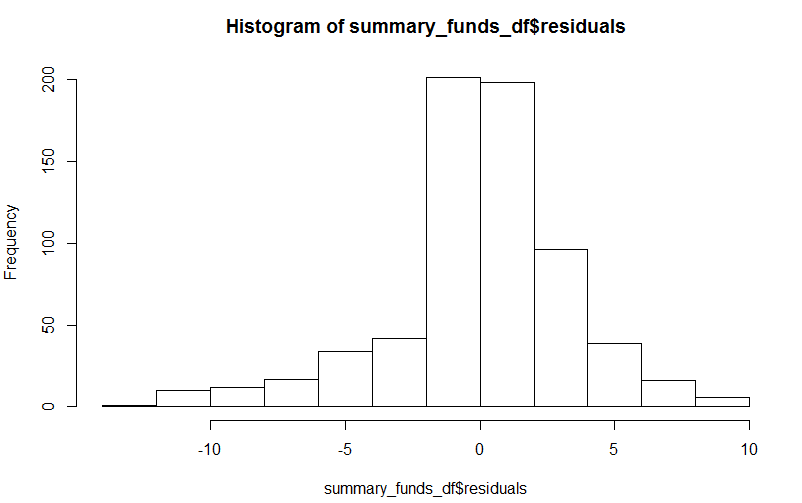
The forecasted range for ticker values portrayed a negative growth outlook for the 90 day period for all tickers. In comparing this observation with the actual data between Jan 2017 – March 2017, it was validated that the negative growth outlook forecasted was in sync with the actual performance of the tickers in the sense that the value of all ticker prices dropped as of March 31 2017, when compared to Jan 1 2017.



In comparing the forecasted vs actual range for all tickers it was found that the forecasted ranges for funds VSCIX, VASVX, LSBDX and VIPSX resembled closely to the actual and therefore were best fitted by the ARIMA models.

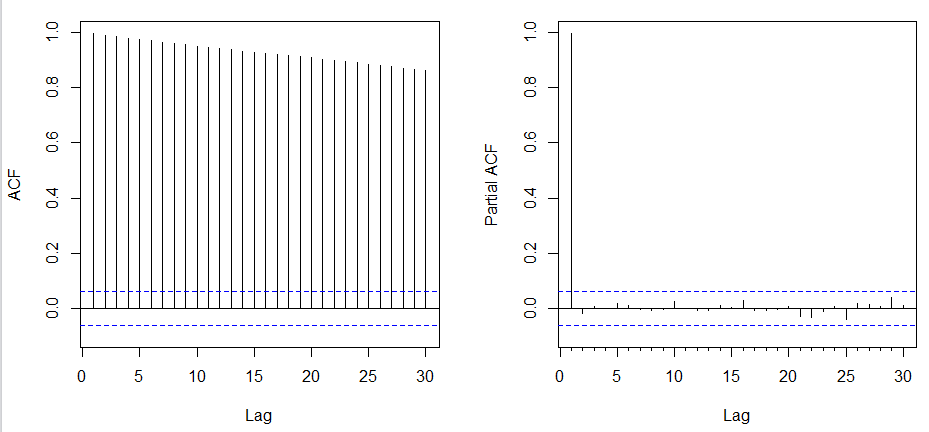
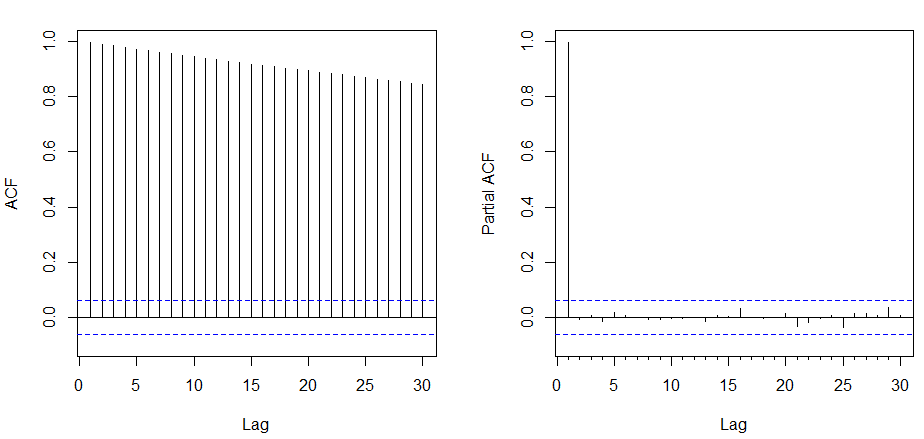
## Appendix

