https://doi.org/10.26219/heal.aueb.6490

ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ
ATHENS UNIVERSITY OF ECONOMICS AND BUSINESS

SCHOOL OF INFORMATION SCIENCES & TECHNOLOGY

DEPARTMENT OF STATISTICS

POSTGRADUATE PROGRAM

Financial Modelling Using Threshold Models

By

Alexandros Savvelis

A THESIS

Submitted to the Department of Statistics
of the Athens University of Economics and Business
in partial fulfilment of the requirements for
the degree of Master of Science in Statistics

Athens, Greece

September 2018

https://doi.org/10.26219/heal.aueb.6490

https://doi.org/10.26219/heal.aueb.6490

ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ
ATHENS UNIVERSITY OF ECONOMICS AND BUSINESS

ΣΧΟΛΗ ΕΠΙΣΤΗΜΩΝ & ΤΕΧΝΟΛΟΓΙΑΣ ΤΗΣ ΠΛΗΡΟΦΟΡΙΑΣ

ΤΜΗΜΑ ΣΤΑΤΙΣΤΙΚΗΣ

ΜΕΤΑΠΤΥΧΙΑΚΟ ΠΡΟΓΡΑΜΜΑ

Χρηματοοικονομική μοντελοποίηση με την χρήση υποδειγμάτων Threshold

Αλέξανδρος Σαββέλης

ΔΙΑΤΡΙΒΗ

Που υποβλήθηκε στο Τμήμα Στατιστικής
του Οικονομικού Πανεπιστημίου Αθηνών
ως μέρος των απαιτήσεων για την απόκτηση
Μεταπτυχιακού Διπλώματος Ειδίκευσης στη Στατιστική

Αθήνα

Σεπτέμβριος 2018

https://doi.org/10.26219/heal.aueb.6490

https://doi.org/10.26219/heal.aueb.6490

Περίληψη

Αλέξανδρος Σαββέλης

Χρηματοοικονομική μοντελοποίηση με την χρήση υποδειγμάτων Threshold
Σεπτέμβριος 2018

Σκοπός της συγκεκριμένης διπλωματικής εργασίας είναι η εξερεύνηση των μοντέλων και αλγόριθμων που χρησιμοποιούν thresholds σε οικονομικά δεδομένα. Οι αλγόριθμοι που χρησιμοποιήθηκαν είναι τα regression trees και τα random forest ενώ τα μοντέλα που εκτιμήθηκαν είναι οι μη γραμμικές GARCH. Προσπαθήσαμε να εξερευνήσουμε τον συνδυασμό μεταξύ της στατιστικής μάθησης και των χρηματοοικονομικών μοντέλων για να προβλέψουμε και να εξηγήσουμε τα χαρακτηριστικά αυτών των δεδομένων. Τα χρηματοοικονομικά δεδομένα που χρησιμοποιήθηκαν είναι αυτά της βιομηχανίας των Hedge funds. Μια βιομηχανία διαφορετική από όλες τις άλλες με τα δικά της χαρακτηριστικά. Συγκεκριμένα ο δείκτης που μελετήσαμε είναι ο CSTHFI όπου είναι αντιπροσωπευτικός όλης της αγοράς των hedge funds. Τα συμπεράσματα που καταλήξαμε είναι ότι το καλύτερο μοντέλο ως προς την πρόβλεψη είναι το random forest χρησιμοποιώντας ως κριτήρια την ρίζα του μέσου τετραγωνικού σφάλματος, τον μέσο του απόλυτου σφάλματος και την διάμεσο του απόλυτου σφάλματος ενώ από τα γραμμικά και μη γραμμικά GARCH μοντέλα το πιο δύσκολο για να εκτιμηθεί είναι το switching volatility GARCH.

https://doi.org/10.26219/heal.aueb.6490

Abstract

Alexandros Savvelis

Financial Modelling Using Threshold Models September 2018

The purpose of this thesis is to explore threshold algorithms and models in financial data. The algorithm that used was regression trees and random forest and the estimated models was nonlinear GARCH models. We try to explore this combination between statistical learning algorithms and financial econometric models to predict and explain the characteristics of the data. The financial data that used was from the hedge fund industry. An industry different from the rest of the market with its own characteristics. Specifically the index that we use is CSTHFI index which represent the overall industry and we can conclude in very interesting fact. Overall the best model for one step ahead out of sample prediction using as a criterion the root mean square error, the median absolute error and the mean absolute error is the random forest and from the linear and nonlinear GARCH models the most difficult to estimate is the switching volatility GARCH.

https://doi.org/10.26219/heal.aueb.6490

Contents

 https://doi.org/10.26219/heal.aueb.6490

List of Figures

https://doi.org/10.26219/heal.aueb.6490

List of Tables

https://doi.org/10.26219/heal.aueb.6490

1. Introduction

1.1 General

Thresholds has a very long history in statistics. In finance thresholds has many different applications. The reason for that is because this models can define different states of the world or regimes and to allow for the possibility that the dynamic behaviour of economic variables depend on the regime that occurs at any given point in time according to Frances and Van Dijk (2000) (something that in practice happen for example economic crisis and periods of high volatility are much different than periods of low volatility). Of course this implies that each regime has different characteristics such as mean, variance and autocorrelation. The regimes can change either deterministically or stochastically which for the first one means that the change point is known in advance and the second that the change point is a stochastic process and unknown. In this thesis we are only consider deterministic regimes.

First of all, the threshold autoregressive (TAR) model developed by Tong (1978), Tong and Lim (1980), an extensive analysis has been made by Tong (1990) in his book and has been enormously influential in economics for the past four decades. Many articles has been written for TAR, the inference and the forecasting that can be made from them. Hansen (1997) in his article explore the statistical properties of threshold estimator and the likelihood ratio statistic for testing the hypotheses concerning the unknown threshold and applied the theory in U.S unemployment rates. Also Hansen (2011) wrote that forecasting of TAR compere to traditional linear model using as a criterion out-of-sample mean square forecast error has mixed reviews in the literature. Another application of TAR can be found in the articles of Gospondinov (2005), who used this model in order to test for Threshold nonlinearity in short term interest rates. The TAR model assumes that the regime is determined by a variable q_t and the threshold value, which is denoted as *c*. The TAR has many different special cases for example the Self-Exciting TAR (SETAR) model that assumes the threshold variable is the lagged variable of the time series $q_t = y_{t-d}$ for a specific integer $d > 0$. The SETAR model with two regimes and *d*=1 has the following form

https://doi.org/10.26219/heal.aueb.6490

$$y_t \begin{cases} \varphi_{0,1} + \varphi_{1,1}y_{t-1} + \varphi_{2,1}y_{t-2} + \ldots + \varphi_{p,1}y_{t-p_1} + \varepsilon_t \text{ if } y_{t-1} \leq c \\ \varphi_{0,2} + \varphi_{1,2}y_{t-1} + \varphi_{2,2}y_{t-2} + \ldots + \varphi_{p,2}y_{t-p_2} + \varepsilon_t \text{ if } y_{t-1} > c \end{cases}$$

So practically there is an indicator function I[·] in the lagged variable and if you are in one regime you can't be in the other. Special cases emerged when we replace the indicator function. Specifically the models called Smooth Transition Autoregressive (STAR) and instead of an indicator function uses a transition function in order to do the switching regime process continuous and smoother. Depending the function, we have LSTAR (Logistic function) and ESTAR (Exponential function) models. Many articles has been written for this models like Terasverta (1994) who estimated the LSTAR and ESTAR model using nonlinear least squares determining the delay parameter and choosing between the two models, Umer, Sevil, Sevil (2018) where they compare the forecasting ability of STAR model against AR for 1-12 step-ahead out- of-sample forecasting the Turkey Travel & Leisure Index, FTSE Global Travel & Leisure Index and MSCI World Index using as a criterion the RMSE, MAE and Diebold-Mariano test and the result was that the nonlinear STAR model doesn't improve the out of sample prediction against the AR and Luukkonen, Saikkonen and Terasvirta (1988) who presented three tests for testing nonlinearity. Practically they used AR against the STAR model and then trying to specify the function of STAR model between ESTAR and LSTAR and discuss their properties of the tests.

Another extension of TAR models is the Mixture Autoregressive (MAR) which discussed by Lanne and Saikkonen (2003). In their articles the MAR model that proposed tries to bring more flexibility in the TAR models. It uses an independent and identical random variable in the threshold in order to randomize the switching regimes. The model has the following form

$$y_t = (v_1 + b_1 y_{t-1}) I(y_{t-1} < c + \eta_t) + (v_2 + b_2 y_{t-1}) I(y_{t-1} \geq c + \eta_t) + \sigma\varepsilon_t$$

which η_t is the independent variable and as we can see even if the threshold variable y_{t-1} is above the threshold value c the switch in the regimes may not be happen for specific values of η_t. For the above models (except maybe MAR) there is a multivariate generalization of the regime switching models. Tsay (1998) propose a procedure for building a multivariate threshold model in order to find the difference between an asset that simultaneously traded in two markets for capitalized the difference between them and used Akaike information Criterion (AIC) in order to determine the thresholds and the model and then use the modelling procedure to analyse the U.S. monthly interest rates and two daily river flow series of Iceland.

https://doi.org/10.26219/heal.aueb.6490

Of course we have also nonlinear models for volatility (Nonlinear GARCH models). The reason that we have this models is due to the fact that typical GARCH model can't capture the leverage effect instead of Nonlinear GARCH models which they can but also the asymmetric effect of variance (h_t). An extensive analysis in different models of that type made by Frances and Van Dijk (2000) in their book (chapter 4) for example ANST-GARCH (Asymmetric Nonlinear Smooth Transition GARCH) which developed by Anderson , Nam and Vahid (1999) , GARCH-GJR introduced by Glosten, Jagannathan and Runkle (1993) which change the coefficient of the shock model according to the sign of the shock, VS-GARCH (volatility-switching GARCH) developed by Fornari and Mele (1996, 1997) and is an extension of GARCH-GJR which change the whole model according to the sign of the shock and many more.

Many articles has been developed for nonlinear GARCH models, some of them is by Audrino and Buhlman (2000) using a tree structured GARCH model for estimating the volatility in stationary financial time series. What they actually did was to use a binary tree model and in each terminal node there is a GARCH(1,1) estimated with maximum likelihood the threshold and the modes. Also comparison between linear and nonlinear GARCH has been done by Gokcan (2000) using GARCH and nonlinear GARCH model specifically EGARCH for modelling the volatility and for the returns the best appropriate model between AR, MA and ARMA for analysing seven different countries the monthly stock market portfolio returns and found out that GARCH performed better even if stock market return series display skewed distributions. Another article about nonlinear GARCH has been written by Wu (2011) and give an extensive analysis in Threshold GARCH model, its theoretical properties and application. Nonlinear GARCH can also change regimes more smoothly. Gonzalez-Rivera (1998) test and estimate Smooth Transition GARCH (ST-GARCH) using quasi maximum likelihood (QML) in exchange rates and stock returns. A very interesting topic is the forecasting performance of nonlinear GARCH model with linear GARCH. Munisamy and Ramasamy (2012) used GARCH , GARCH-GJR and EGARCH to evaluate the predictive performance in four selected exchange rates. Specifically Australian Dollar, Singapore Dollar, Thailand Bhat and Philippine Peso used and found that the leverage effect incorporated in EGARCH and GARCH-GJR do not improve the results.

Threshold models doesn't exist only in time series framework. One of the most famous algorithm is the Classification and regression algorithm (CART) developed by Breiman et. al

https://doi.org/10.26219/heal.aueb.6490

(1984) and the algorithm to work divides the space into different regimes and in each regime there is a prediction sometimes more complicate such as a linear model or the average. Tree models also has been used in finance for example Dingli and Fournier (2017) use various machine learning algorithms for regression problems (and for classification) one of them was regression trees and random Forest and reported that the support vector machine and linear regression gives better results in the next day prediction of stock markets in financial and tech industry. Abe and Nakayama (2018) use deep learning technique to predict one step ahead the stock returns in the Cross-Section index and compered them with the SVM and the Random Forest and discover that a deep neural network outperform both of them.

All the above models estimated using a frequentists approach. Of course can be estimated by a Bayesian framework for example Chiang , Chen and So (2002) analyse asymmetries in global stock market using a double threshold autoregressive GARCH model (DTAR-GARCH) estimated with a Bayesian method and with that model detect significant evidence to sustain the asymmetrical hypothesis of stock index returns based on four different major factor indices. Also they found that negative news can cause a larger decline in a national stock index than positive with the same magnitude. Vrontos and Giannikis (2008) detecting nonlinear risk exposure in hedge fund strategies using a flexible threshold regression model with GARCH model estimated with a Bayesian method. In order to find the important risk factors and thresholds they presented a computationally flexible Markov Chain Monte Carlo and conclude that the proposed threshold regression model can improve the ability to evaluate the hedge fund performance.

