

The Buffet Bet - Basket of Funds vs. All Passive

Stats 107 Final Project

David Wihl

December 6, 2016

Abstract

Type: Simulating a trading strategy. Eight years ago, Warren Buffett made a famous bet: the ultimate passive investment, the Vanguard S&P 500 Index Admiral Mutual Fund (VFIAX), vs. a basket of funds determined by a team of hedge fund experts. Whoever had a lower return after ten years would donate \$1 million to the charity of choice of the winner. After eight years, it looks increasingly likely that Buffett will win the bet.

Contents

1	Introduction	1
2	Baseline	2
2.1	Importing List of Vanguard Mutual Funds	2
2.2	Summary Statistics	2
2.3	Evaluation of VFIAX	3
2.4	Expenses	5
3	Trading Methods	6
3.1	Age Based Balanced Fund	6
3.2	Better Single Fund	7
3.3	Diversified Static Set of Funds	8
3.4	Minimum Variance	8
3.5	CAPM / Sharpe Ratio / Efficient Frontier	9
4	Future Ideas	9
5	Conclusions and Discussions	9
6	References	9
7	Appendix A - Code	9
7.1	Preparing Report	9
7.2	Caching Stock Data	12

1 Introduction

Can a basket of funds beat the market in the long term?

Our project aims to create a Mutual Fund selector for semi-passive investor. A typical Vanguard customer may have to choose from over one hundred mutual funds without any guidance. We will create a portfolio to maximize returns over the long term using several common trading strategies, such as maximizing the Sharpe ratio. In order to make this more realistic, we will take into account fees and transaction costs. Since this is aimed at a typical Vanguard retail customer, no shorting will be allowed.

We will not account for taxes on the assumption that the funds are in a tax deferred account such as 401(k) or IRA. For this reason, we will not evaluate ETFs which are at a disadvantage in a tax deferred account.

The portfolio will start with \$100,000, which is the current average 401(k) balance ([source](#)).

The overall time period will be evaluated whenever possible on a sixteen year history which is the maximum provided by Yahoo via `quantmod`.

The baseline or benchmark will be investing this amount in VFIAX and leaving it there for the duration. The first section examines VFIAX to determine if it is a valid proxy for market both in terms of similar results and reasonable expenses. We will also examine how VFIAX compares to other Vanguard mutual funds in terms of CAGR and Volatility in order to motivate several trading strategies.

The starting fund in the basket will be VTHR, which is one of the typical Vanguard Age-Adjusted balanced funds and a common default fund for many retirement plans.

We will simulate whether rebalancing on a monthly, quarterly, semi-annual or annual basis is the best strategy.

Time permitting, we will evaluate these trading strategies over six windows of ten years each for a more robust estimation of performance.

2 Baseline

2.1 Importing List of Vanguard Mutual Funds

A list was obtained of the 168 currently offered Vanguard mutual funds including published expense ratios. The list was converted into a CSV for easy import into R.

Once the list of funds was imported, `quantmod` was used to read in historical daily adjusted data for the entire date range available from Yahoo. Since the earliest available date was November 13, 2000, an even 16 year end period was chosen, ending November 13, 2016. While more data would have been preferable, it was felt that this was a reasonable sample given it included two significant market corrections in 2001 and again in 2009.

Given the size of data (168 securities, 16 years of daily data), the data was cached locally to allow fast iteration on many models. See appendix for all code.

2.2 Summary Statistics

Several simple descriptive statistics were created: Total Return, Average Daily Return, Standard Deviation / Risk, Sharpe Ratio and CAGR. The following tables summarize the best and worst mutual funds.

Interestingly, in performing these summary statistics, it was determined that `quantmod` does not use adjusted price to calculate returns. Without explicitly using adjusted pricing, the default calculation is wrong. This was first discovered by going through the `quantmod` sources and then confirming at <http://stackoverflow.com/questions/34772616/r-function-periodreturn-not-computing-returns-using-adjusted-closing-prices>. As an extra precaution, values were adjusted upon retrieval using `getSymbol`.

