

Thesis

John Gilheany

2017-05-18

Contents

| | | |
|----------|---|-----------|
| 1 | Opening Comments | 5 |
| 2 | Introduction | 7 |
| 3 | Literature Review | 9 |
| 4 | Data Collection and Summary Statistics | 11 |
| 5 | Data Analysis | 17 |
| 6 | Bibliography | 19 |

Chapter 1

Opening Comments

Chapter 2

Introduction

Chapter 3

Literature Review

Chapter 4

Data Collection and Summary Statistics

Data was downloaded from www.ishares.com for EUSA (iShares MSCI USA Equal Weighted ETF) and USMV (iShares Edge MSCI Min Vol USA ETF), from Oct 31, 2011 to December 31, 2016. iShares are a type of ETF managed by BlackRock, and www.ishares.com contains the month end data for the two ETFs of interest in this dissertation. The data sets included information for the constituents of each ETF of the time, and some other characteristics of them, including: ticker, company name, asset class, weight of the stock relative to the entire index, price per share, number of shares, market value of the position, notional value of the position, sector, sedol number, isin number, exchange that the stock is listed on, and the month end date for the data. An R function was created for both ETFs that would combine each month end data set into one aggregated one. Thus, each month end data set was individually downloaded, then aggregated to create the data sets “usa” and “minvol”. These were stored in the data-raw folder, for safekeeping.

However, after having a quick overview of the data, there were many issues with it that needed to be fixed before the analysis could begin. First, looking at unique exchanges of the data on R, it was seen that there were many foreign exchanges like the Swiss Exchange and the Mexican Exchange, which did not make sense, given the ETF constituents are supposed to be listed on US-based exchanges. One clear discrepancy between the company name and ticker was “AAPL*” (yes, the ticker had an asterisk) which was listed on the Mexican Exchange, instead of being listed as AAPL on the NYSE. This would cause issues for reading the data later on, because AAPL with the asterisk was not read in as “AAPL” due to the asterisk. Other tickers were very easy to notice as incorrect because they included numbers. Seeing as this is a US index with US stocks, this was a clear issue. One example was the company Aflac, Inc. which was recorded by the ticker 8686. This corresponded to Aflac, Inc. on the Tokyo exchange, but when checking the recorded price of the stock for corresponding dates, it matched up with the Aflac, Inc. stock in the US, with ticker AFL. Other tickers were read in correctly, but corresponded to non-US exchanges. One example was BAC, Bank of America, which is listed on the NYSE. Though the price for each date corresponded to the price of BAC in the NYSE, it was listed on the Swiss Exchange. BAC did not correspond to Bank of America on the Swiss Exchange either, which made me question how these companies were being read in.

Instead of looking through 800+ tickers at once, I created a subset of the usa data that looked at all of the companies not listed on the NYSE or NASDAQ exchanges. This included stocks listed on the Boerse Berlin, Bolsa Mexicana De Valores, and London Stock Exchange. There were 16 exchanges that were not the NYSE or NASDAQ, which seemed problematic. There were 74 tickers within these, that I went through and looked at by hand. These included many tickers with numbers and asterisks. In the end, I was able to figure out 67 of these tickers pretty definitively. Some of these tickers were recorded properly, just with the wrong exchange listed (i.e. F, Ford, listed on the Swiss Exchange). Thus, these tickers were kept as is. Some other tickers were harder to find because of M&A activity causing the company to no longer be traded. One example was LinkedIn, which was incorrectly listed as LKI. Before it was acquired by Microsoft, the ticker