1.2 Data Characteristics

Forecasting in finance is one of the most difficult and demanding things. For the last decades many different methods has been used to forecast stock prices, financial indices and many more financial products. From this long term analysis many different characteristics has been discovered but some of the characteristics are met most of the times (stylized facts) in empirical applications according to Vrontos (2015). This characteristics are the following:

- Fat tails
- Volatility Clustering

 https://doi.org/10.26219/heal.aueb.6490

- Leverage effect
- Non-trading days
- Co-movements of volatility

Specifically fat tails is that the distribution of the returns has higher probability occurring extreme values compare to the probability of normal distribution gives. This characteristic also called leptokurtosis and can be found using descriptive statistics when the kurtosis of the distribution of the returns is bigger than 3 which means it is bigger compared to the kurtosis that should have if the data was normally distributed. When this happens the proper distribution to describe the returns is usually Student-t or Generalized Error distribution.

Another interesting phenomenon that financial returns/indices have is the volatility clustering phenomenon. According to Mandelbrot (1963) '…large changes tend to be followed by large changes - of either sign – and small changes tend to be followed by small changes…' which means that if the returns have high/low volatility will keep have high/low volatility in this specific point in time. So clusters appear in financial returns depend from two things: the risk and the time. As the time interval changes the volatility will change and that's why we need to use models to capture this characteristic. The volatility clustering phenomenon is an indication of heteroscedasticity which means that some data in the dataset have different volatility from the rest. Of course this phenomenon can be detected using time series analysis diagrams or using specific test in the squared /absolute returns of model residuals.

The leverage effect is that the volatility is different depending of the sign of the returns. If the index/return is high positive, the volatility will be different compare to high negative returns because the market in the second case are more unstable. In other words negative returns can produce higher volatility than positive and this means that market evaluate the news differently. In other words news can be characterized as bad or good depend the sign and the value of the returns. An example of bad news is a trade war which can shake an economy and we expect high volatility the next days. So there is a need for modelling the volatility with models that can capture this characteristic.

Non-trading days or the 'Holidays Effect' is a phenomenon which information is gathered and hasn't yet capitalized by the market. This means that there is news for indices/stocks that involve the market and because the stock exchange is closed the brokers/market doesn't capitalized it yet. This happens usually after weekend or holidays.

 https://doi.org/10.26219/heal.aueb.6490

Last but not least is the co-movements in volatilities of returns. It is known that market assets are correlated in the sense that two or more stocks on the same industry can be affected by the same factors. So if the volatility increase in one of them there will be probably increase in the others. Even if the markets are in different countries there still can be co-movements in volatilities, so there is a homogeneity in volatility changes of the market and in order to create models more accurate we need to take into account this information.

1.3 Hedge funds

The history of hedge funds is estimated in early 1920 where the market was rising, many private investments was starting to create with the most known the Graham-Newman Partnership founded by Benjamin Graham and his partner Jerry Newman according to Wikipedia. The term "hedge fund" is created by the sociologist Alfred W. Jones, who used the well-known phrase "hedged" in wall streets. The name hedge funds comes from hedging which is actually the practice of reducing the risk, but nowadays the purpose of hedge funds primarily the maximization of returns on investments. Hedge funds are practically pooled funds and as I already mention try to explore numerus strategies to maximize the profits using various techniques like leverage. Of course using these techniques the risk is increasing dramatically and so a hedge funds can led to bankruptcy and that's why they need risk management. So because using these techniques, hedge funds can turn a minor loss to significant loss (and minor profit to huge obviously), they require from the investors to lock up their money. For this reason and various others the investors of hedge funds needs to be "qualified" investors and that mean to have a specific income in order to handle the potential risks according to Investopedia. Also hedge funds can invest in anything from stock market (short positions and long positions) to land, real estate and they don't have limitations such as the mutual funds.

The explanatory variables of the hedge funds that used are determined along the lines and discussed in Agarwal and Naik (2004) and used by Vrontos, Vrontos and Giamouridis (2007) for estimating hedge fund pricing and model uncertainty using several different criteria in their analysis such as AIC, BIC and Bayesian Model Averaging.

https://doi.org/10.26219/heal.aueb.6490

The following matrix contains the indices that we use for the analysis:

Table 1: Hedge funds explanatory variables

RUS	Russel 3000 equity index excess return
RUS(-1)	Russel 3000 equity index excess return lagged
MXUS	Morgan Stanley Capital International world excluding USA index excess return
MEM	Morgan Stanley Capital International emerging markets index excess return
SMB	Fama and French (1993) 'size'
HML	Book-to-Market
MOM	Momentum factors by Carhart (1997)
SBGC	the Salomon Brothers world government and corporate bond index excess return
SBWG	the Salomon Brothers world government bond index excess return
LHY	Lehman high yield index excess return
DEFSPR	the difference between the yield on the BAA-rated corporate bonds and the 10-year Treasury bonds
GSCI	the Goldman Sachs commodity index excess returns
FRBI	Federal Reserve Bank competitiveness weighted dollar-index excess return
VIX	the change in S&P 500 implied volatility index

All the above factors are used to predict the Credit Suisse/Tremont Hedge Fund Index (CSTHFI) which is monthly data and is appropriate representative of the entire hedge fund industry. Our main goal is to use the mean, linear regression, regression tree and random forest for predicting one step ahead the returns of CSTHFI and describe the volatility using linear and nonlinear GARCH models. Also another important aspect of the analysis is to use the above models to evaluate the important variables.

https://doi.org/10.26219/heal.aueb.6490

1.4 Data Analysis

Analysing the structure of CSTHFI we calculate the ACF for the returns, absolute returns and squared returns, Q-Q plot, Ljung-Box Statistic, descriptive statistics, lagged value plot, Augmented Ducky Fuller and Jarque-Bera test. The following diagrams are the ACF of CSTHFI. As we can observe the returns don't have a structure but the squared and absolute values have something that point us one of the main characteristics of financial data, volatility clustering. Volatility Clustering can be observed by the fact that autocorrelation of squared and absolute returns are statistical significant but also by Ljung-Box statistic which in our case is statistical insignificant for the returns and statistical significant for the absolute and squared returns). Using the ACF plots also we can observe that we don't have an explosive time series because we would expect all the sticks to be above the line and don't decline at all.

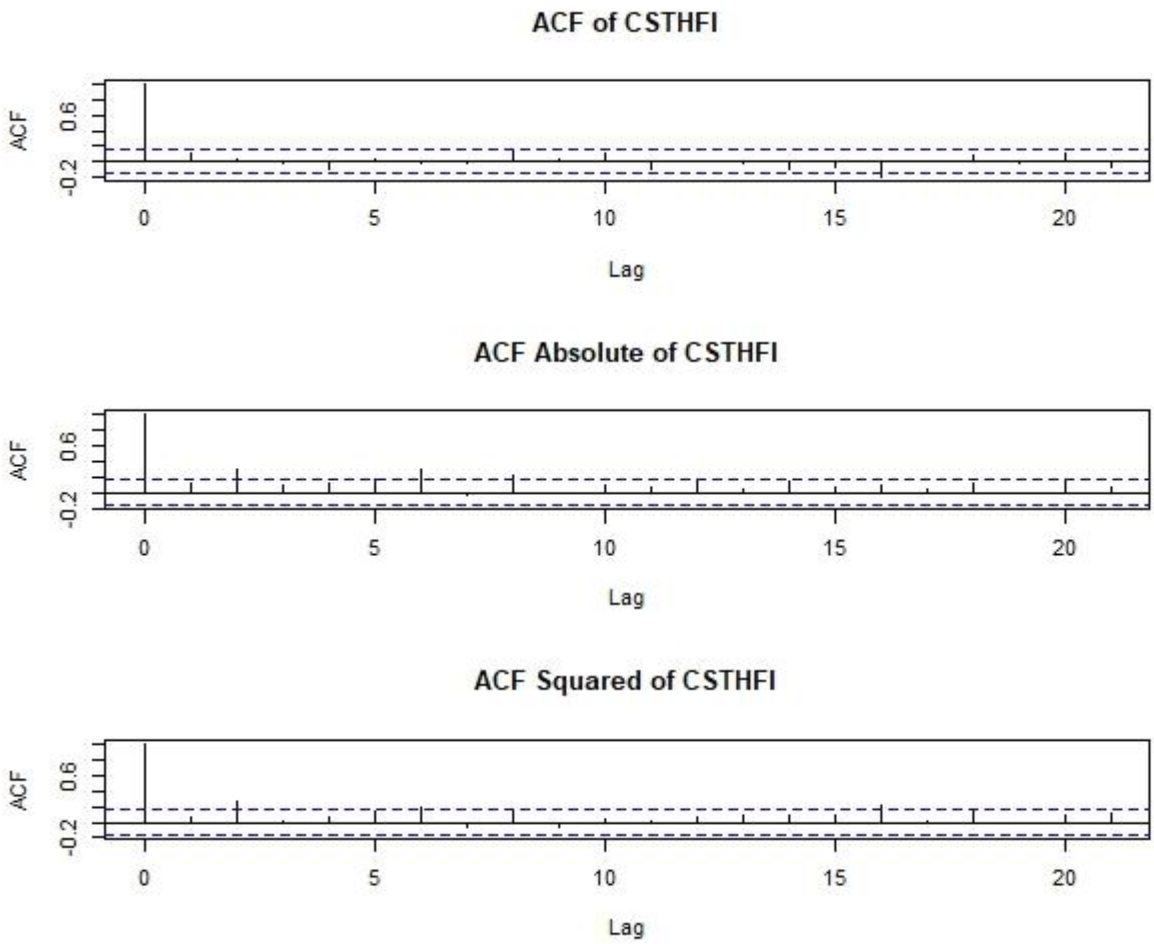

Figure 1 ACF of returns of CSTHFI, ACF of absolute returns of CSTHFI, ACF of squared returns of CSTHFI

https://doi.org/10.26219/heal.aueb.6490

Also another interesting fact is that large return values tend to occur in clusters and we can observe that from the scatterplot of the returns and the lagged returns. As we see from the figure 2 when the returns leave the main cloud of values they travel for an extended period of time. Specifically we see the returns travel outside the cloud in observations seventy-one until seventy-seven which takes from high positive returns to high negative returns. From the other hand we see the observations like fifty five to fifty seven which are large negative returns. This is a classic behaviour of financial returns where large negative returns proceed with large negative returns. One explanation for that is because investors afraid the turbulence of the market and so they wait until the market is more stable.