Table 1: Best Sharpe Ratio

Ticker	Fund.Name	Expenses	CAGR	Sharpe
VSCSX	Vanguard Short Term Corp Bd Index Adm	0.10	0.03	0.11
VFSTX	Vanguard Short-Term Investment Grade Inv	0.20	0.04	0.11
VTABX	Vanguard Total Intl Bond Index Adm	0.14	0.04	0.11

Table 2: Worst Sharpe Ratio

Ticker	Fund.Name	Expenses	CAGR	Sharpe
VTIPX	Vanguard Shrt-Term Infl-Prot Sec Idx Inv	0.17	0.00	0
VDVIX	Vanguard Developed Mkts Index Inv	0.20	-0.01	0
VMMSX	Vanguard Emerging Mkts Select Stock Inv	0.93	-0.01	0

Table 3: Best CAGR

Ticker	Fund.Name	Expenses	CAGR	Sharpe
VSIAX	Vanguard Small-Cap Value Index Adm	0.08	0.16	0.03
VMVAX	Vanguard Mid-Cap Value Index Adm	0.08	0.16	0.07
VSGAX	Vanguard Small-Cap Growth Index Adm	0.08	0.13	0.05

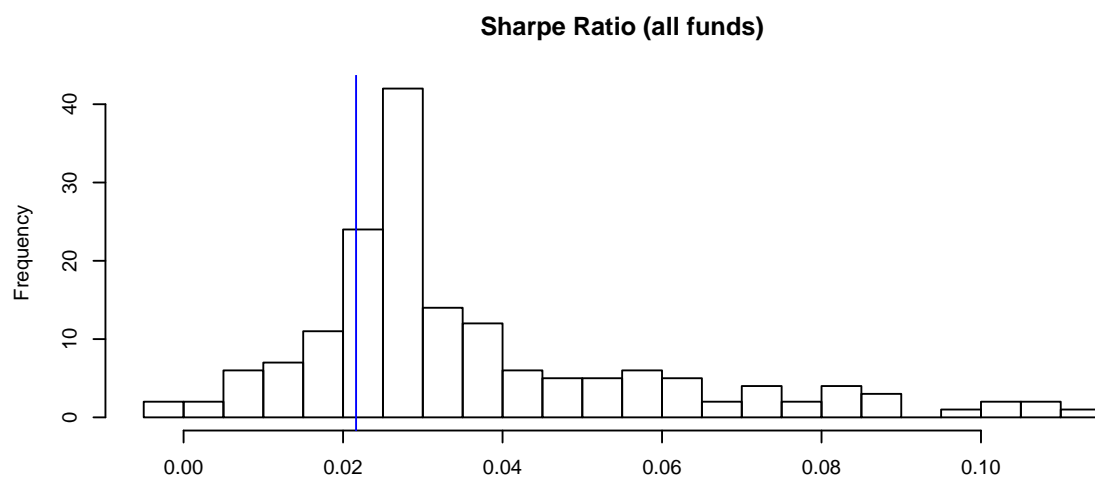
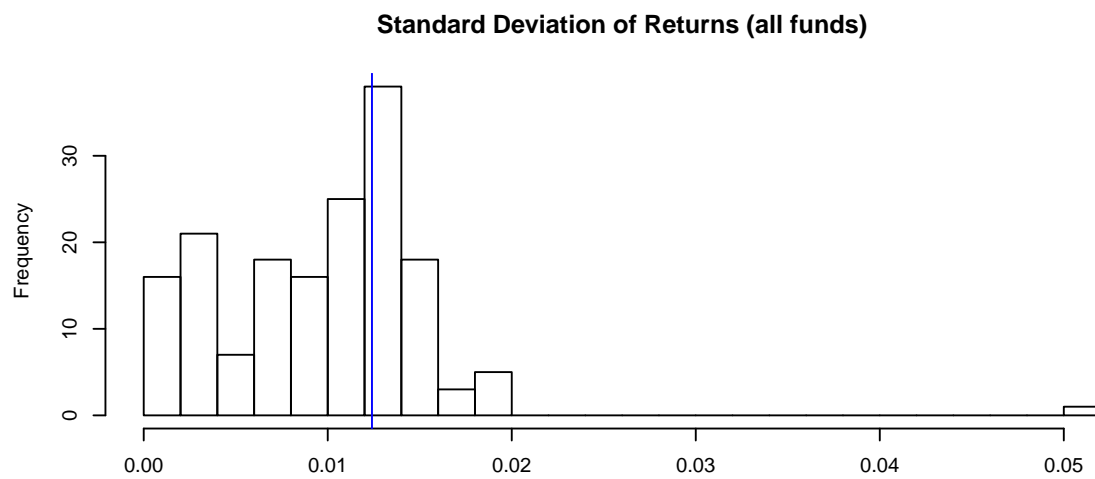
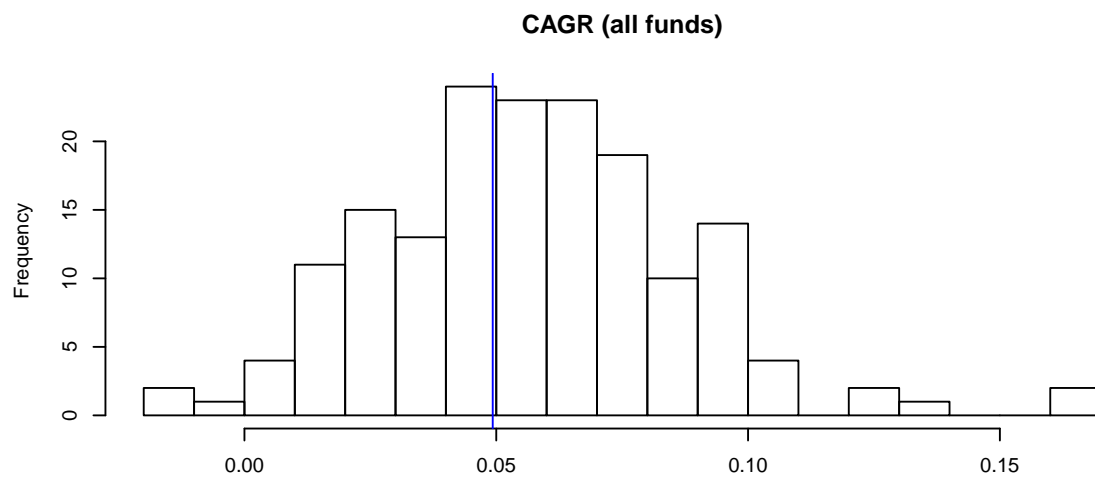
Table 4: Worst CAGR