1. Linear Regression Diagnostics

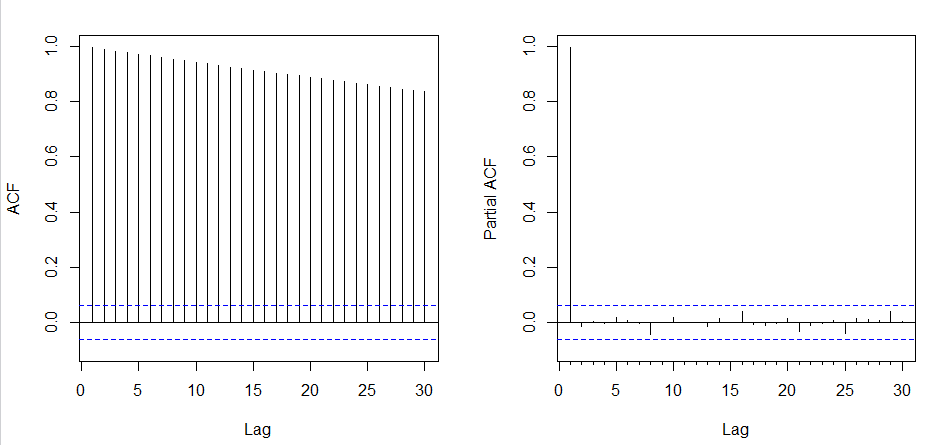
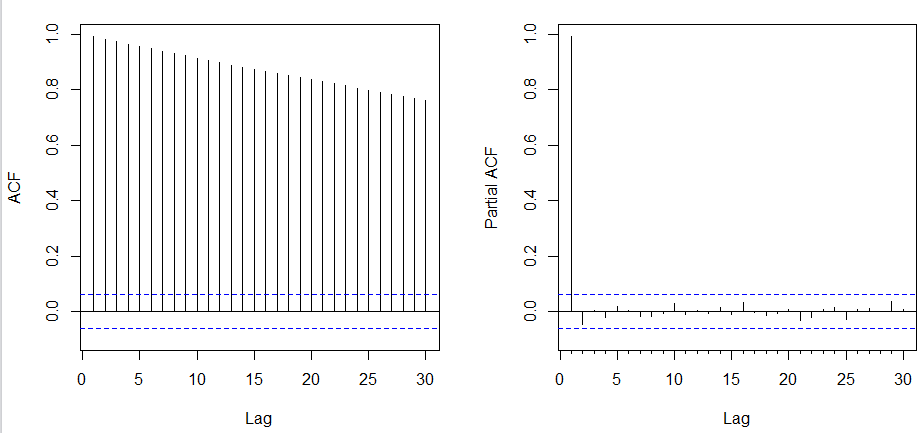