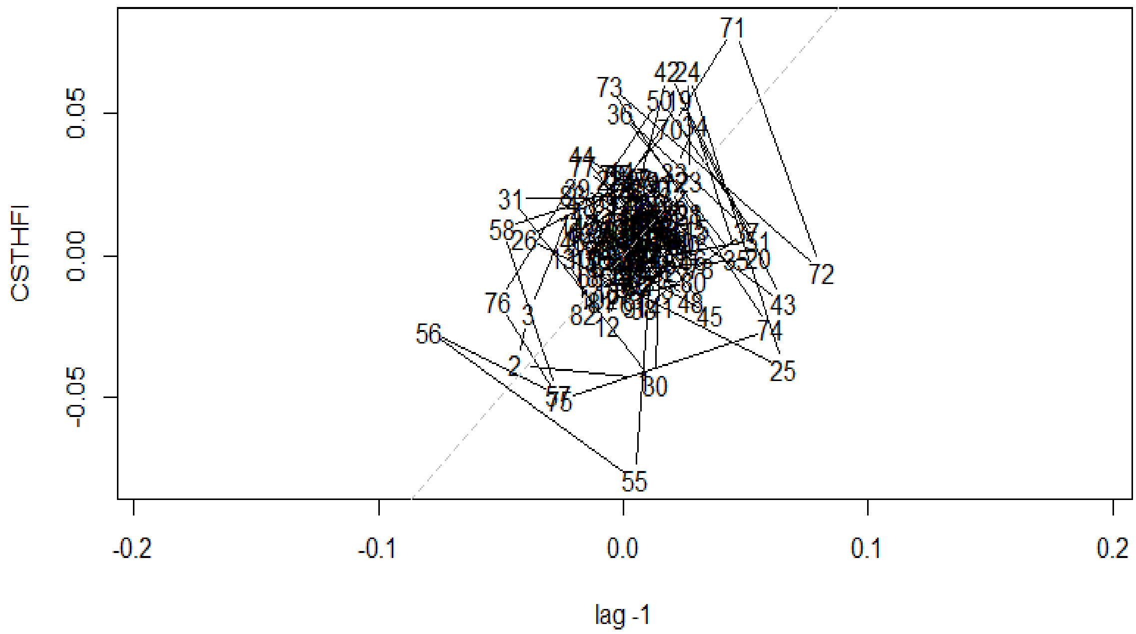

Figure 2 Scatterplot of CSTHFI value and CSTHFI lagged value

Table 2: Descriptive Statistics of CSTHFI and explanatory variables

	Mean	sd	Median	Min	Max	Range	Skew	Excess Kurtosis
CSTHFI	0.01	0.02	0.01	-0.08	0.08	0.16	0	2.11

https://doi.org/10.26219/heal.aueb.6490

RUS	0.00	0.04	0.01	-0.17	0.08	0.25	-0.84	1.22
RUS(-1)	0.00	0.04	0.01	-0.17	0.08	0.25	-0.84	1.23
MXUS	0.00	0.04	0.00	-0.14	0.09	0.24	-0.58	0.50
MEM	0.00	0.07	0.00	-0.35	0.12	0.47	-1.19	3.67
SMB	0.00	0.04	0.00	-0.12	0.15	0.26	0.40	1.47
HML	0.00	0.04	0.00	-.0.21	0.15	0.36	-0.77	5.73
MOM	0.01	0.07	0.00	-0.46	0.14	0.60	-2.52	15.50
SBGC	0.00	0.01	0.01	-0.04	0.04	0.08	-0.41	0.74
SBWG	0.00	0.02	0.00	-0.04	0.05	0.10	0.35	0.23
LHY	0.00	0.03	0.00	-0.12	0.10	0.23	-0.33	5.02
DEFSPR	0.00	0.00	0.00	0.00	0.00	0.01	0.91	1.95
GSCI	0.00	0.01	0.00	-0.03	0.03	0.07	0.04	1.01
FRBI	0.00	0.06	0.01	-0.16	0.15	0.31	-0.05	-0.03
VIX	0.00	0.04	0.00	-0.13	0.19	0.32	0.78	4.89

From the above table we see that excess kurtosis has a value of 2.11 and for a distribution to be normal we expect zero (which normal distribution has). So as we expected the distribution of the returns has fat tails and a very interesting statistics is the skewness, which is zero meaning that there is no asymmetry in the distribution of the CSTHFI. To test the assumption of normality in kurtosis and skewness the test we use is the Jarque-Bera which gives statistical significant result. Also observing the other factors the minimum values are larger than the positive ones and excess kurtosis has a very large values, larger than normal for example the MOM has the largest excess kurtosis in the dataset. We see a slightly right asymmetry in the most factors as a property of financial data and all the variables has mean equal approximately to zero.

Also using a correlation plot we can observe the correlation of CSTHFI with different factors to identify important relations. An interesting fact is that we don’t have factors to be highly correlated with CSTHFI and the highest is the Morgan Stanley Capital International emerging markets index excess return (MXUS) and Russel 3000 equity index excess return (RUSS) with correlation 0.54 and 0.45. Russel 3000 is also highly correlated with Volatility Index (VIX) and from this we can possible conclude that Russel 3000 as an index has a strong connection

 https://doi.org/10.26219/heal.aueb.6490

with the stock market because the VIX express the volatility of S&P500, the market portfolio. From the below figure we can also conclude that there is no reason for collinearity in linear regression because the correlation among the factors is not so high.

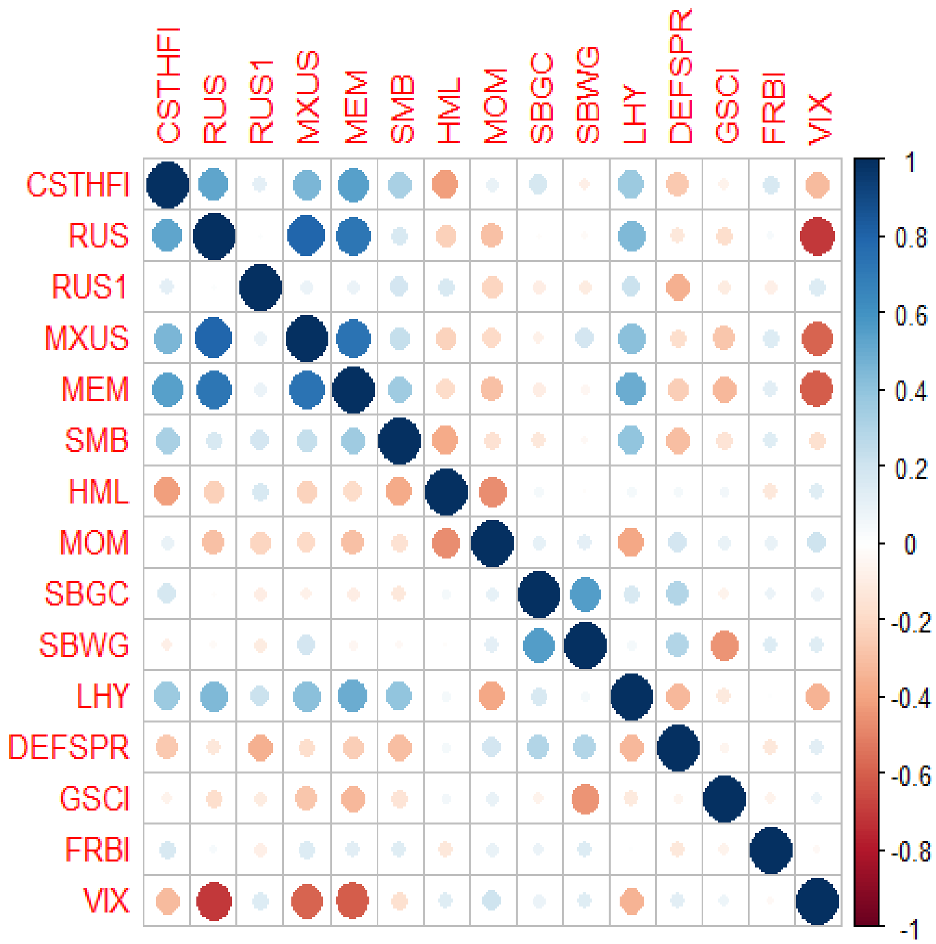

Figure 3 the correlation plot between CSTHFI and the explanatory variables

Also observing the Q-Q plot we can see that the returns aren't distributed normally, as we expected for a financial time series. We see deviations from the normality especially in the tails of the distributions and so the assumption of normality is violated. Also we can spot the distribution has fat tails something that we expected to have.

https://doi.org/10.26219/heal.aueb.6490

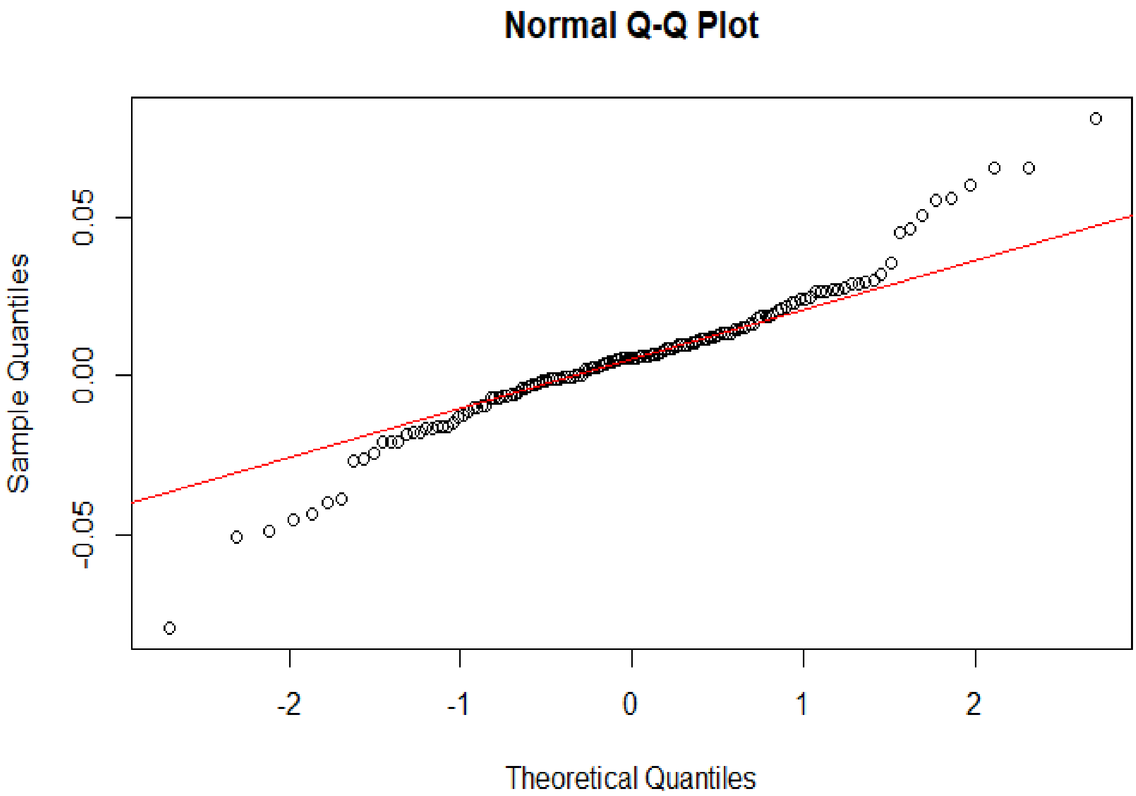

Figure 4 Q-Q plot of CSTHFI

Finally we have the time series and the distribution of CSTHFI. The distribution of CSTHFI confirms us that it's leptokurtic and fat tailed. Also we can spot that it is symmetric and that the maximum and minimum returns are almost equal. From the series we can observed the difference in volatility and that the variability of the distribution changes from the eighty eight month and after. Also we can see outliers which are outliers created by something outside the data generating process for example bad or good news that can cause the returns to take large positive and/or negative values. Also we can observe that we don't have level shifts in the distribution of the returns. After that we can see that the volatility is reduced significantly compare to the previous year's meaning maybe that the sector is more stable.