Ticker	Fund.Name	Expenses	CAGR	Sharpe
VMMSX	Vanguard Emerging Mkts Select Stock Inv	0.93	-0.01	0
VDVIX	Vanguard Developed Mkts Index Inv	0.20	-0.01	0
VTIPX	Vanguard Shrt-Term Infl-Prot Sec Idx Inv	0.17	0.00	0

VMMSX is particularly bad: high expenses, low returns and low Sharpe Ratio.

2.3 Evaluation of VFIAX

The following histograms show the distribution of all funds. VFIAX is the blue line.



Based on these summary statistics and distributions, it appears that there are better mutual fund choices than VFIAX as a combination of higher CAGR and lower volatility.

2.4 Expenses

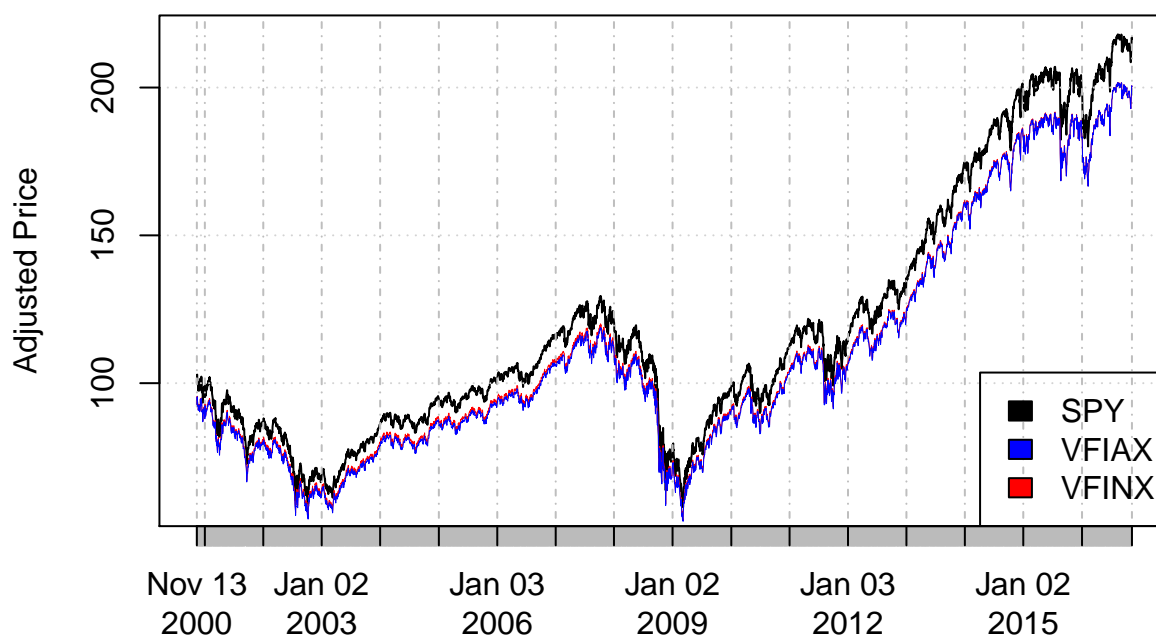
Every mutual fund has expense ratios which is the principle way that a company like Vanguard makes revenue. An expense ratio is a mutual fund's annual operating expense, expressed as a percentage of the fund's average net assets. It's calculated annually and removed from the fund's earnings before they're distributed to investors, directly reducing investors' returns. Since expenses were taken directly from assets, no special calculations needed to be made for expenses, since the costs were already reflected in share prices, and therefore the returns.