1. ACF and PACF Plots of all Ticker Symbols

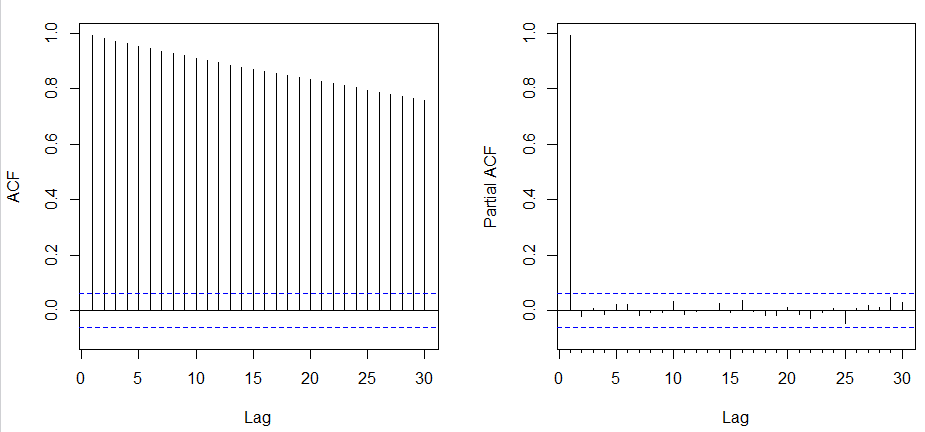
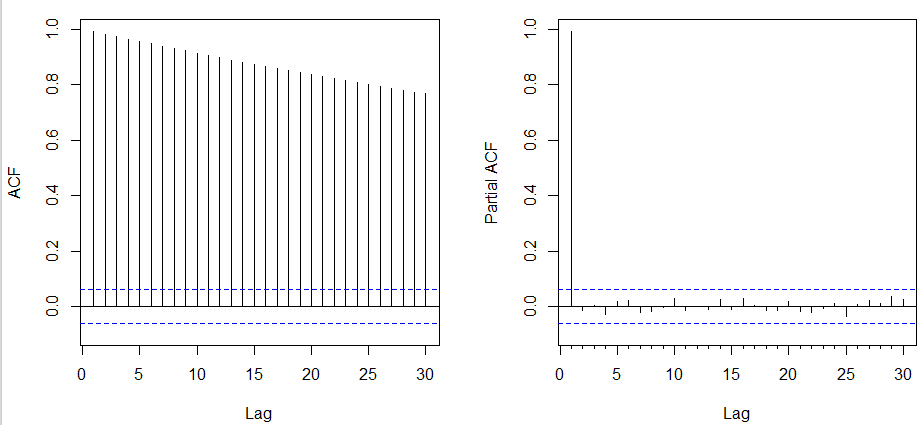
VIGIX SWPPX

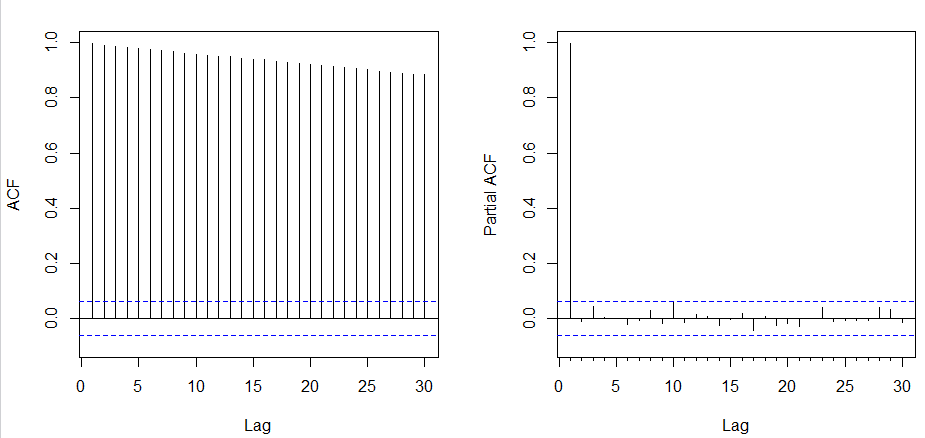
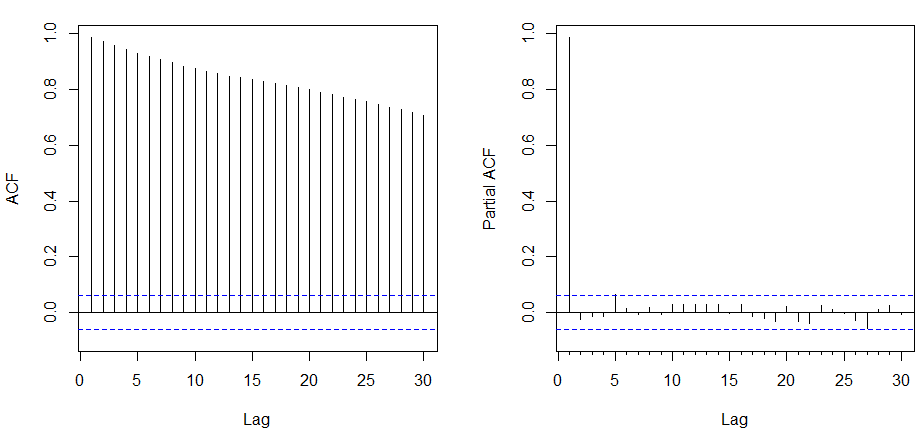
RFNGX VASVX