https://doi.org/10.26219/heal.aueb.6490

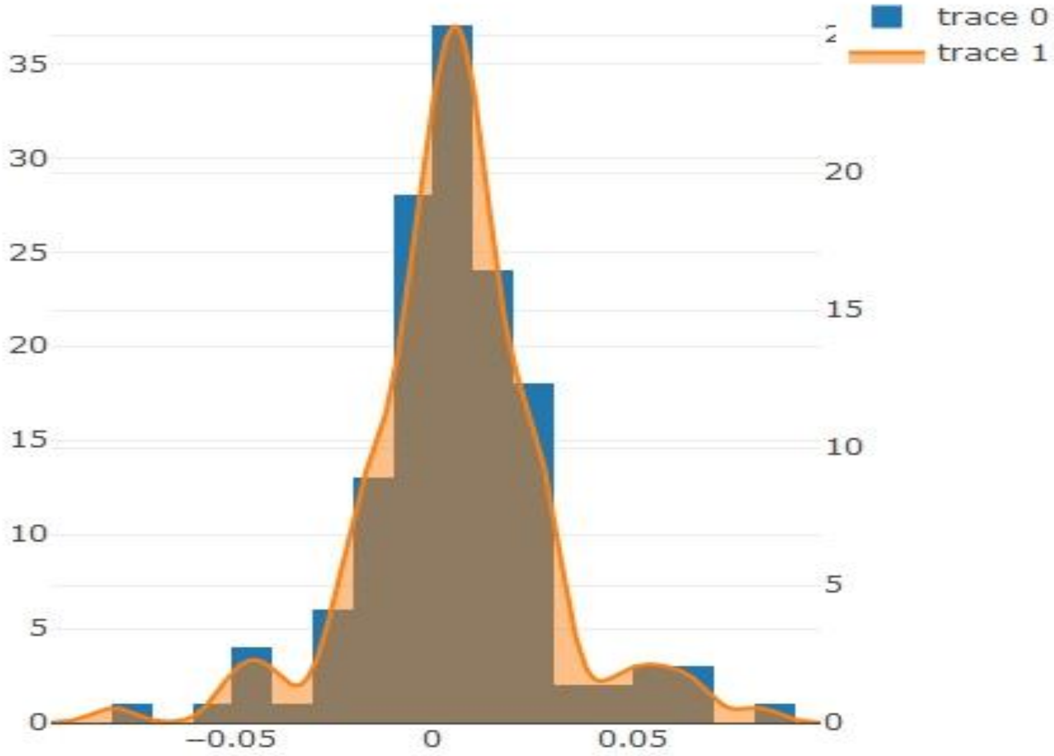

Figure 5 Distribution of CSTHFI

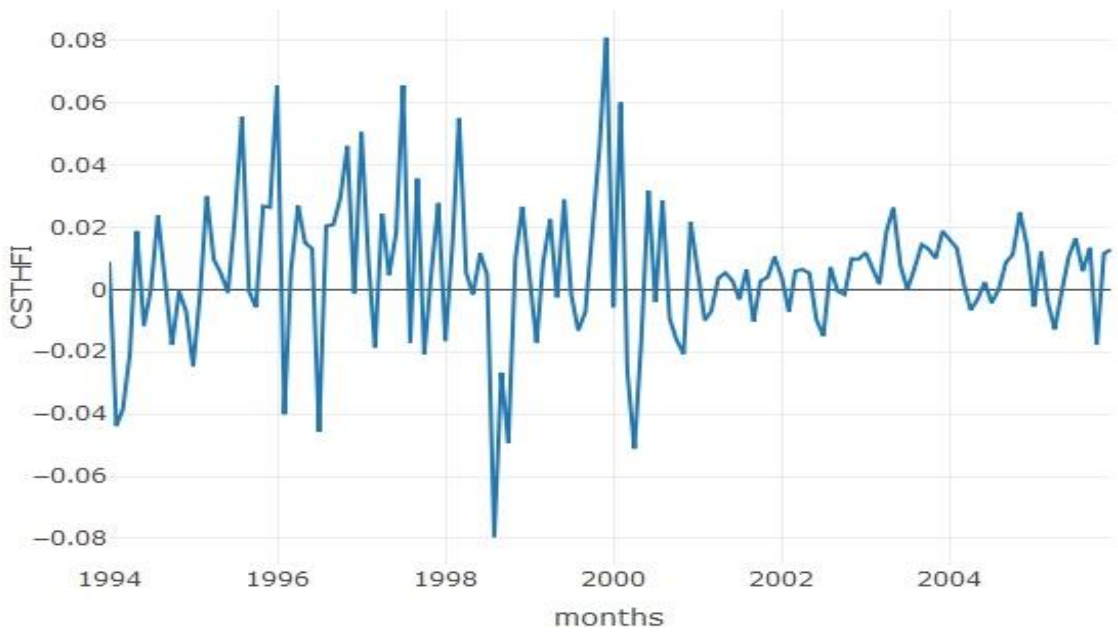

Figure 6 Time series plot of monthly returns of CSTHFI

2. Models

2.1 Multiple Linear Regression

Multiple Linear Regression is one of the most basic models in statistic and it has many applications. One of them is in Finance for predicting or explaining the relationship of financial assets. The model has the following form

$$Y = b_0 + b_1X_1 + b_2X_2 + b_3X_3 \dots . +b_nX_n + e$$

$$e \sim N(0, \sigma^2)$$

Or can be described like this

$$Y \sim N(b_0 + b_1X_1 + b_2X_2 + b_3X_3 \dots . +b_nX_n, \sigma^2)$$

Where Y is the dependent variable and particularly in finance Y is usually the returns, X_i $i = 1,2 \dots n$ is the independent variables and in finance are financial assets b_i $i = 1,2, \dots, n$ are the regression coefficients and e is an identical and independent random variable (iid) which called shocks (residuals) and follows normal distribution. The term linear is used because the above equation is linear in terms of the coefficients b_i $i = 1,2, \dots, n$ and has nothing to do with the order of the independent variables.

In order to find the regression coefficients $\hat{b}$ that minimizes the sum of squared residuals we end up with the least-squares normal equations which is

$$X'X\hat{b} = X'y$$

And solving for $\hat{b}$ we have the least squares estimators of $\hat{b}$

$$\hat{b} = (X'X)^{-1}X'y$$

Provided that the inverse matrix $(X'X)^{-1}$ exists. After finding the $\hat{b}$ and evaluating the model we can find the shocks which is the difference between the observed values and the fitted values

$$e = y - \hat{y}$$

 https://doi.org/10.26219/heal.aueb.6490

We can observe from the model that variance doesn't change over time something that in financial assets is violated (time varying volatility) and also that shocks are independent with each other something that usually doesn't happen in finance (volatility clustering). So to deal with this problem we model the residuals (or we can say the variance) with models like GARCH and others that we explain them latter.

Some of the factors aren't important or can't describe or explain the dependent variable. So we usually choose a criterion in order to decide which variable to include in the model and which to exclude. There are many different criteria. The one we use is Akaike Information Criterion (AIC) created by Akaike (1973) because it tries to pick the best model for forecasting purposes. AIC uses penalized log likelihood and has the following form

$$AIC = -2\ln(L) + 2p$$

Where p is the number of parameters in the model and L is the Log Likelihood. The procedure that we use for estimating the regression model is stepwise. More extensive analysis about multiple linear regression models can be found for example in Montgomery (2012).

2.2 Regression Trees

Tree based methods partition the feature space into a set of rectangles and then fit a simple model for example a constant in each one. The most popular algorithm for classification and regression is the CART (Classification and Regression Trees) algorithm developed by Breiman et al. (1984) and for implementation we use the Rpart package. Regression Trees can described like nested if else statements as Kuhn et al. (2013) described it. For example

1. If predictor A >= 1.7 then
    a. if predictor B>= 202.1 then outcome = 1.3
    b. else Outcome = 2.5
2. else Outcome = 5.6

They divide the feature space into two splits and within this splits, tree predict the outcome as the example above which is numerical and constant

 https://doi.org/10.26219/heal.aueb.6490

$$f(x) = \sum_{m=1}^{M} C_m I(x \in R_m)$$

Where m is the number of partitioning. In the above example the feature space divided in three and in each partition the constant assigned is the prediction for this regime.

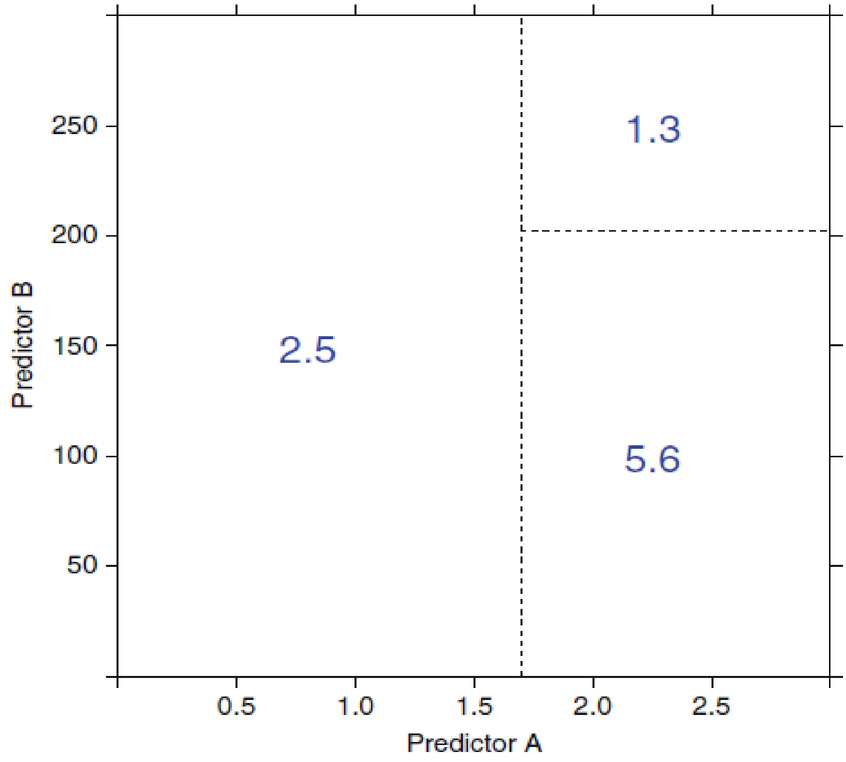

Figure 7 Example of the paradigm for the subspace created by a regression tree (picture from the book applied predictive modelling)

For prediction we can use many different criteria. One of the basic criteria is to minimize the sum of squared error and so it is easy to observe that the best C_m is just the average of $y_i$in region R_m:

$$\hat{C}_m = ave(y_i | x_i \in R_i)$$

In other models C_m can be the median or even a linear model. The most important thing for the algorithm is the splitting variable and the splitting point. It is computationally infeasibly to find the best binary partition using the minimum sum of squares so the algorithm proceed very

 https://doi.org/10.26219/heal.aueb.6490

greedily. What the algorithm does is starting with all the data uses a splitting variable j and split point s and define the pair of half planes as Hastie et al. (2009) described it

$$R_1(j,s) = \{X|X_j \leq s\} \text{ and } R_2(j,s) = \{X|X_j > s\}$$

Then to find j and s we minimize the following function:

$$\min_{j,s}[\min_{c_1} \sum_{x_i \in R_1(j,s)} (y_i - c_1)^2 + \min_{c_2} \sum_{x_i \in R_2(j,s)} (y_i - c_2)^2]$$

For any choice j and s, the inner minimization is solved by

$$\widehat{c_1} = ave(y_i|x_i \in R_1(j,s)) \text{ and } \widehat{c_2} = ave(y_i|x_i \in R_2(j,s))$$

After the algorithm continuous until a stopping criterion is met. Here the problem is when to stop because a large tree is probe to overfitting which means having a very good in sample predictions but not out of sample. To deal with that we can tune the parameters of the tree and found out which tree is the optimal tree. There is three parameters for tuning.

- Minsplit is the minimum number of observations that required to have a regime before the algorithm attempt to split. The default in rpart is twenty. Controlling that parameter we force the algorithm to have reasonable number of observations before a split.
- Maxdepth is the maximum number of internal nodes between the root node and the terminal node. Practically is number of nodes a tree could have before ends at the terminal nodes. The default is thirty and so a very large tree can be constructed leading to overfitting.
- Complexity parameter (Cp) is a parameter that penalize trees objective function in order to control complexity and depth of the tree for better out of sample prediction. The form of the objective function with the complexity parameter is the following:

$$SSE_{c_p} = SSE + c_p|T|$$

Where T is the number of terminal nodes. When the value of c_p= 0 then we have the original tree and as c_p becomes bigger then it is more expensive to create a large tree.

 https://doi.org/10.26219/heal.aueb.6490

Regression trees has many advantage but many disadvantages. Regression trees are considered to be easy model in the sense that they don't need to know the relationship of the variables or scaling the parameters. Also they can handle the missing values and make predictor selection very easily something that we want in real life problems. Of course there is no free lunch. Regression trees do not have good predictive performance. Also if the data changed then all the structure of the tree changes and it loses all the information and the interpretation. For the prediction there are many techniques that give regression trees predictive power for example bagging and boosting but loses interpretation. More extensive analysis about regression tree algorithms can be found for example in Hastie (2009) and Kuhn (2013).