was LNKD. Thus, the ticker was changed to show this, as we will need to get old data for companies based off their tickers at the time. Other tickers were for US companies, and included their tickers for overseas exchanges. One example was Intel Corp, which was listed as INCO. INCO is the correct ticker for Intel on the Amsterdam exchange, but not on the NYSE, which lists it as INTC. The general methodology to make sure a change in ticker was appropriate was to check the price of the stock at a specific date, in the usa data set, and then comparing it to the new ticker I was going to assign it. If the price matched, the change was made. If the price did not match up and was very different, then I looked to see if a stock-split might be the cause of this. If there was no evidence of a stock-split, then the stock was highlighted and left alone for further analysis. An entire list of what was changed and what was not is shown in the appendix. In the end, the ticker named “1015736” and Orchard Supply Hardware Stores were removed from the data set. This accounted for less than 0.2% of the data from one month-end date. I also removed cash assets as well by removing values if the sector was cash, since this is not important for our purposes.

Thus, after removing a couple tickers and cash, I wanted to check how accurate the data set was. This was accomplished by comparing the weighted returns of the constructed index we have (looking at each constituent’s monthly return, multiplied by its weight), and comparing it to the ETF return. Because the iShares price data did not reflect stock splits, it was very problematic in some returns. For example, when AAPL split its stock in 2014, from 4-30-2014 to 5-30-2014, the price went from \$633.00 to \$92.93 per share. This would have been improperly recorded as an 85% decrease in the price of the stock, when in reality this was due to a stock split. Since WRDS generally gives price information adjusted for stock splits and dividends, those prices were used instead. However, this was not always the case, and for some instances, these issues were changed manually. This was done by sorting the highest deltas in the data set, with delta corresponding to the monthly change in stock price. I investigated stocks with deltas greater than 0.3 in magnitude for the given month to see if it needed to be adjusted. This same procedure was used for the minvol data set, as it is a subset of the usa data. It is available on the package as usa and minvol.

Before taking the data as given, some checks were done to ensure the data was accurate. First, we wanted to check how our weighted returns compared to the ETF returns. Though we did not expect it to be perfectly correlated, we wanted to aim for a 98% or higher correlation between the weighted returns we calculated, and the ETF returns, on a monthly basis. The results for the usa data are shown below.