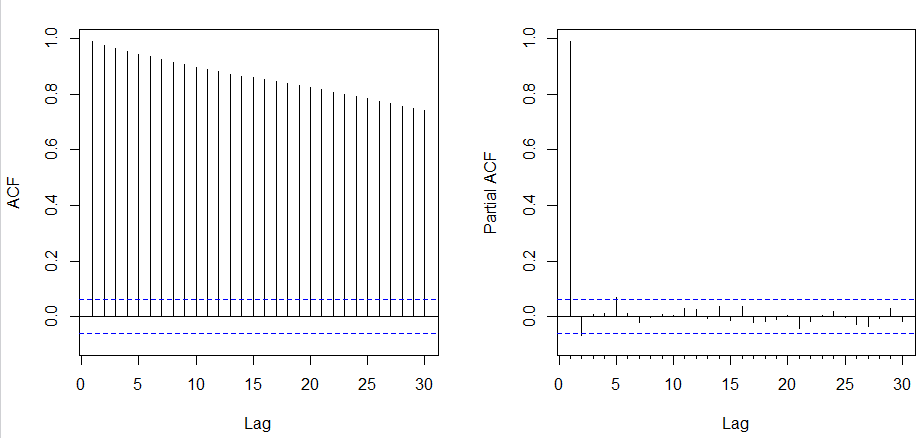
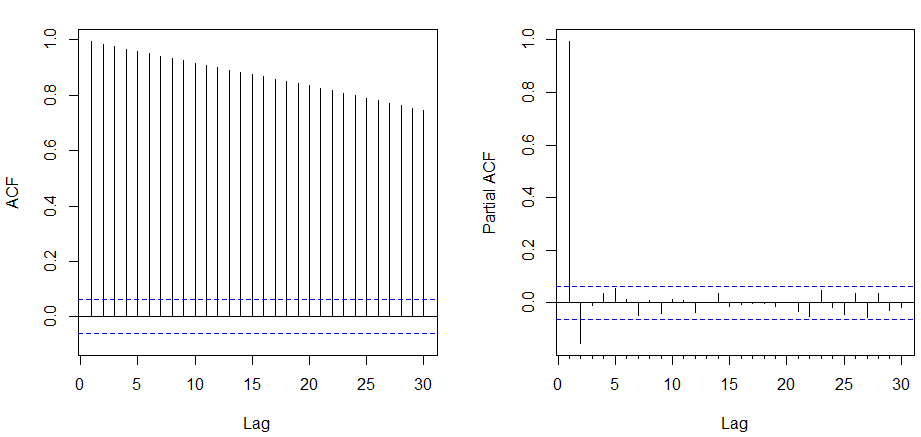
VSGIX VSCIX