For VFIAX, Vanguard claims a 0.05% expense ratio. These are considered "admiral" class shares. For smaller investors, there are "investor" class shares (VFINX) with a 0.16% expense ratio. This expense has gone down over time and will likely continue to do so. We shall now examine if this matches reality and how returns deviate from the benchmark S&P 500.

Metric	VFINX (Investor)	VFIAX (Admiral)
Claimed Expense	0.15%	0.05%
Cum. Exp. 16 Yrs	2.4%	0.8%
Actual Expense	0.8%	0.4%
Difference	67.1%	50%
CAGR	4.9166%	4.9299%

SPY CAGR: 4.9446%

S&P 500 vs. VFIAX, VFINX



By these calculations, expenses add up over time but are lower than advertised. Admiral shares do not offer significant benefit over Investor shares even though the expenses are purported to be 1/3. To be precise, there are less than 2 basis points separating their CAGR. For a very small cost, typically 1.5 basis points of CAGR for Admiral shares, Vanguard index funds offer significant convenience.

3 Trading Methods

From the previous section, we have established that VFIAX:

- represents the S&P 500 market without significant overhead
- may not be an optimal choice. We should be able to do better.

We will now iterate through a number of trading strategies to determine which trading models can best achieve this potential.

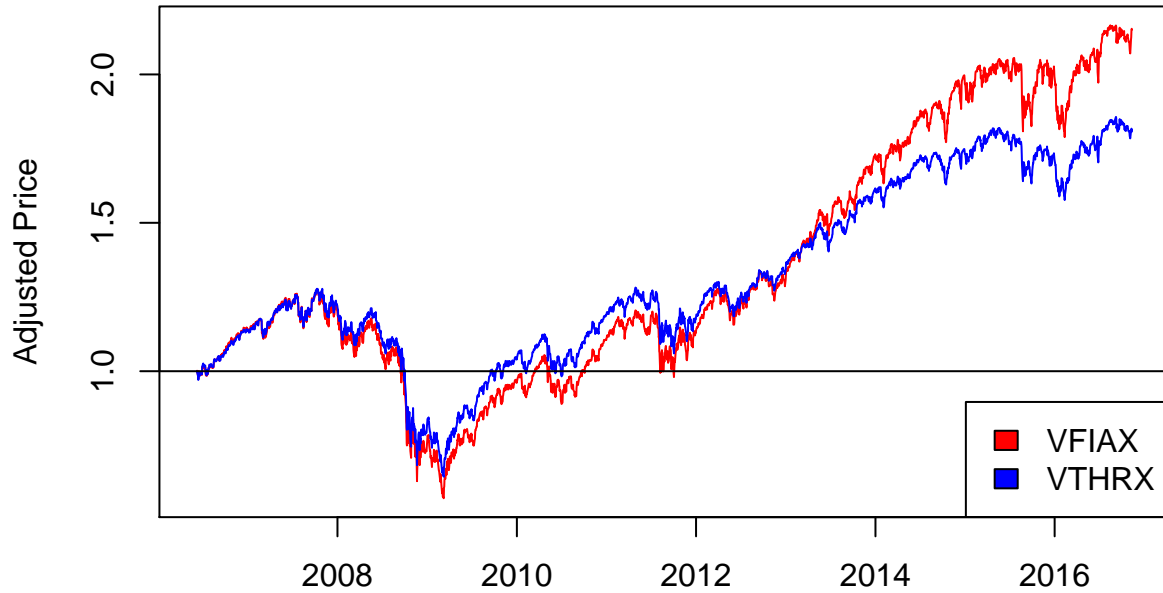
3.1 Age Based Balanced Fund

The default mutual fund for many Vanguard-based retirement plans is age adjusted balanced fund, such as VTHR. Since VTHR is a newer fund, having started on 2006-06-07, we will calculate VFIAX performance over the same time period.