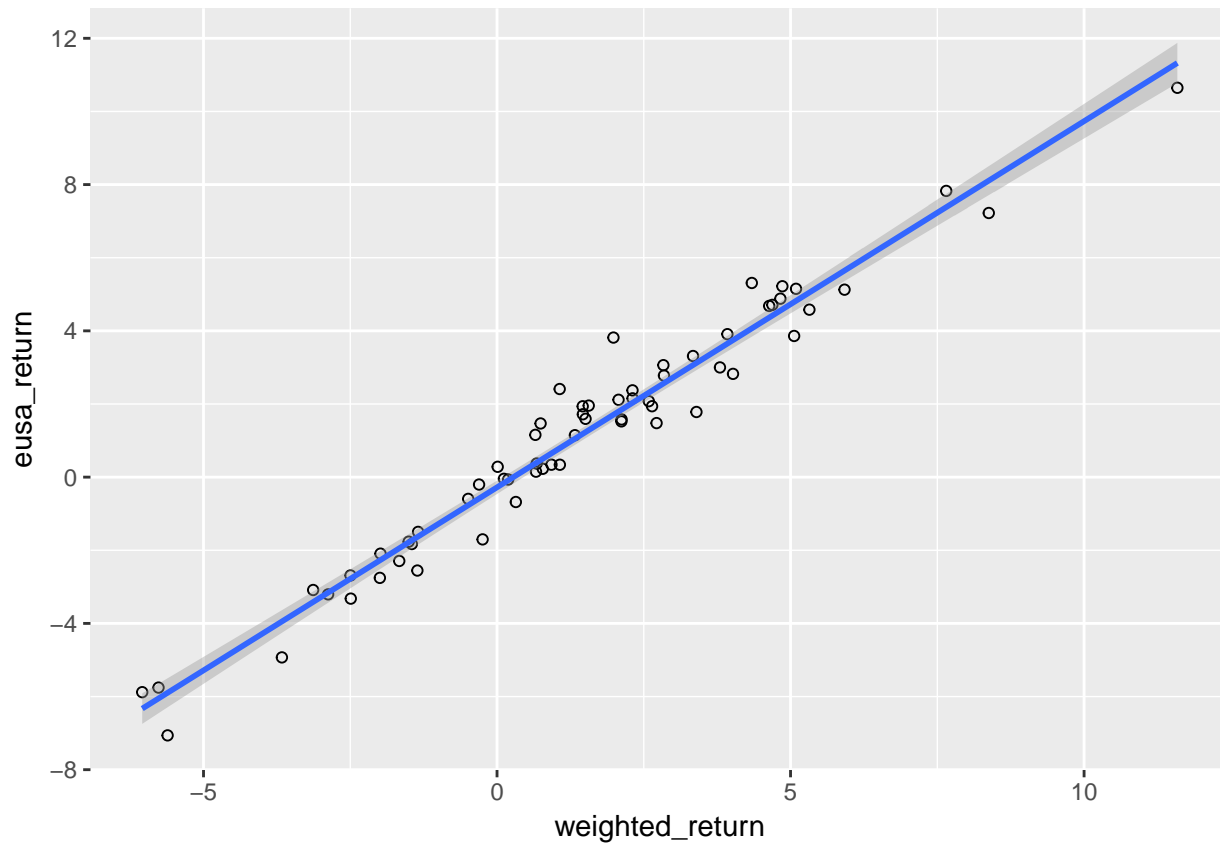
```
# USA data
data(returns1)
summary(returns1)
```

```
##      date      weighted_return  eusa_return.Delt.1.arithmetic
##  Min.   :2011-10-31   Min.      :-6.0439   Min.      :-7.061848
##  1st Qu.:2013-02-14   1st Qu.  :-0.3983   1st Qu.  :-1.088819
##  Median :2014-05-30   Median    : 1.4592   Median    : 1.524721
##  Mean   :2014-05-30   Mean      : 1.3886   Mean      : 1.111701
##  3rd Qu.:2015-09-15   3rd Qu.   : 3.3678   3rd Qu.   : 3.031121
##  Max.   :2016-12-30   Max.      :11.5892   Max.      :10.647011
```

```
library(ggplot2)
ggplot(returns1, aes(x=weighted_return, y=eusa_return)) + geom_point(shape=1) + geom_smooth(method=lm)
```