2.3 Random Forest

One of the most powerful technique is bagging or bootstrap aggregation because it reduce the variance of an estimation prediction function. It works especially well when the base learner has high variance low bias something that trees do have. So random forests is a substantial modification of bagging that creates a large collection of de-correlated trees and then average them but in order to work it needs the model to be approximately unbiased and have a high variance or else it doesn't work. The logic behind it is because creating many different unbiased trees and averaging them, the bias will still be the same for the bagged trees but the variance will be decreased. If the trees are independent and identical the variance of the average will be $\frac{\sigma^2}{B}$ but if the trees are simply identically distributed (but not independent) with positive pairwise correlation ρ (positive because the trees are created by the same variables) the variance of the average is

$$\rho\sigma^2 + \frac{1-\rho}{B}\sigma^2$$

From the above we can observe that as B increased the second term goes to zero and so the size of correlation (ρ) of pairs of bagged trees limits the benefit of the averaging. The algorithm deal with this problem by choosing random m variables (m<p) from the input data and by that reducing the correlation of the bagged trees and so the variance. The random forest regression prediction of a new observation is the average of all the trees:

https://doi.org/10.26219/heal.aueb.6490

$$\widehat{f_{rf}^{B}}(x) = \frac{1}{B}\sum_{b=1}^{B} T(x;\Theta_b)$$

Where Θ_b characterizes the bth random forest tree in terms of split variables, cut points at each node and terminal-node values. The algorithm of random forest as described it by Kuhn (2013) is the following:

1. Select the number of models to Built , B
2. For i = 1 to B do
    a. Generate a bootstrap sample of the original data
    b. Train a tree model on this sample
    c. For each split do
        i. Randomly select k(<P) of the original predictors
        ii. Select the best predictor among the k predictors and partition the data
    d. End
    e. Use typical tree model stopping criteria to determine when a tree is complete (but do not prune)
3. end

As we can see Random Forest has two tuning parameters, the randomly selected predictors and the number of trees. For regression usually preferred to be the one-third of the original parameters but this is not always the case and the best value for this parameter will depend from the data. Because the random forest can't over fit we can select as much trees as we want but the problem arises with the computational cost. Although it is not so important, if you choose a reasonably number of trees. An interesting fact is that we do not prune the tree but we can use stopping criteria such as the number of observation in terminal node. In this thesis for the tuning of random forest the R package that used was the CARET and specifically the train function. The results was that indeed the best model use the one-third of the data and the number of trees was one thousand and fifty hundred. Random forest was retrained after each new observation but the hyper parameters was kept steady in the whole forecasting procedure. More extensive analysis about random forest algorithm can be found for example in Hastie (2009) and Kuhn (2013).

 https://doi.org/10.26219/heal.aueb.6490

2.4 GARCH

GARCH (Generalized Autoregressive Conditional Heteroscedasticity) model is the most empirically used model in Finance created by Bollorslev (1986). It is a generalization of ARCH model and the reason of that is because it can be more parsimonious and more easily applicable. The model has the following form (the model can have explanatory variable in the mean equation) as described it by Vrontos (2015):

$$Y_t = b_0 + b_1X_{1t} + b_2X_{2t} + b_3X_{3t} \dots . + b_nX_{nt} + e_t$$

$$e_t|\Phi_{t-1} \sim N(0, \sigma_t^2)$$

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^{p} \alpha_i e_{t-i}^2 + \sum_{j=1}^{q} b_j \sigma_{t-j}^2$$

Otherwise

$$Y_t|\Phi_{t-1} \sim N(b_0 + b_1X_{1t} + b_2X_{2t} + b_3X_{3t} \dots . + b_nX_{nt}, \sigma_t^2)$$

As we can see GARCH is a linear combination of previous squared error terms and variances. The advantages of the model except the parsimonious form is that can capture the volatility clustering which is that variance can divided in periods of high volatility or low volatility. This means that if the previous time you have high volatility then now you will have high volatility and vice versa for example Greece has financial crisis and we expect high volatility in this period of time instead of a country which has more stable economy like Germany. Empirically the order of the model is usually GARCH(1,1). Of course a disadvantages of GARCH is that it can't capture the leverage effect which is that volatility is different if you have negative or positive shocks and for that reason threshold models like Switching Volatility GARCH (SV-GARCH) and GARCH-GJR developed.

https://doi.org/10.26219/heal.aueb.6490

2.5 GARCH-GJR

GJR – GARCH introduced by Glosten, Jagannathan and Runkle (1993) and it has the following form as described it by Franses (2000):

$$h_t = \omega + \alpha_1 \varepsilon_{t-1}^2 (1 - I[\varepsilon_{t-1} > 0]) + \gamma_1 \varepsilon_{t-1}^2 I[\varepsilon_{t-1} > 0] + \beta_1 h_{t-1}$$

Where $I\,[\cdot]$ is an indicator function. To ensure non-negativeness of the conditional variance the coefficients need to have the following restrictions: $\omega > 0$, $(\alpha_1 + \gamma_1)/2 \geq 0$ and $\beta_1 > 0$. In order to have covariance-stationarity the restriction is $(\alpha_1 + \gamma_1)/2 + \beta_1 < 1$. As we can see the model is a generalization of GARCH(1,1). The difference is that using the sign of the shock it can determine the coefficient of the shock and this model captures the leverage effect. Of course also it captures the volatility clustering that classical GARCH model do. One disadvantage of this model is the variance stay the same even if the coefficient of the shock changes. For this reason the Switching Volatility GARCH created.

An interesting point is the reason that the sign of the residual is used. If we look close why the sign is used is because it express the following

$$\begin{cases} e_t < 0 \\ e_t > 0 \end{cases} <=> \begin{cases} y - \hat{y} < 0 \\ y - \hat{y} > 0 \end{cases} <=> \begin{cases} y < \hat{y} \\ y > \hat{y} \end{cases}$$

So what this models is suggesting is that when the prediction value is bigger than the actual value then probably is because of bad news and the opposite when the news is good. So using this models we can create smaller confidence intervals depending which regime is less volatile and larger depending the regime with more volatility.

2.6 SV-GARCH

Switching Volatility GARCH created by Fornari and Mele (1996, 1997) in order to capture the asymmetric behaviour of the volatility. The model is an extension of GARCH-GJR. Rabemananjara and Zakoian (1993) discovered that the size and the sign of the shock are important. Something that is intuitively right for example if a stock has positive shock the variance should be different instead of a negative shock. So this model practically changes the

https://doi.org/10.26219/heal.aueb.6490

GARCH models if the error term e_{t-1} is positive or negative and allow the coefficient of each regime to be different. With this adjustment the model can capture not only the leverage effect but the size of variance at each regime. The Volatility Switching GARCH (VS-GARCH) model of order (1,1) has the following form as described it by Franses (2000):

$$h_t = (\omega + \alpha_1 \varepsilon_{t-1}^2 + \beta_1 h_{t-1})(1 - I[\varepsilon_{t-1} > 0]) + (\zeta + \gamma_1 \varepsilon_{t-1}^2 + \delta_1 h_{t-1})(I[\varepsilon_{t-1} > 0])$$

Also here there isn't known conditions for non-negativiness and covariance stationarity but knowing that variance can't be negative this implies the coefficient to be positive. As we can see the model can have different slopes and it can be discontinuous at $\varepsilon_t = 0$. Also an important note is that the coefficient can point out the difference of a negative and positive shock and the magnitude of volatility for each regime.

The forecasting procedure has two stages. The first stage is calculating the model for the conditional expectation and then iteratively we forecast the next month. After we re-train the model including the true value of the month we forecast and then we forecast the next month. Depend the model the hyperparameters are steady or changing at each iteration. Regression tree and Random forests hyperparameters weren't recalculate because of computational cost and we calculate them in sample. The volatility coefficient from each model was estimated in sample using QML with distribution Normal the optimization technique that used was the BFGS algorithm using Box constrains with optim but also Nelder-Mead procedure using Box constrains using the package dfoptim in R. The prediction period is thirteen months and the in sample is one hundred thirty one months. The reason we didn't use larger prediction period is because the models for volatility and the regression tree for tuning needed enough data to converge.

 https://doi.org/10.26219/heal.aueb.6490

3. Empirical analysis

As we already mention the purpose of this thesis is to explore threshold algorithms for the conditional mean and threshold models for the conditional variance. The first model to estimate for the mean is a simple mean model:

$$r_t = \mu_t + e_t$$

$$e_t | \Phi_{t-1} \sim N(0, \sigma_t^2)$$

$$\sigma_t^2 = \alpha_0 + \alpha_1 e_{t-1}^2 + b_1 \sigma_{t-1}^2$$

or

$$\sigma_t^2 = \omega + \alpha_1 \varepsilon_{t-1}^2 (1 - I[\varepsilon_{t-1} > 0]) + \gamma_1 \varepsilon_{t-1}^2 I[\varepsilon_{t-1} > 0] + \beta_1 \sigma_{t-1}^2$$

or

$$\sigma_t^2 = (\omega + \alpha_1 \varepsilon_{t-1}^2 + \beta_1 \sigma_{t-1}^2)(1 - I[\varepsilon_{t-1} > 0]) + (\zeta + \gamma_1 \varepsilon_{t-1}^2 + \delta_1 \sigma_{t-1}^2)(I[\varepsilon_{t-1} > 0])$$

Iteratively we predict the returns of the last year in our data set using the mean updated each time with the new observation. In order to see the predictability of the model we using out of sample statistics such as Root mean square error (RMSE), mean absolute deviations (MAE) and Median absolute deviations (MDAE).The result are presenting in the table

Table 3 In-sample and out of sample statistics of mean model

	In sample	Out of sample
RMSE	0.023	0.010
MAE	0.016	0.009
MDAE	0.012	0.007

Something worth mention is that the model gives better result out of sample period instead of the in sample period. The reason for that is because the in sample time period has larger

https://doi.org/10.26219/heal.aueb.6490

volatility instead of the out of sample period. Also because the out of sample period is a relative small consisting by only thirteen observations (one year and one month). In figure 8 and 9 we can observe the actual values and the predicted values of the mean model for the in sample and out of sample prediction. We see that there is high volatility in the training data and with time the volatility is reducing, making the statistics look better for the out of sample period.