Metric	VFIAX	VTHR
Expense Ratio	0.05%	0.15%
CAGR	7.61%	4.93%

Not looking so good: lower CAGR and higher expenses.

VTHR vs. VFIAX (Normalized)



The age adjusted balanced fund slightly outperforms the market during downturns, but over a longer time period significantly underperforms the market.

3.2 Better Single Fund

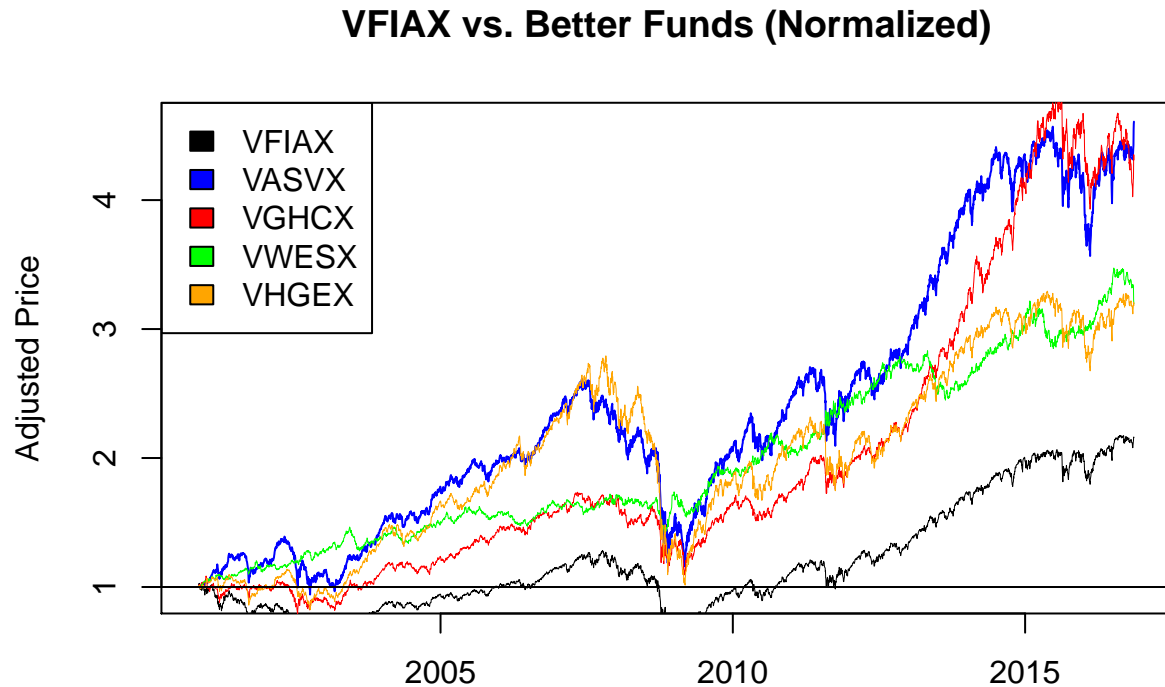
From the selection of funds, is there a fund that outperforms the market by 20%, has been in existence at least 10 years and has lower volatility?

It turns out that there are 26 out of 168 that meet these criteria. Here are the top ten:

Table 7: Top 10 Overall better funds than VFIAX

Ticker	Fund.Name	startDate	Expenses	CAGR	stdDev
VASVX	Vanguard Selected Value Inv	2000-11-13	0.39	0.100	0.012
VCSAX	Vanguard Consumer Staples Index Adm	2004-01-30	0.10	0.097	0.009
VGHAX	Vanguard Health Care Adm	2001-11-12	0.31	0.096	0.010
VGHCX	Vanguard Health Care Inv	2000-11-13	0.36	0.096	0.010
VUIAX	Vanguard Utilities Index Adm	2004-02-02	0.10	0.092	0.011
VHCIX	Vanguard Health Care Index Adm	2004-02-02	0.10	0.091	0.011
VWESX	Vanguard Long-Term Investment Grade Inv	2000-11-13	0.21	0.075	0.007
VHGEX	Vanguard Global Equity Inv	2000-11-13	0.57	0.075	0.012
VLACX	Vanguard Large-Cap Index Inv	2004-01-30	0.20	0.075	0.012
VBLTX	Vanguard Long-Term Bond Index Inv	2000-11-13	0.16	0.073	0.006
VFIAX	Vanguard 500 Index Adm	2000-11-13	0.05	0.049	0.012

Let's plot a selection:



We've discovered that approximately 15% of funds consistently beat the market over a 16 year period across a large number of different asset classes: Health Care Sector, Consumer Staples Sector, even several bond funds.