Table 4.1: USA Returns

| date | weighted_return | eusa_return |
|------------|-----------------|-------------|
| 2011-10-31 | 11.5892350 | 10.64701065 |
| 2011-11-30 | -0.2456645 | -1.70244264 |
| 2011-12-30 | 1.0671034 | 2.40963855 |
| 2012-01-31 | 5.0610248 | 3.86029412 |
| 2012-02-29 | 4.3400696 | 5.30973451 |
| 2012-03-30 | 3.3970199 | 1.78151261 |
| 2012-04-30 | -0.4906636 | -0.59445178 |
| 2012-05-31 | -6.0439231 | -5.88039867 |
| 2012-06-29 | 4.0198762 | 2.82386163 |
| 2012-07-31 | 1.5618633 | 1.95674562 |
| 2012-08-31 | 2.3092650 | 2.15488215 |
| 2012-09-28 | 2.5853910 | 2.07646671 |
| 2012-10-31 | -1.6654955 | -2.29254117 |
| 2012-11-30 | 0.6531763 | 1.15664243 |
| 2012-12-31 | 0.7794097 | 0.22868344 |
| 2013-01-31 | 5.0931233 | 5.14993481 |
| 2013-02-28 | 1.3265195 | 1.14693118 |
| 2013-03-28 | 3.7977193 | 3.00030647 |
| 2013-04-30 | 2.0699705 | 2.11550478 |
| 2013-05-31 | 1.9830989 | 3.81701632 |
| 2013-06-28 | -1.3567937 | -2.55402750 |
| 2013-07-31 | 5.3210605 | 4.57949309 |
| 2013-08-30 | -2.8771461 | -3.20572845 |
| 2013-09-30 | 3.3385738 | 3.31189894 |
| 2013-10-31 | 4.6359683 | 4.68190581 |
| 2013-11-29 | 2.8436501 | 2.77821626 |
| 2013-12-31 | 2.7175078 | 1.47826755 |
| 2014-01-31 | -3.1322783 | -3.08625626 |
| 2014-02-28 | 4.6903592 | 4.71108798 |
| 2014-03-31 | 0.6637837 | 0.14914243 |
| 2014-04-30 | 0.6756760 | 0.37230082 |
| 2014-05-30 | 2.3076650 | 2.37363996 |
| 2014-06-30 | 1.5081588 | 1.59444830 |
| 2014-07-31 | -1.3446626 | -1.49738469 |
| 2014-08-29 | 3.9244135 | 3.90970838 |
| 2014-09-30 | -1.5046634 | -1.76538908 |
| 2014-10-31 | 2.6410836 | 1.93899267 |
| 2014-11-28 | 2.8336728 | 3.06193459 |
| 2014-12-31 | -0.3058795 | -0.20256583 |
| 2015-01-30 | -2.4954545 | -2.68831755 |
| 2015-02-27 | 5.9196423 | 5.12654121 |
| 2015-03-31 | -1.4516707 | -1.83002646 |
| 2015-04-30 | 0.9272040 | 0.34066998 |
| 2015-05-29 | 1.4618722 | 1.71972356 |
| 2015-06-30 | -1.9847732 | -2.09240924 |
| 2015-07-31 | 2.1216832 | 1.57801299 |
| 2015-08-31 | -5.7662469 | -5.75471448 |
| 2015-09-30 | -3.6642971 | -4.92957746 |
| 2015-10-30 | 7.6518942 | 7.82716049 |
| 2015-11-30 | 0.1175917 | -0.04579803 |
| 2015-12-31 | -2.4910425 | -3.32187858 |
| 2016-01-29 | -5.6115203 | -7.06184834 |
| 2016-02-29 | 0.7419356 | 1.46813225 |
| 2016-03-31 | 8.3766775 | 7.22289092 |
| 2016-04-29 | 2.1189998 | 1.52472106 |



```
cor(returns1$eusa_return, returns1$weighted_return)
```

```
##                                [,1]
## Delt.1.arithmetic 0.9805503
```

As we can see, the returns seem pretty consistent and have a correlation greater than 0.98.

Shown below is the data for the min vol index.