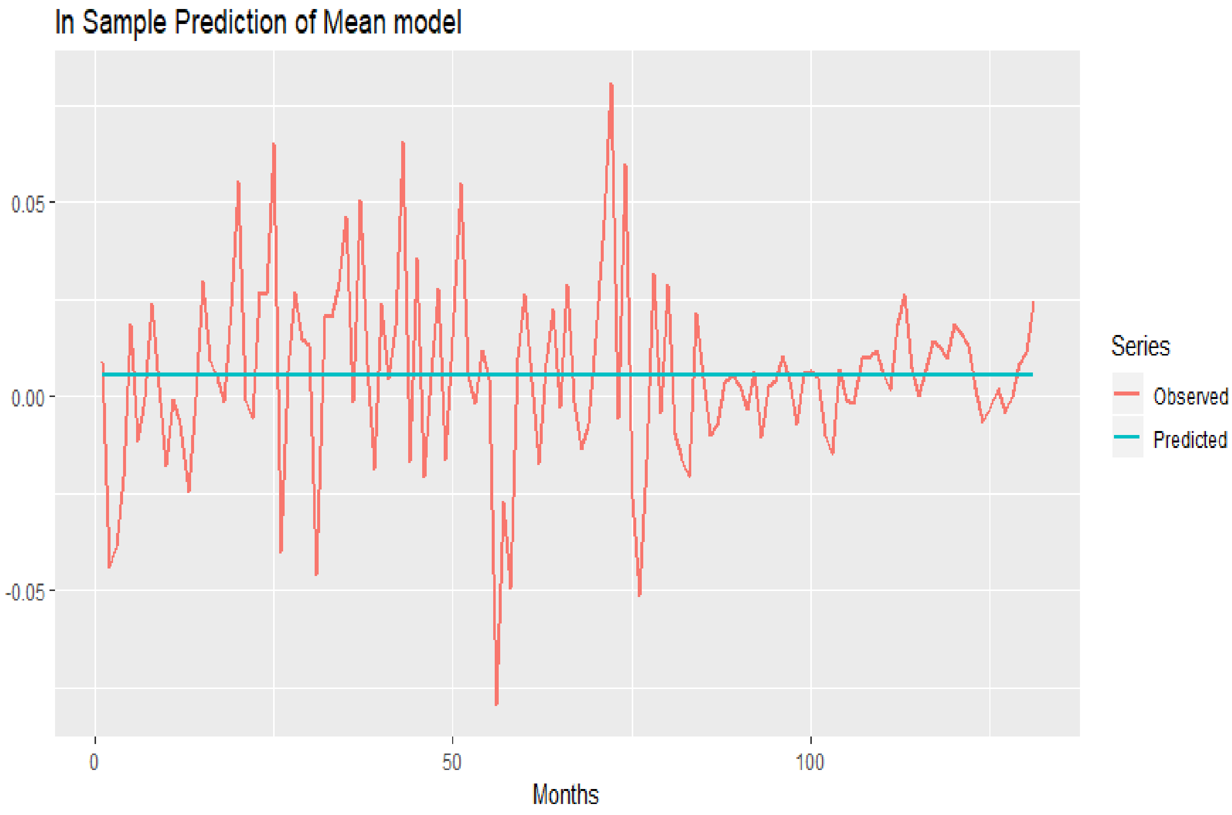

Figure 8 In sample prediction of the mean model

 https://doi.org/10.26219/heal.aueb.6490

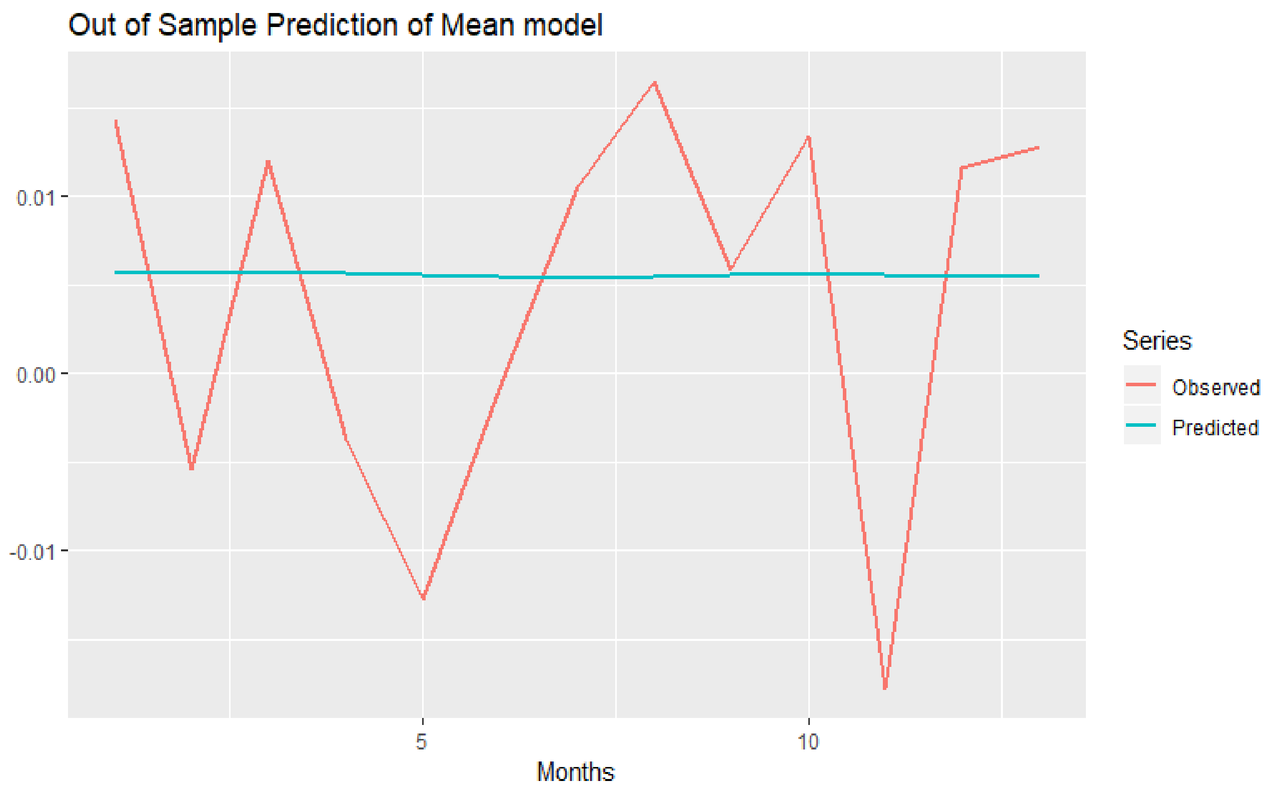

Figure 9 Out of sample prediction of the mean model

Analysing the squared and absolute residuals we found that the PACF and ACF are statistical significant. Also the assumption of normality is violated. Of course this can be caused because the mean model does not fit well the data.

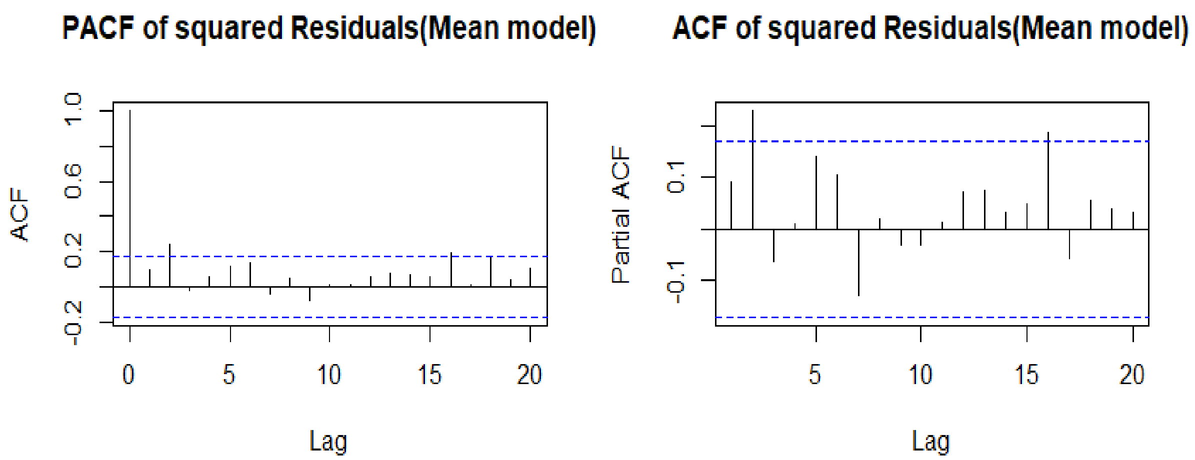

Figure 10 PACF and ACF of the squared residuals

So we model the residual variance using three different models. The first one is a linear GARCH model, the second is GARCH-GJR and third is SV-GARCH with the forms as we

 https://doi.org/10.26219/heal.aueb.6490

described in the previous chapter. Using QMLE estimation and optimization techniques BFGS and Nelder-Mead algorithm for the coefficients of the in sample period. The coefficients are the following:

GARCH: $\alpha_0 = 2 \times 10^{-6}$ $\alpha_1 = 0.12$ $b_1 = 0.87$

GARCH-GJR: $\omega = 1 \times 10^{-5}$ $\alpha_1 = 0.03$ $\gamma_1 = 0.35$ $\beta_1 = 0.80$

VS-GARCH: $\omega = 1 \times 10^{-9}$ $\zeta = 5 \times 10^{-5}$ $\alpha_1 = 0.14$ $\gamma_1 = 0.03$ $\beta_1 = 0.999$ $\delta_1 = 0.81$

Firstly we can observe that GARCH is covariance stationary because the parameters satisfy all the proper properties. We can observe that it has very high volatility coefficient something that we expect in financial data even if the data are monthly and not daily. Second and more interesting fact is in the GARCH-GJR were we can observe that the positive shocks have more impact than the negative ones. This might be because we have small dataset or because in the dataset more than 60% is positive values or another explanation might be because the hedge funds when positive news are in the market tend to take position for future profits and also possibly because the negative news does not affect so much the bad market news. We should not forget that hedge funds have unique and special characteristics, different from the rest of the market. Now looking the coefficients of SV-GARCH we can see that the negative news have large negative impact but the negatives regime volatilities coefficient is nearly one so the model can be trusted.

Next model is linear regression. The fitted models has the following form:

$$r_t = b_0 + b_1X_1 + b_2X_2 + b_3X_3 \dots . + b_nX_n + e_t$$

$$e_t|\Phi_{t-1} \sim N(0, \sigma_t^2)$$

$$\sigma_t^2 = \alpha_0 + \alpha_1 e_{t-1}^2 + b_1\sigma_{t-1}^2$$

or

$$\sigma_t^2 = \omega + \alpha_1\varepsilon_{t-1}^2(1 - I[\varepsilon_{t-1} > 0]) + \gamma_1\varepsilon_{t-1}^2 I[\varepsilon_{t-1} > 0] + \beta_1\sigma_{t-1}^2$$

or

 https://doi.org/10.26219/heal.aueb.6490

$$\sigma_t^2 = (\omega + \alpha_1 \varepsilon_{t-1}^2 + \beta_1 \sigma_{t-1}^2)(1 - I[\varepsilon_{t-1} > 0]) + (\zeta + \gamma_1 \varepsilon_{t-1}^2 + \delta_1 \sigma_{t-1}^2)(I[\varepsilon_{t-1} > 0])$$

We use the explanatory variables to make a better prediction of the CSTHFI and then we use linear and nonlinear GARCH for the residual variance if necessary. The variables that are important using the AIC criterion is the following:

Table 4 Variables that are important using AIC criterion

Variable	Coefficient
Intercept	0.002
RUS	0.22
MXUS	0.001
MEM	0.09
SMB	0.07
HML	-0.05
MOM	0.10
SBGC	0.51
SBWG	-0.26
LHY	0.06
DEFSPR	-1.99
VIX	0.05
LHY2	1.13
RUS2	2.20
RUS3	-2.93
MXUS2	1.25
DEFSPR2	-1361
RUS:VIX	3.65

The information variables for the linear model are the following: RUSSEL 3000 (RUS), Morgan Stanley Capital International emerging markets index excess return (MEM), Fama and French (1993) 'size' (SMB), Momentum factors by Carhart (1997) (MOM), Morgan Stanley Capital international world excluding USA index excess return(MXUS),Salomon Brothers world government bond index excess return (SBGC), Lehman high yield index excess

 https://doi.org/10.26219/heal.aueb.6490

return(LHY) ,the Salomon Brothers world government bond index excess return (SBWG), the difference between the yield on the BAA-rated corporate bonds and the 10-year Treasury bonds (DEFSPR), the change in S&P 500 implied volatility index (VIX),the interaction between RUS and VIX, the squared and cubic term of RUS, the squared term of LHY, the squared term of MXUS and , the squared term of DEFSPR. The statistics checking the in sample and out of sample prediction of the linear model presented in the next table

Table 5 In-sample and out of sample statistics of linear regression

	In sample	Out of sample
RMSE	0.013	0.008
MAE	0.010	0.006
MDAE	0.0089	0.004

As we can see the model gives better result for the out of sample period instead of the in sample period. Also the explanatory variables are indeed help us in the prediction of the CSTHFI. Below is the plots of the in sample and out of sample prediction. We can detect that in the period of large volatility the model can’t capture the dynamic of the index even using explanatory variables. This can be caused because there is a need for nonlinear relationship between the explanatory variables and the CSTHFI something that we will explore with the next models.