Should the entire portfolio be put into one of these such as VASVX? Even though the volatility is lower, this would not be wise. Let's now examine more complex strategies that involve diversification to see if risks can be lowered even further at while potentially increasing in CAGR.

3.3 Diversified Static Set of Funds

Vanguard and many other consumer general advice experts recommend a diversified set of static funds.

3.4 Minimum Variance

The original Buffet Bet was the market vs. a hedge fund consisting of a basket of funds. In retrospect, it is not surprising that the market is beating the hedge fund. The primary purpose of the hedge fund is the eponymous hedging - that is reducing risk. Hedge funds outperform the market during downturns but underperform over the long term. Their primary *raison d'être* is capital preservation. This strategy will be mimicked by selecting a minimum variance portfolio. The portfolio will be compared against market downturn periods vs. long term.

3.5 CAPM / Sharpe Ratio / Efficient Frontier

As highlighted earlier in this paper, there are funds such as VSIAX which offer CAGR three times the market albeit at higher risk (See Table 3). By using CAPM, a selection of funds will be made to maximize returns while minimizing risk.

4 Future Ideas

This paper has illustrated several techniques to beating the market over the long term using well known trading techniques. It would be interesting to explore more complex novel techniques that apply modern Machine Learning methods. One area particularly worth exploring would be Reinforcement Learning which is well suited to problems such as maximizing returns while minimizing risks. Furthermore, additional factors like restrictions on number of trades, and cost of trades are relatively easy to model.

5 Conclusions and Discussions

6 References

This entire project code including source data can be found on GitHub (<https://github.com/wihl/stats107-project>)

Ang, Clifford S., *Analyzing Financial Data and Implementing Financial Models Using R*, Springer 2015, ISBN 978-3-319-14075-9 (eBook)

List of Vanguard funds, including fees and transaction costs, supplied by Vanguard <https://investor.vanguard.com/mutual-funds/list#/mutual-funds/asset-class/month-end-returns>

The Buffet Bet, <http://longbets.org/362/>

Fortune, The Buffett Bet <http://fortune.com/2016/05/11/warren-buffett-hedge-fund-bet/>

7 Appendix A - Code

7.1 Preparing Report

The following is the code used to prepare this report:

```
# Load libraries
library(ggplot2)
library(quantmod)
library(knitr)

percent <- function(x, digits = 2, format = "f", ...) {
  paste0(formatC(100 * x, format = format, digits = digits, ...), "%")
}

#
# Import List of Securities
#
loadRData = function(filename) {
  # Source: http://stackoverflow.com/questions/5577221/how-can-i-load-an-object-into-a-variable-name-th
  load(filename)
```

```

    get(ls()[ls()!="filename"])
}
fromDate = "2000-11-13"
toDate   = "2016-11-13"
funds     = read.csv("data/vanguard.csv",header = T, stringsAsFactors = F)
# Load from previously cached download
stockDataEnv= loadRData("data/stockData.RData")
fundData = mget(funds$Ticker,stockDataEnv)
#
# Generate Summary Statistics
#
for (i in 1:nrow(funds)) {
  sym = funds$Ticker[i]
  data = eval(parse(text=paste("fundData$",sym,sep="")))
  if(!is.null(data)){
    # we have the stockdata; Calculate summary statistics
    funds$startDate[i] = as.Date(index(data[1]))
    funds$endDate[i] = as.Date(index(last(data[])))
    startPrice = as.numeric(Ad(data[1]))
    endPrice = as.numeric(Ad(last(data[])))
    funds$totalRet[i] = (endPrice - startPrice) / startPrice
    dailyRets = dailyReturn(Ad(data))
    funds$stdDev[i] = sd(dailyRets)
    funds$avgRet[i] = mean(dailyRets)
    funds$Sharpe[i] = funds$avgRet[i] / funds$stdDev[i]
    # CAGR
    years = as.numeric((as.Date(funds$endDate[i]) -
                           as.Date(funds$startDate[i])))/365
    funds$CAGR[i] = ((endPrice / startPrice)^(1/years)) - 1
  }
}
#
# Generate Tables of Top Performers
#
colsToDisplay = c("Ticker","Fund.Name","Expenses","CAGR","Sharpe")
kable(head(funds[order(-funds$Sharpe),colsToDisplay],n=3), caption="Best Sharpe Ratio",
       row.names = F,digits=2)

kable(head(funds[order(funds$Sharpe),colsToDisplay],n=3), caption="Worst Sharpe Ratio",
       row.names = F,digits=2,n=3)

kable(head(funds[order(-funds$CAGR),colsToDisplay],n=3), caption="Best CAGR",
       row.names = F,digits=2,n=3)

kable(head(funds[order(funds$CAGR),colsToDisplay],n=3), caption="Worst CAGR",
       row.names = F,digits=2,n=3)

#
# Show histograms with base VFIA case
#
par(mfrow=c(3,1))