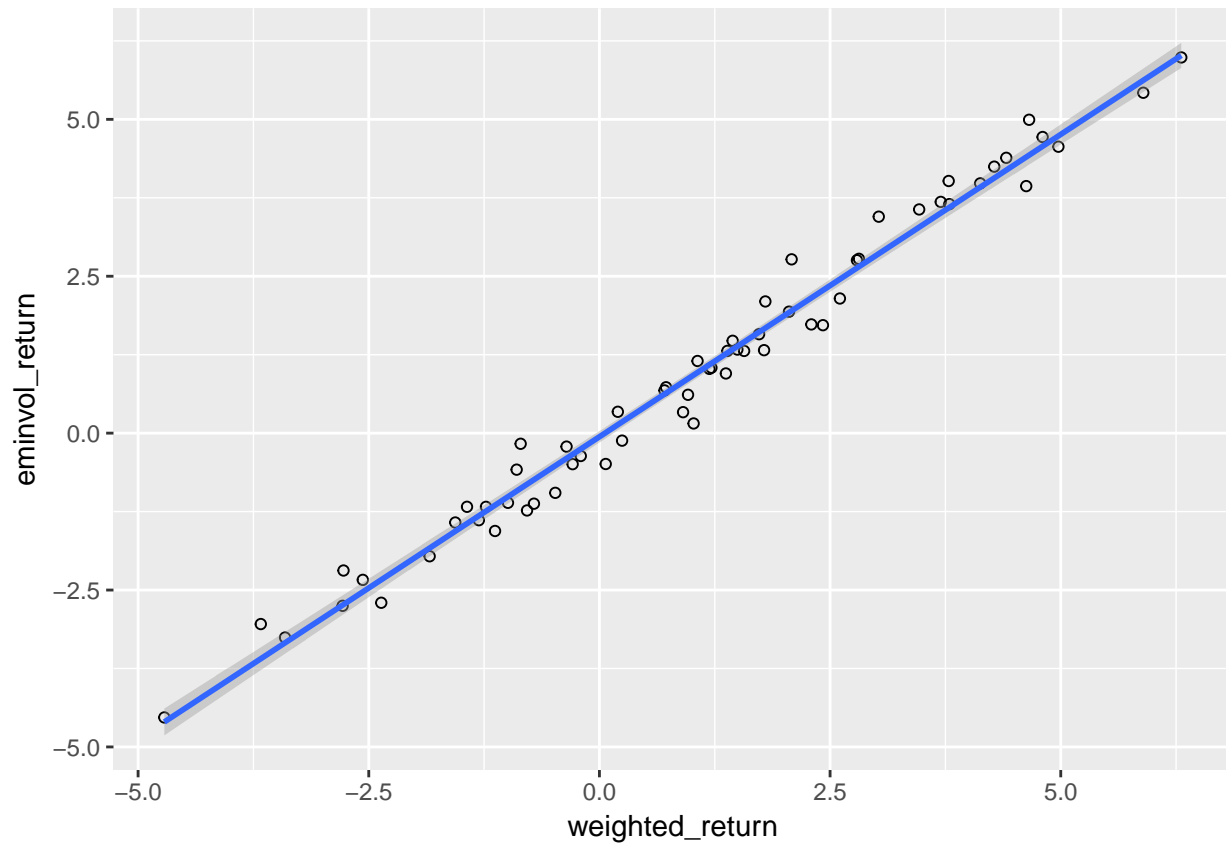
```
# USA data
data(returns2)
summary(returns2)
```

```
##      date      weighted_return  eminvol_return.Delt.1.arithmetic
## Min.   :2011-11-30   Min.   : -4.7156   Min.   : -4.530281
## 1st Qu.:2013-03-07   1st Qu.: -0.8364   1st Qu.: -1.070307
## Median :2014-06-14   Median :  1.1278   Median :  0.987327
## Mean   :2014-06-15   Mean    :  1.0234   Mean    :  0.927892
## 3rd Qu.:2015-09-22   3rd Qu.:  2.7451   3rd Qu.:  2.765673
## Max.   :2016-12-30   Max.    :  6.3064   Max.    :  5.986360
```

```
ggplot(returns2, aes(x=weighted_return, y=eminvol_return)) +
  geom_point(shape=1) + geom_smooth(method=lm)
```