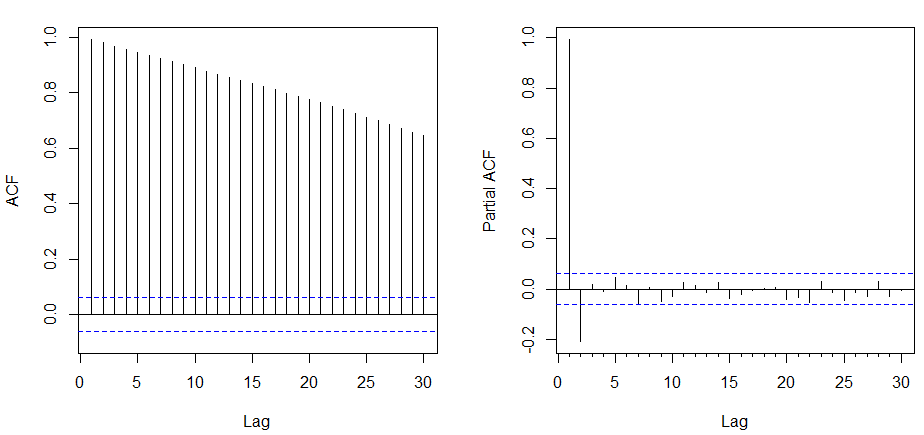
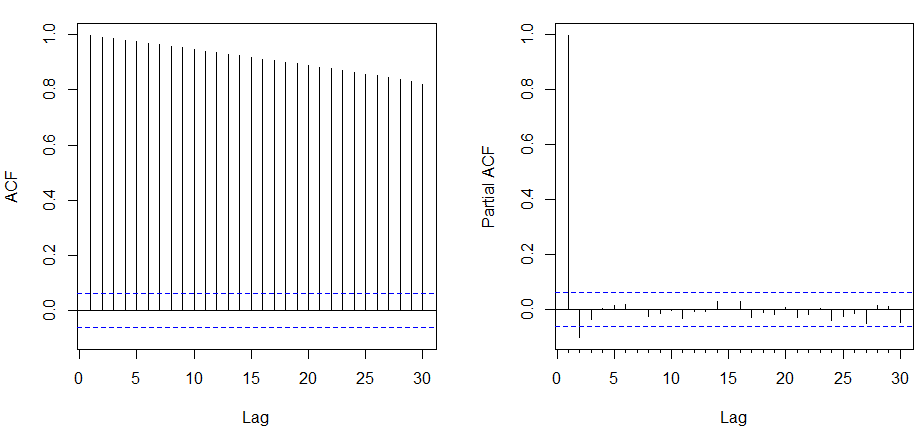
CSRIX SWISX

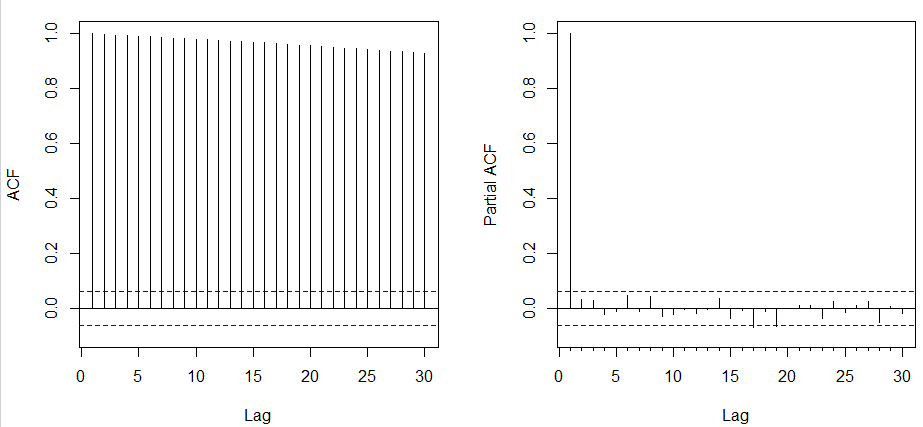
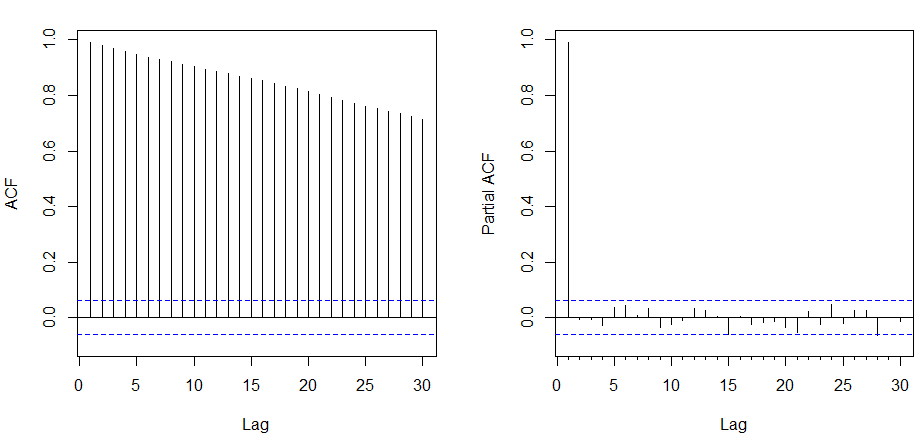
RERGX DFEMX

DEMSX LSBDX

VBTIX VIPSX

1. Residual Diagnostics