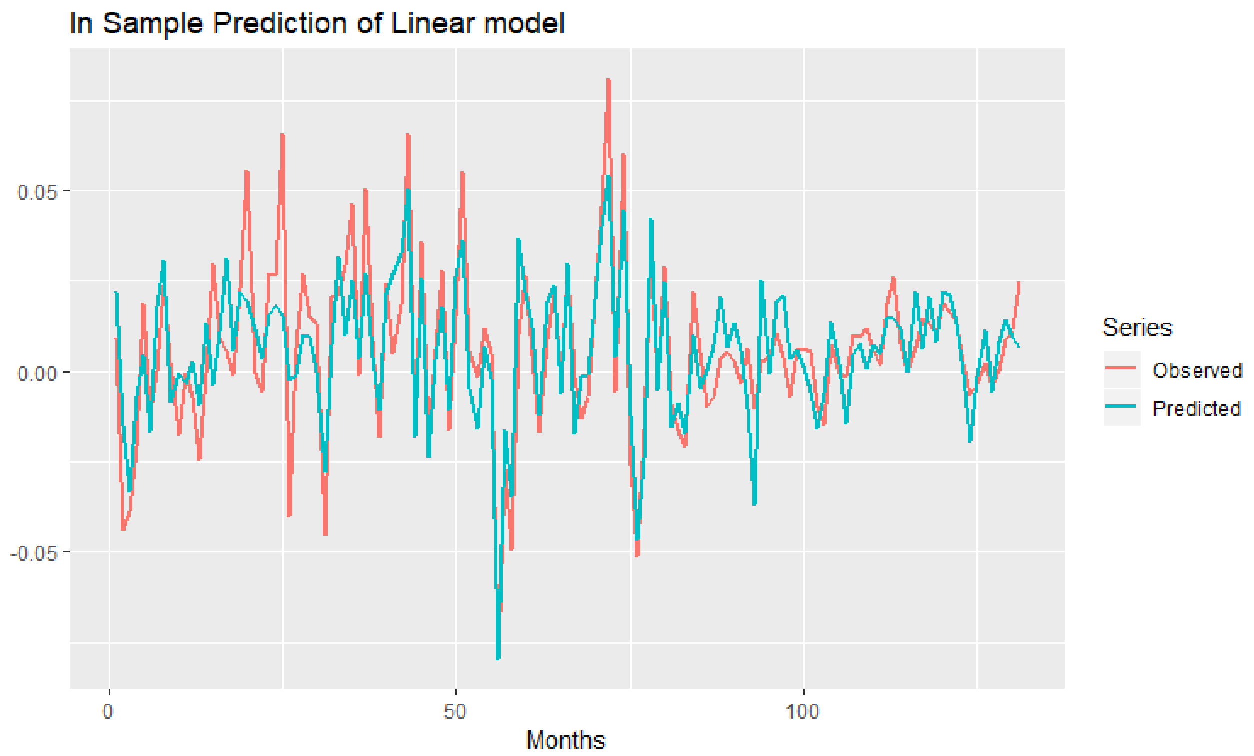

Figure 11 In-Sample prediction of linear regression

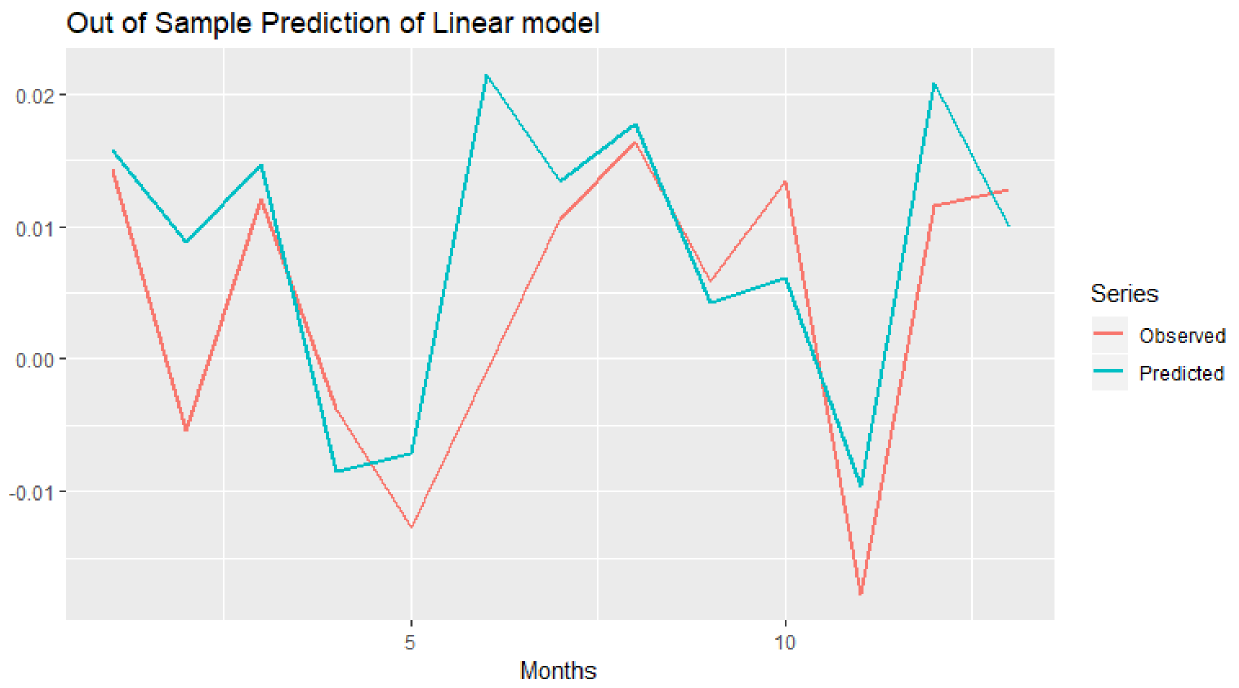

Figure 12 Out-of-Sample prediction of linear regression

https://doi.org/10.26219/heal.aueb.6490

The PACF and ACF of the absolute and squared in sample residuals are statistical significant so we use GARCH models in the residuals but the assumption of the normality is not violated as the Kolmogorov Smirnov test implies with p value nonsignificant.

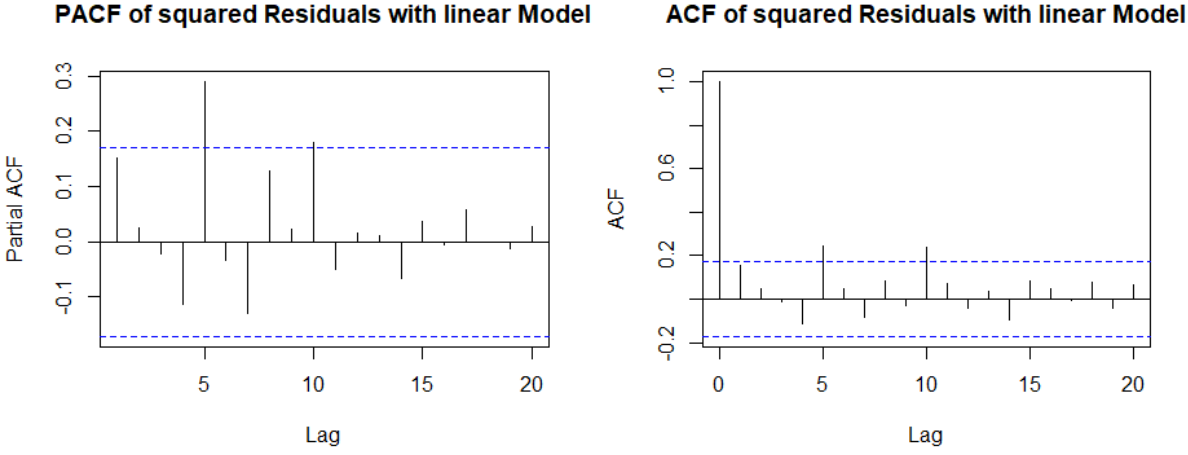

Figure 13 PACF and ACF of squared residuals from linear regression

The GARCH model coefficients are presented next and we can see that the three models are practically the same model. The coefficients are all the same with both BFGS and Nelder-Mead giving the same results. The coefficient of variance is the largest coefficient by far and we observe that all the other coefficient independent from the model are nearly zero. Also we can observe that the news have no impact in the variance and probably because the explanatory variables. The linear models with this explanatory variables explains better the variance of the index.

$$\text{GARCH: } \alpha_0 = 9 \times 10^{-8} \;\; \alpha_1 = 1 \times 10^{-8} \;\; b_1 = 0.97$$

$$\text{GARCH-GJR: } \omega = 1 \times 10^{-5} \;\; \alpha_1 = 1 \times 10^{-8} \;\; \gamma_1 = 1 \times 10^{-8} \;\; \beta_1 = 0.97$$

$$\text{VS-GARCH: } \omega = 1 \times 10^{-9} \;\; \zeta = 5 \times 10^{-5} \;\; \alpha_1 = 1 \times 10^{-4} \;\; \gamma_1 = 1 \times 10^{-4} \;\; \beta_1 = 0.93$$
$$\delta_1 = 0.999$$

The third model is the regression tree model. The model is the following

$$r_t = regression\ tree + e_t$$

$$e_t|\Phi_{t-1} \sim N(0, \sigma_t^2)$$

https://doi.org/10.26219/heal.aueb.6490

$$\sigma_t^2 = \alpha_0 + \alpha_1 e_{t-1}^2 + b_1 \sigma_{t-1}^2$$

or

$$\sigma_t^2 = \omega + \alpha_1 \varepsilon_{t-1}^2 (1 - I[\varepsilon_{t-1} > 0]) + \gamma_1 \varepsilon_{t-1}^2 I[\varepsilon_{t-1} > 0] + \beta_1 \sigma_{t-1}^2$$

or

$$\sigma_t^2 = (\omega + \alpha_1 \varepsilon_{t-1}^2 + \beta_1 \sigma_{t-1}^2)(1 - I[\varepsilon_{t-1} > 0]) + (\zeta + \gamma_1 \varepsilon_{t-1}^2 + \delta_1 \sigma_{t-1}^2)(I[\varepsilon_{t-1} > 0])$$

As we already mention there is a need for a model that can capture the nonlinear relationships. In the next figure we can see the in sample period tree that calculated and tuned in order to have the best possible prediction, in sample and out of sample. The plot and the algorithm implemented in R using the library Rpart. The important variables are recognized by the algorithm using the Residual sum of squared as we already mentioned.

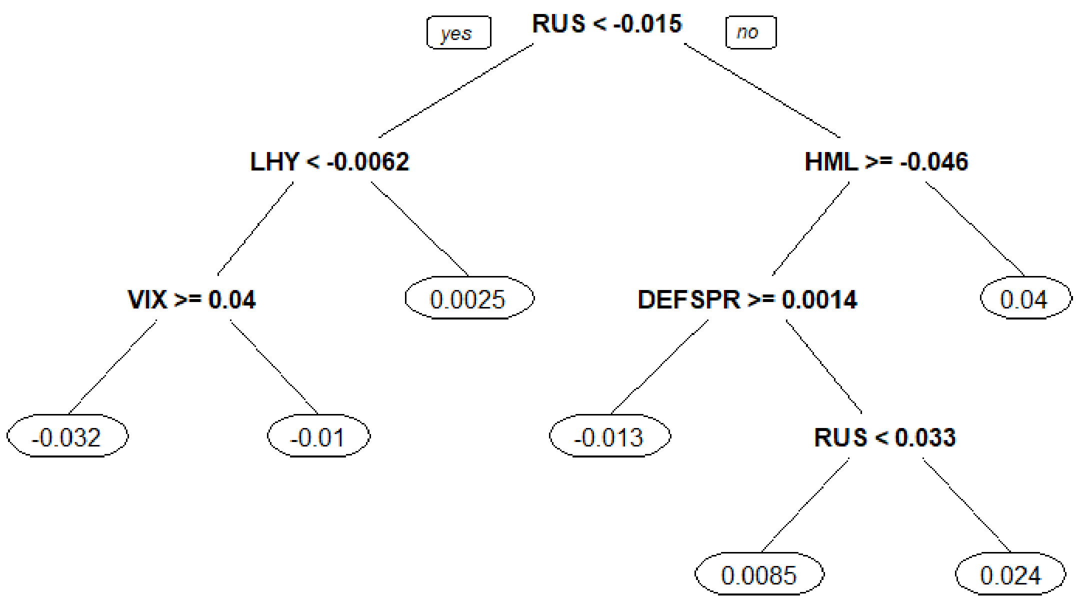

Figure 14 Regression Tree of the Hedge fund data

As we can see the variables that used after pruning presented is six. Specifically the following: RUSSEL 3000 (RUS), Morgan Stanley Capital International emerging markets index excess return (MEM), Fama and French (1993) 'size' (SMB), the Salomon Brothers world government bond index excess return (SBWG), Lehman high yield index excess return (LHY), Book-to-Market (HML). An interesting fact is the difference between the regression tree and

https://doi.org/10.26219/heal.aueb.6490

the linear model in the important variables. Indices that consider to be important in the linear regression such as DEFSPR and SMB do not consider important for the prediction in the regression tree. In the next table the statistics presented for the predictive power of the regression trees in the hedge funds and we can notice that the model is not better than the linear model both for the in sample and out of sample period.

