Exercise Set 2: A $300 Billion Strategy

T. Evgeniou

The purpose of this exercise is to become familiar with:

1. Some time series analysis tools;
2. Correlation matrices and principal component analysis (PCA) (see [readings of sessions 3-4](http://inseaddataanalytics.github.io/INSEADAnalytics/Report_s23.html));
3. More data manipulation and reporting tools (including Google Charts).

As always, while doing this exercise we will also see how to generate replicable and customizable reports. For this purpose the exercise uses the R Markdown capabilities (see [Markdown Cheat Sheet](https://www.rstudio.com/wp-content/uploads/2015/02/rmarkdown-cheatsheet.pdf) or a [basic introduction to R Markdown](http://rmarkdown.rstudio.com/authoring_basics.html)). These capabilities allow us to create dynamic reports. For example today's date is 2016-06-08 (you need to see the .Rmd to understand that this is *not* a static typed-in date but it changes every time you compile the .Rmd - if the date changed of course).

Before starting, make sure you have pulled the [exercise set 2 souce code files](https://github.com/InseadDataAnalytics/INSEADAnalytics/tree/master/Exercises/Exerciseset2) on your github repository (if you pull the course github repository you also get the exercise set files automatically). Moreover, make sure you are in the directory of this exercise. Directory paths may be complicated, and sometimes a frustrating source of problems, so it is recommended that you use these R commands to find out your current working directory and, if needed, set it where you have the main files for the specific exercise/project (there are other ways, but for now just be aware of this path issue). For example, assuming we are now in the "Data Analytics R version/INSEADAnalytics" directory, we can do these:

getwd()  
setwd("Exercises/Exerciseset2/")  
list.files()

**Note:** as always, you can use the help command in Rstudio to find out about any R function (e.g. type help(list.files) to learn what the R function list.files does).

Let's now see the exercise.

**IMPORTANT:** You should answer all questions by simply adding your code/answers in this document through editing the file ExerciseSet2.Rmd and then clicking on the "Knit HTML" button in RStudio. Once done, please post your .Rmd and html files in your github repository.

### The Exercise: Introduction

For this exercise we will use the Futures' daily returns to develop what is considered to be a *"classic" hedge fund trading strategy*, a **futures trend following strategy**. There is a lot written about this, so it is worth doing some online search about "futures trend following", or "Managed Futures", or "Commodity Trading Advisors (CTA)". There is about [**$300 billion**](http://www.barclayhedge.com/research/indices/cta/Money_Under_Management.html) invested on this strategy today, and is considered to be one of the **oldest hedge fund strategies**. Some example links are:

* [A fascinating report on 2 centuries of trend following from the CFM hedge - a $6 billion fund](https://www.trendfollowing.com/whitepaper/Two_Centuries_Trend_Following.pdf)
* [Another fascinating report on 1 century of trend following investing from AQR - a $130 billion fund](https://www.aqr.com/library/aqr-publications/a-century-of-evidence-on-trend-following-investing)
* [Wikipedia on CTAs](https://en.wikipedia.org/wiki/Commodity_trading_advisor)
* [Morningstar on CTAs](http://www.morningstar.co.uk/uk/news/69379/commodity-trading-advisors-(cta)-explained.aspx)
* [A report](http://perspectives.pictet.com/wp-content/uploads/2011/01/Trading-Strategies-Final.pdf)
* [Man AHL (a leading hedge fund on CTAs - among others) - an $80 billion fund](https://www.ahl.com)

Of course there are also many starting points for developing such a strategy (for example [this R bloggers one](http://www.r-bloggers.com/system-from-trend-following-factors/) (also on [github](https://gist.github.com/timelyportfolio/2855303)), or the [turtle traders website](http://turtletrader.com) which has many resources.

In this exercise we will develop our own strategy from scratch.

*Note (given today's market conditions):* **Prices of commodities, like oil or gold, can be excellent indicators of the health of the economy and of various industries, as we will also see below**.

### Getting the Futures Data

There are many ways to get futures data. For example, one can use the [Quandl package,](https://www.quandl.com/browse) or the [turtle traders resources,](http://turtletrader.com/hpd/) or (for INSEAD only) get data from the [INSEAD library finance data resources](http://sites.insead.edu/library/E_resources/ER_subject.cfm#Stockmarket) website. One has to pay attention on how to create continuous time series from underlying contracts with varying deliveries (e.g. see [here](https://www.quantstart.com/articles/Continuous-Futures-Contracts-for-Backtesting-Purposes) ). Using a combination of the resources above, we will use data for a number of commodities.