Table 4.2: Min Vol Returns

| | date | weighted_return | eminvol_return |
|----|------------|-----------------|----------------|
| 2 | 2011-11-30 | 1.0207211 | 0.1540832 |
| 3 | 2011-12-30 | 2.0832643 | 2.7692308 |
| 4 | 2012-01-31 | 1.5682330 | 1.3098802 |
| 5 | 2012-02-29 | 1.7985326 | 2.0982638 |
| 6 | 2012-03-30 | 2.6064008 | 2.1455966 |
| 7 | 2012-04-30 | 1.3875986 | 1.3106160 |
| 8 | 2012-05-31 | -2.7737073 | -2.1887347 |
| 9 | 2012-06-29 | 3.0274691 | 3.4491510 |
| 10 | 2012-07-31 | 2.0550335 | 1.9353902 |
| 11 | 2012-08-31 | -0.8537533 | -0.1694915 |
| 12 | 2012-09-28 | 1.3709153 | 0.9507640 |
| 13 | 2012-10-31 | -0.9913378 | -1.1099899 |
| 14 | 2012-11-30 | 0.1997206 | 0.3401361 |
| 15 | 2012-12-31 | -1.1313053 | -1.5593220 |
| 16 | 2013-01-31 | 4.6572484 | 4.9931129 |
| 17 | 2013-02-28 | 2.7913166 | 2.7550016 |
| 18 | 2013-03-28 | 4.9746988 | 4.5643154 |
| 19 | 2013-04-30 | 2.8126037 | 2.7777778 |
| 20 | 2013-05-31 | -2.5643787 | -2.3379863 |
| 21 | 2013-06-28 | -0.4782116 | -0.9512572 |
| 22 | 2013-07-31 | 3.6996474 | 3.6843721 |
| 23 | 2013-08-30 | -3.4070922 | -3.2573290 |
| 24 | 2013-09-30 | 2.4226946 | 1.7202326 |
| 25 | 2013-10-31 | 4.8035140 | 4.7183438 |
| 26 | 2013-11-29 | 1.7845074 | 1.3218391 |
| 27 | 2013-12-31 | 0.7058349 | 0.6806580 |
| 28 | 2014-01-31 | -3.6699298 | -3.0422535 |
| 29 | 2014-02-28 | 4.4096304 | 4.3869843 |
| 30 | 2014-03-31 | 0.9065562 | 0.3339827 |
| 31 | 2014-04-30 | 1.4951304 | 1.3314840 |
| 32 | 2014-05-30 | 1.0635132 | 1.1497399 |
| 33 | 2014-06-30 | 0.7224467 | 0.7307172 |
| 34 | 2014-07-31 | -1.5620464 | -1.4239656 |
| 35 | 2014-08-29 | 4.1272757 | 3.9792859 |
| 36 | 2014-09-30 | -0.7844910 | -1.2319790 |
| 37 | 2014-10-31 | 4.2794278 | 4.2462845 |
| 38 | 2014-11-28 | 3.4656171 | 3.5641548 |
| 39 | 2014-12-31 | 0.0684333 | -0.4916421 |
| 40 | 2015-01-30 | -0.2904233 | -0.4940711 |
| 41 | 2015-02-27 | 3.7922662 | 3.6494538 |
| 42 | 2015-03-31 | -1.4350491 | -1.1736527 |
| 43 | 2015-04-30 | -0.8969496 | -0.5816772 |
| 44 | 2015-05-29 | 1.1920250 | 1.0238908 |
| 45 | 2015-06-30 | -2.3645382 | -2.7027027 |
| 46 | 2015-07-31 | 3.7854982 | 4.0178571 |
| 47 | 2015-08-31 | -4.7156114 | -4.5302814 |
| 48 | 2015-09-30 | -0.7096045 | -1.1238761 |
| 49 | 2015-10-30 | 6.3064130 | 5.9863602 |
| 50 | 2015-11-30 | -0.3562819 | -0.2144900 |
| 51 | 2015-12-31 | 0.2454774 | -0.1194172 |
| 52 | 2016-01-29 | -1.3065154 | -1.3868962 |
| 53 | 2016-02-29 | 1.2139996 | 1.0426770 |
| 54 | 2016-03-31 | 5.8946529 | 5.4235661 |
| 55 | 2016-04-29 | -0.2021128 | -0.3642158 |
| 56 | 2016-05-31 | 1.7304525 | 1.5764222 |



```
cor(returns2$eminvol_return, returns2$weighted_return)
```

```
##           [,1]  
## Delt.1.arithmetic 0.9906848
```

The correlation is 0.99, which is also quite good.

Chapter 5

Data Analysis

Chapter 6

Bibliography