```

```

# CAGR
hist(funds$CAGR,main="CAGR (all funds)",xlab="",breaks=20)
abline(v=funds[funds$Ticker=="VFIAX",]$CAGR,col="blue")

# Standard Deviation
hist(funds$stdDev,main="Standard Deviation of Returns (all funds)",xlab="",breaks=20)
abline(v=funds[funds$Ticker=="VFIAX",]$stdDev,col="blue")

# Sharpe Ratio
hist(funds$Sharpe,main="Sharpe Ratio (all funds)",xlab="",breaks=20)
abline(v=funds[funds$Ticker=="VFIAX",]$Sharpe,col="blue")
#
# Compare claimed Expense Ratios vs. Actual Expense Ratios
#
getSymbols("SPY", src="yahoo", from=as.Date(fromDate),to=toDate,adjust=T)
p0 = as.numeric(Ad(SPY[1]))
p1 = as.numeric(Ad(last(SPY[])))
tot = (p1 - p0)/p0

# Compound expenses over an expected 16 year period
expectedVFIAX = round((1.0005 ^ 16 - 1) * 100, 2)
expectedVFINX = round((1.0015 ^ 16 - 1) * 100, 2)
# Determine the percentage difference in total return
spyOverVFIAX = round((tot - funds[funds$Ticker=="VFIAX",]$totalRet)/ funds[funds$Ticker=="VFIAX",]$totalRet, 2)
spyOverVFINX = round((tot - funds[funds$Ticker=="VFINX",]$totalRet)/ funds[funds$Ticker=="VFINX",]$totalRet, 2)
pctOverVFIAX = round((expectedVFIAX - spyOverVFIAX) / expectedVFIAX * 100,1)
pctOverVFINX = round((expectedVFINX - spyOverVFINX) / expectedVFINX * 100,1)

years = as.numeric((as.Date(funds[funds$Ticker=="VFIAX",]$endDate)) -
                    as.Date(funds[funds$Ticker=="VFIAX",]$startDate))/365
spy.CAGR = (p1 / p0)^(1/years) - 1

# Plot Difference in Adjusted Price based on Expense Ratio
plot(Ad(SPY), main="S&P 500 vs. VFIAX, VFINX",ylab="Adjusted Price")
lines(Ad(fundData$VFINX),col="red",lwd=.1)
lines(Ad(fundData$VFIAX),col="blue",lwd=.5)
legend("bottomright",c("SPY","VFIAX","VFINX"),fill=c("black","blue","red"))
#
# Compare VFIAX to VTHR over equivalent time periods.
#
# Since VTHR is younger than VFIAX, recalculate CAGR on matching dates
sDate = as.Date(funds[funds$Ticker=="VTHR",]$startDate)
startPrice = as.numeric(Ad(fundData$VFIAX[sDate]))
endPrice = as.numeric(Ad(last(fundData$VFIAX[])))
totRet = (endPrice - startPrice) / startPrice
years = as.numeric((as.Date(funds[funds$Ticker=="VTHR",]$endDate) -
                    as.Date(funds[funds$Ticker=="VTHR",]$startDate)))/365
CAGR = ((endPrice / startPrice)^(1/years)) - 1
f1 = as.numeric(Ad(fundData$VFIAX[paste0(sDate,"::"),])) / startPrice
f2 = as.numeric(Ad(fundData$VTHR)) / as.numeric(Ad(fundData$VTHR[1]))
plot(index(fundData$VTHR),f1,type="l",main="VTHR vs. VFIAX (Normalized)",xlab="",ylab="Adjusted Price")
lines(index(fundData$VTHR),f2,col="blue",lwd=1)
abline(h=1)