### Data description

Let's load the data and see what we have.

source("helpersSet2.R")  
library(googleVis)  
load("data/FuturesTrendFollowingData.Rdata")

We have data from 2001-01-02 to 2015-09-24 of daily returns for the following 64 futures:

show\_data = data.frame(colnames(futures\_data))  
m1 <- gvisTable(show\_data, options = list(showRowNumber = TRUE, width = 1920,   
 height = min(400, 27 \* (nrow(show\_data) + 1)), allowHTML = TRUE, page = "disable"))  
print(m1, "chart")

### Basic data analysis

Let's see how these are correlated. Let's also make it look nicer (than, say, what we did in Exercise Set 1), using [Google Charts](https://code.google.com/p/google-motion-charts-with-r/wiki/GadgetExamples) (see examples online, e.g. [examples](https://cran.r-project.org/web/packages/googleVis/vignettes/googleVis_examples.html) and the [R package used used](https://cran.r-project.org/web/packages/googleVis/googleVis.pdf) ).The correlation matrix is as follows (note that the table is "dynamic": for example you can sort it based on each column by clicking on the column's header)

We see quite high correlations among some of the futures. Does it make sense? Why? Do you see some negative correlations? Do those make sense?

Given such high correlations, we can try to see whether there are some "principal components" (see [reading on dimensionality reduction](http://inseaddataanalytics.github.io/INSEADAnalytics/Report_s23.html)). This analysis can also indicate whether all futures (the global economy!) are driven by some common "factors" (let's call them **"risk factors"**).

Variance\_Explained\_Table\_results <- PCA(futures\_data, graph = FALSE)  
Variance\_Explained\_Table <- cbind(paste("component", 1:ncol(futures\_data), sep = " "),   
 Variance\_Explained\_Table\_results$eig)  
Variance\_Explained\_Table <- as.data.frame(Variance\_Explained\_Table)  
colnames(Variance\_Explained\_Table) <- c("Component", "Eigenvalue", "Percentage\_of\_explained\_variance",   
 "Cumulative\_percentage\_of\_explained\_variance")

Here is the scree plot (see Sessions 3-4 readings):

eigenvalues <- Variance\_Explained\_Table[, 2]

Let's now see how the 20 first (**rotated**) principal components look like. Let's also use the *rotated* factors (note that these are not really the "principal component", as explained in the [reading on dimensionality reduction](http://inseaddataanalytics.github.io/INSEADAnalytics/Report_s23.html)) and not show any numbers less than 0.3 in absolute value, to avoid cluttering. Note again that you can sort the table according to any column by clicking on the header of that column.

corused = cor(futures\_data[, apply(futures\_data != 0, 2, sum) > 10, drop = F])  
Rotated\_Results <- principal(corused, nfactors = 20, rotate = "varimax", score = TRUE)  
Rotated\_Factors <- round(Rotated\_Results$loadings, 2)  
Rotated\_Factors <- as.data.frame(unclass(Rotated\_Factors))  
colnames(Rotated\_Factors) <- paste("Component", 1:ncol(Rotated\_Factors), sep = " ")  
  
sorted\_rows <- sort(Rotated\_Factors[, 1], decreasing = TRUE, index.return = TRUE)$ix  
Rotated\_Factors <- Rotated\_Factors[sorted\_rows, ]  
Rotated\_Factors[abs(Rotated\_Factors) < 0.3] <- NA

#### Questions:

1. How many principal components ("factors") do we need to explain at least 50% of the variance in this data?
2. What are the highest weights (in absolute value) of the first principal component portfolio above on the 64 futures?
3. Can we interpret the first 10 components? How would you call these factors?
4. Can you now generate the principal components and scree plot using only: a) the pre-crisis bull market years (e.g. only using the data between November 1, 2002, and October 1, 2007)? b) the financial crisis years (e.g. only using the data between October 1, 2007 and March 1, 2009), (Hint: you can select subsets of the data using for example the command `crisis\_data = futures\_data[as.Date(rownames(futures\_data)) > "2007-10-01" & as.Date(rownames(futures\_data)) < "2009-03-01", ])
5. Based on your analysis in question 3, please discuss any differences you observe about the futures returns during bull and bear markets. What implications may these results have? What do the results imply about how assets are correlated during bear years compared to bull years?
6. (Extra - optional) Can you create an interactive (shiny based) tool so that we can study how the "**risk factors**" change ove time? (Hint: see [Exercise set 1](https://github.com/InseadDataAnalytics/INSEADAnalytics/blob/master/Exercises/Exerciseset1/ExerciseSet1.Rmd) and online resources on [Shiny](http://rmarkdown.rstudio.com/authoring_shiny.html) such as these [Shiny lessons](http://shiny.rstudio.com/tutorial/lesson1/). Note however that you may need to pay attention to various details e.g. about how to include Google Charts in Shiny tools - so keep this extra exercise for later!).

**Your Answers here:**

### A Simple Futures Trend Following Strategy

We can now develop a simple futures trend following trading strategy, as outlined in the papers in the Exercise Introduction above. There are about $300 billion invested in such strategies! Of course we cannot develop here a sophisticated product, but with some more work...

We will do the following:

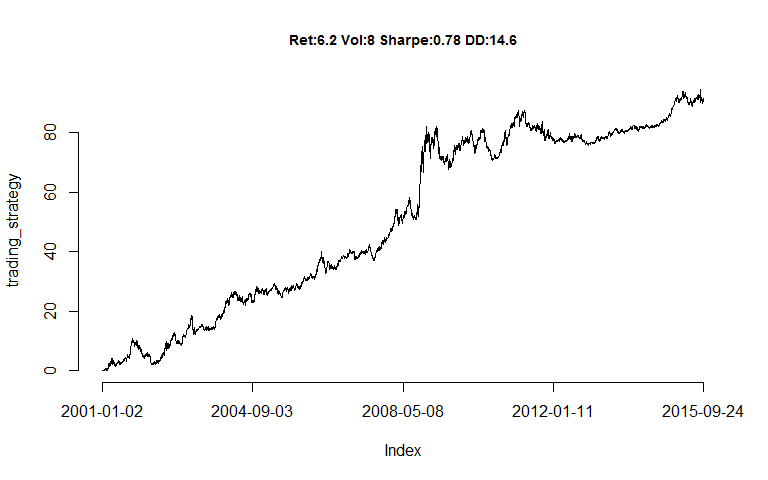
1. Calculate a number of moving averages of different "window lengths" for each of the 64 futures - there are [many](http://www.r-bloggers.com/stock-analysis-using-r/) so called [technical indicators](http://www.investopedia.com/active-trading/technical-indicators/) one can use. We will use the "moving average" function ma for this (try for example to see what this returns ma(1:10,2) ).
2. Add the signs (can also use the actual moving average values of course - try it!) of these moving averages (as if they "vote"), and then scale this sum across all futures so that the sum of their (of the sum across all futures!) absolute value across all futures is 1 (hence we invest $1 every day - you see why?).
3. Then invest every day in each of the 64 an amount that is defined by the weights calculated in step 2, using however the weights calculated using data until 2 days ago (why 2 days and not 1 day?) - see the use of the helper function shift for this.
4. Finally see the performance of this strategy.

Here is the code.

signal\_used = 0 \* futures\_data # just initialize the trading signal to be 0  
# Take many moving Average (MA) Signals and let them 'vote' with their sign  
# (+-1, e.g. long or short vote, for each signal)  
MAfreq <- seq(10, 250, by = 20)  
for (iter in 1:length(MAfreq)) signal\_used = signal\_used + sign(apply(futures\_data,   
 2, function(r) ma(r, MAfreq[iter])))  
# Now make sure we invest $1 every day (so the sum of the absolute values of  
# the weights is 1 every day)  
signal\_used = t(apply(signal\_used, 1, function(r) {  
 res = r  
 if (sum(abs(r)) != 0)   
 res = r/sum(abs(r))  
 res  
}))  
colnames(signal\_used) <- colnames(futures\_data)  
# Now create the returns of the strategy for each futures time series  
strategy\_by\_future <- scrub(shift(signal\_used, 2) \* futures\_data) # use the signal from 2 days ago  
# finally, this is our futures trend following strategy  
trading\_strategy = apply(strategy\_by\_future, 1, sum)  
names(trading\_strategy) <- rownames(futures\_data)