```

```

legend("bottomright",c("VFIAX","VTHRX"),fill=c("red","blue"))
#
# Find a better single fund that beats the market with higher CAGR and lower volatility
#
betterFunds = c("VFIAX")
VFIAX.CAGR = funds[funds$Ticker=="VFIAX",]$CAGR
VFIAX.stdDev = funds[funds$Ticker=="VFIAX",]$stdDev
for (i in 1:nrow(funds)) {
  if (funds$CAGR[i] > (1.2 * VFIAX.CAGR)) {
    if (funds$stdDev[i] < VFIAX.stdDev) {
      if (funds$startDate[i] < as.Date("2006-01-01")) {
        #cat(funds$Ticker[i], " ", funds$Fund.Name[i], "\n")
        betterFunds = c(betterFunds, funds$Ticker[i])
      }
    }
  }
}

colsToDisplay = c("Ticker","Fund.Name","startDate","Expenses","CAGR","stdDev")
bf = funds[funds$Ticker %in% betterFunds,colsToDisplay ]
bf$startDate = as.Date(bf$startDate)
kable(rbind(head(bf[order(-bf$CAGR),],n=10), tail(bf[order(-bf$CAGR),],n=1)),
      caption="Top 10 Overall better funds than VFIAX",
      row.names = F,digits=3)

f1 = as.numeric(Ad(fundData$VFIAX)) / as.numeric(Ad(fundData$VFIAX[1]))
f2 = as.numeric(Ad(fundData$VASVX)) / as.numeric(Ad(fundData$VASVX[1]))
f3 = as.numeric(Ad(fundData$VGHCX)) / as.numeric(Ad(fundData$VGHCX[1]))
f4 = as.numeric(Ad(fundData$VWESX)) / as.numeric(Ad(fundData$VWESX[1]))
f5 = as.numeric(Ad(fundData$VHGEX)) / as.numeric(Ad(fundData$VHGEX[1]))

plot(index(fundData$VFIAX),f2,type="l",main="VFIAX vs. Better Funds (Normalized)",xlab="",ylab="Adjusted")
lines(index(fundData$VFIAX),f1,col="black",lwd=0.5)
lines(index(fundData$VFIAX),f3,col="red",lwd=0.5)
lines(index(fundData$VFIAX),f4,col="green",lwd=0.5)
lines(index(fundData$VFIAX),f5,col="orange",lwd=0.5)
abline(h=1)
legend("topleft",c("VFIAX","VASVX","VGHCX","VWESX","VHGEX"),fill=c("black","blue","red","green","orange"))
##
##

```

7.2 Caching Stock Data

Stock data was downloaded and cached with the following code:

```

# Harvard Stats 107, Fall 2016, Dr. Parzen
# by David Wahl
# Download and cache stock symbol data
# Inspired from http://gekkquant.com/2012/06/01/stock-data-download-saving-r/
library(quantmod)
fromDate = "2000-11-13"
toDate   = "2016-11-13"
funds    = read.csv(

```

```

"data/vanguard.csv",
header = T,
colClasses = c("character", "character", "numeric")
)
stocksLst = funds$Ticker
savefilename <- "data/stockdata.RData" #The file to save the data in
startDate = as.Date(fromDate) #Specify what date to get the prices from
maxretryattempts <- 5 #If there is an error downloading a price how many times to retry

#Load the list of ticker symbols from a csv, each row contains a ticker
stockData <- new.env() #Make a new environment for quantmod to store data in
nrstocks = length(stocksLst) #The number of stocks to download

#Download all the stock data
for (i in 1:nrstocks) {
  for (t in 1:maxretryattempts) {
    tryCatch({
      #This is the statement to Try
      #Check to see if the variables exists
      #NEAT TRICK ON HOW TO TURN A STRING INTO A VARIABLE
      #SEE http://www.r-bloggers.com/convertng-a-string-to-a-variable-name-on-the-fly-and-vice-versa-
      if (!is.null(eval(parse(
        text = paste("stockData$", stocksLst[i], sep = "")
      )))) {
        #The variable exists so dont need to download data for this stock
        #So lets break out of the retry loop and process the next stock
        #cat("No need to retry")
        break
      }

      #The stock wasnt previously downloaded so lets attempt to download it
      cat("(",i,":",nrstocks,") ", "Downloading ", stocksLst[i], "\t\t Attempt: ", t, "/", maxretryattempts,
      getSymbols(
        stocksLst[i],
        env = stockData,
        src = "yahoo",
        from = startDate,
        to = toDate,
        adjust = T
      )
    }
    #Specify the catch function, and the finally function
    , error = function(e)
      print(e))
  }
}

#Lets save the stock data to a data file
tryCatch({
  save(stockData, file = savefilename)
  cat("Sucessfully saved the stock data to %s", savefilename)
},
, error = function(e)

```

```
print(e))
```