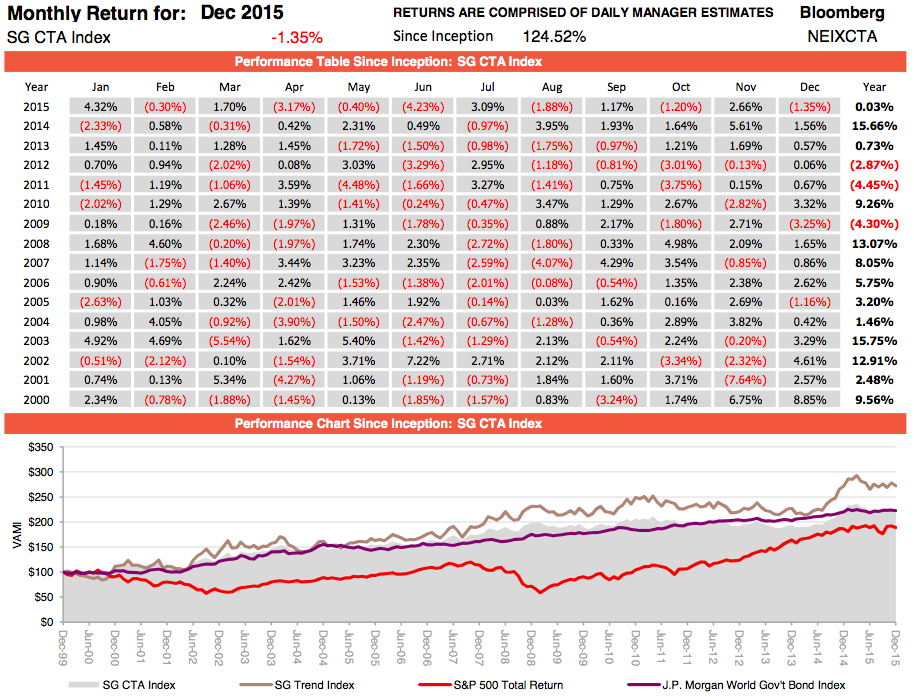
### Reporting the performance results

Let's see how this strategy does:



Here is how this strategy has performed during this period.

How does this compare with **existing CTA products** such as [this one from Societe Generale?](https://cib.societegenerale.com/fileadmin/indices_feeds/SG_CTA_Monthly_Report.pdf) (Note: one can easily achieve a correlation of more than 0.8 with this specific product - as well as with many other ones)



Compare our strategy with this product

#### Questions

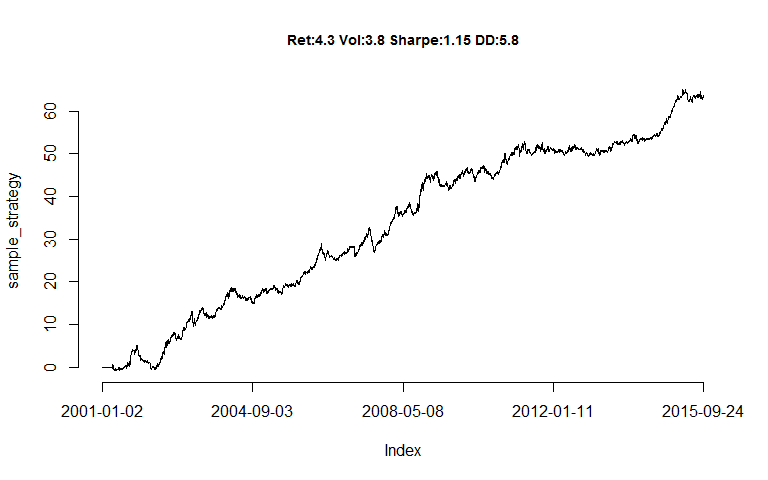
1. Can you describe in more detail what the code above does?
2. What happens if you use different moving average technical indicators in the code above? Please explore and report below the returns of a trading strategy you build. (Hint: check that the command line MAfreq<-seq(10,250,by=20) above does for example - but not only of course, the possibilities are endless)

**Your Answers here:**

### A class competition

Now you have seen how to develop some trading strategies that hedge funds have been using for centuries. Clearly this is only the very first step - as many of the online resources on technical indicators also suggest. Can you now explore more such strategies? How good a **futures trend following hedge fund strategy** can you develop? Let's call this.... a **class competition**! Explore as much as you can and report your best strategy as we move along the course...

Here is for example something that can be achieved relatively easily...



Here is how this strategy has performed during this period.

**Finally**: One can develop (shiny based) interactive versions of this report and deploy them using shinyapps::deployApp('ExerciseSet2.Rmd') (you need a [shinyapps.io](https://www.shinyapps.io) account for this). This is for example an [interactive version of this exercise.](https://inseaddataanalytics.shinyapps.io/ExerciseSet2/)

As always, **have fun**