Table 6 In-Sample and Out-of-Sample Statistics of the regression Tree

	In sample	Out of sample
RMSE	0.0159	0.0095
MAE	0.0118	0.0064
MDAE	0.008	0.0042

Figure 15 In-sample prediction of Regression Tree

https://doi.org/10.26219/heal.aueb.6490

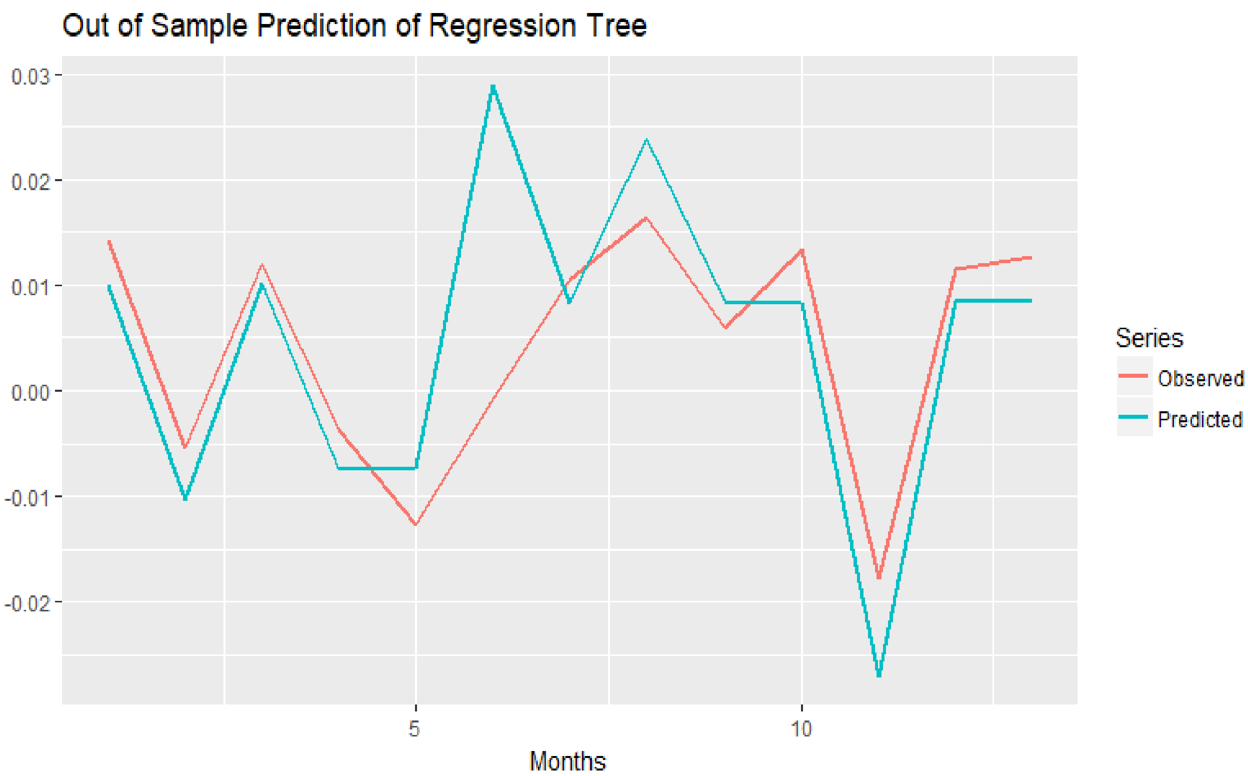

Figure 16 out-of- sample prediction of Regression Tree

From the above plots we can conclude that the regression trees cannot predict extreme values that appear in the dataset. Looking the residuals of the model, the normality assumption using the Kolmogorov Smirnov test isn't clear. The p-value is not statistical significant but only for 0.056 which means that is ambiguous. Also the PACF and ACF of the squared and absolute residuals are not statistical significant and also the residuals PACF and ACF are non-statistical significant as we can spot from the above plot. So there is no autocorrelation so there is no need for linear and nonlinear GARCH.

https://doi.org/10.26219/heal.aueb.6490

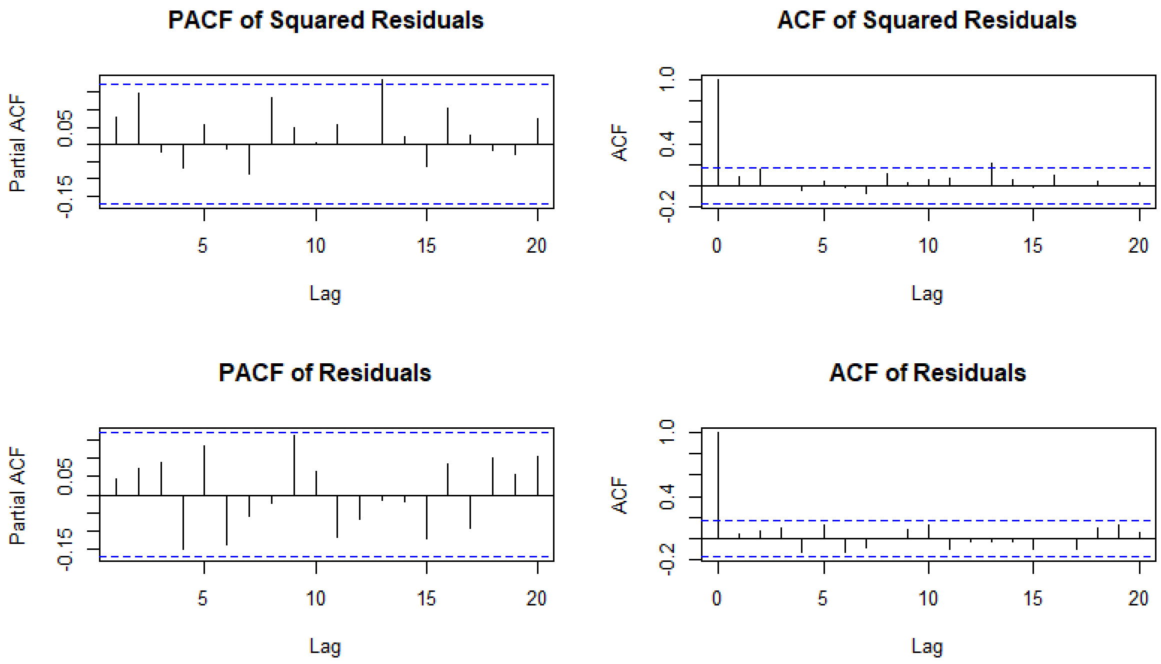

Figure 17 PACF and ACF of residuals from regression tree

The fourth and last model is the random forest model for estimating the returns and for the volatility we using linear and nonlinear GARCH models if necessary. The form of the models is the following:

$$r_t = Random\ Forest + e_t$$

$$e_t|\Phi_{t-1} \sim N(0, \sigma_t^2)$$

$$\sigma_t^2 = \alpha_0 + \alpha_1 e_{t-1}^2 + b_1 \sigma_{t-1}^2$$

or

$$h_t = \omega + \alpha_1 \varepsilon_{t-1}^2 (1 - I[\varepsilon_{t-1} > 0]) + \gamma_1 \varepsilon_{t-1}^2 I[\varepsilon_{t-1} > 0] + \beta_1 h_{t-1}$$

or

$$h_t = (\omega + \alpha_1 \varepsilon_{t-1}^2 + \beta_1 h_{t-1})(1 - I[\varepsilon_{t-1} > 0]) + (\zeta + \gamma_1 \varepsilon_{t-1}^2 + \delta_1 h_{t-1})(I[\varepsilon_{t-1} > 0])$$

First of all we can use the random Forest for variable importance in order to see the variables that are most important and to compare them with the previous models. Using the caret package in R we can see the important variables.

https://doi.org/10.26219/heal.aueb.6490

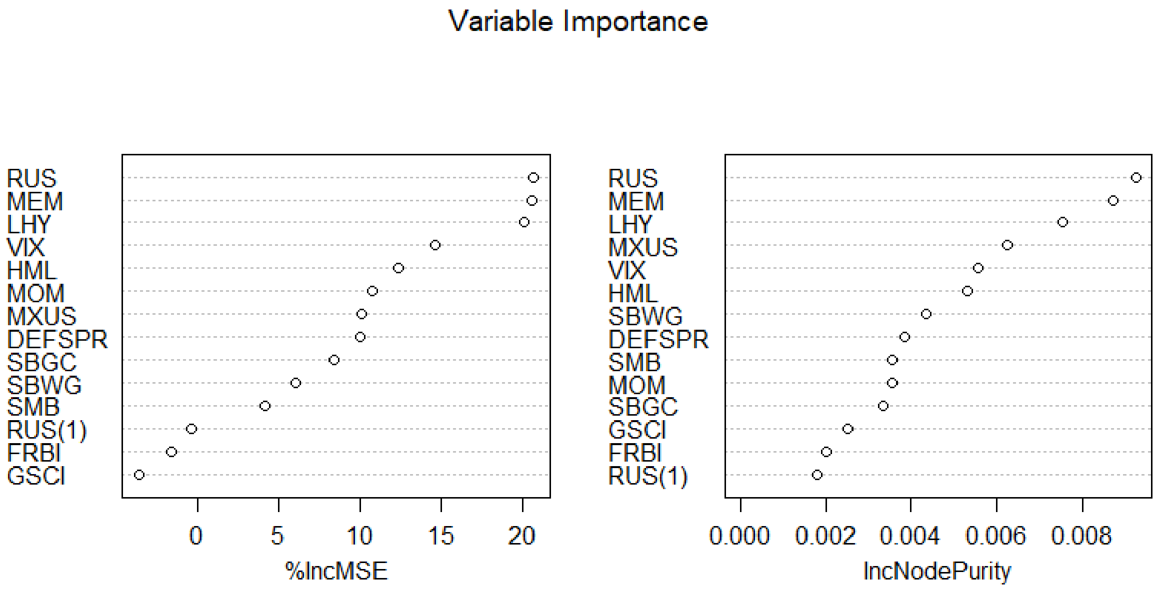

Figure 18 Variance Importance using %incMSE (Left) and incNodePurity (right)

The first diagram is the %IncMSE and the second the IncNodePurity. As we can observe the two diagrams are different. The most trust worthy diagram is the %IncMSE which practically use the OOB (out of bag) sample to calculate the MSE for a tree, after that uses randomization for each individual variable and then recalculate the new MSE. After taking the difference between the MSE with the randomize variable and the MSE of whole model we expect the first MSE to be larger and then we sum them over all the trees of the forest for each variable and scale them. Using this method we can calculate the important variables from the perspective of the prediction.

In the next table we can see the in sample and out of sample prediction for the Random Forest model.

Table 7 In sample and out of sample statistics of Random forest

	In sample	Out of sample
RMSE	0.007	0.006
MAE	0.005	0.005
MDAE	0.003	0.004

 https://doi.org/10.26219/heal.aueb.6490

The random forest gives better result for the in sample and out of sample period that the other models as we see from the above table. The model is also very good at prediction as we see from the figure 19 and 20.

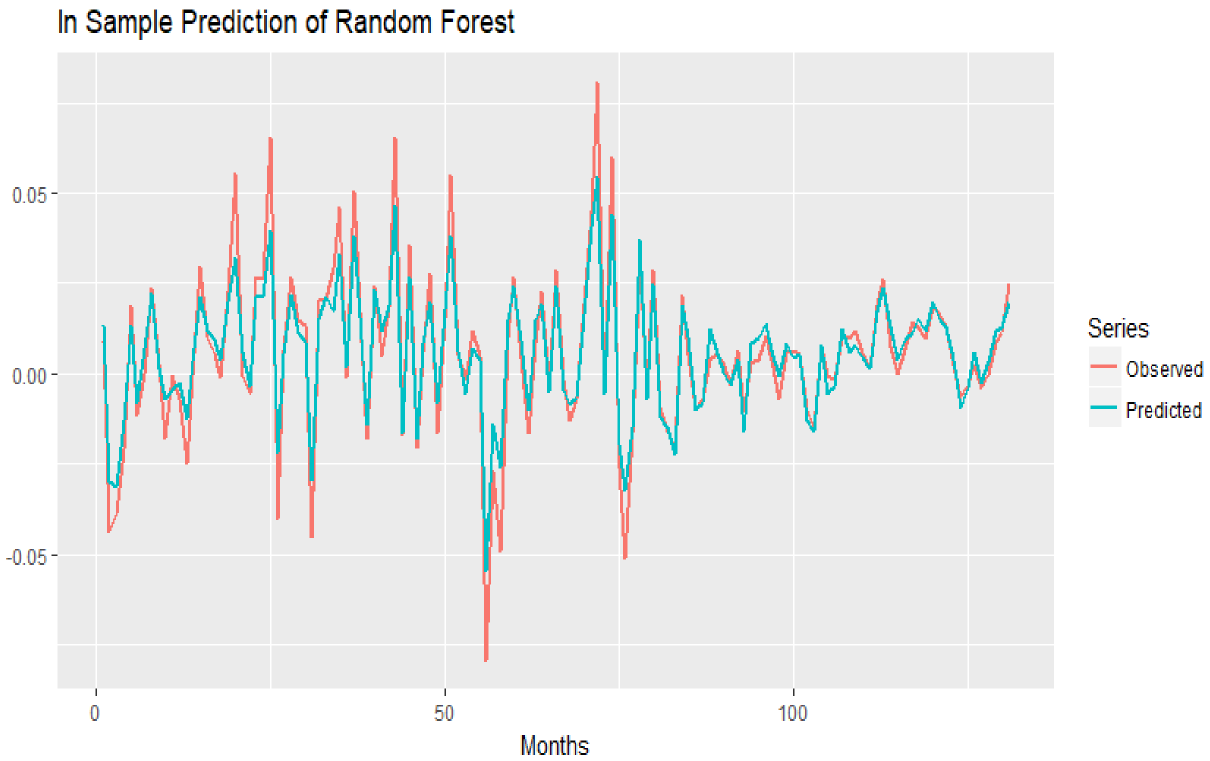

Figure 19 in sample prediction of Random forest

https://doi.org/10.26219/heal.aueb.6490

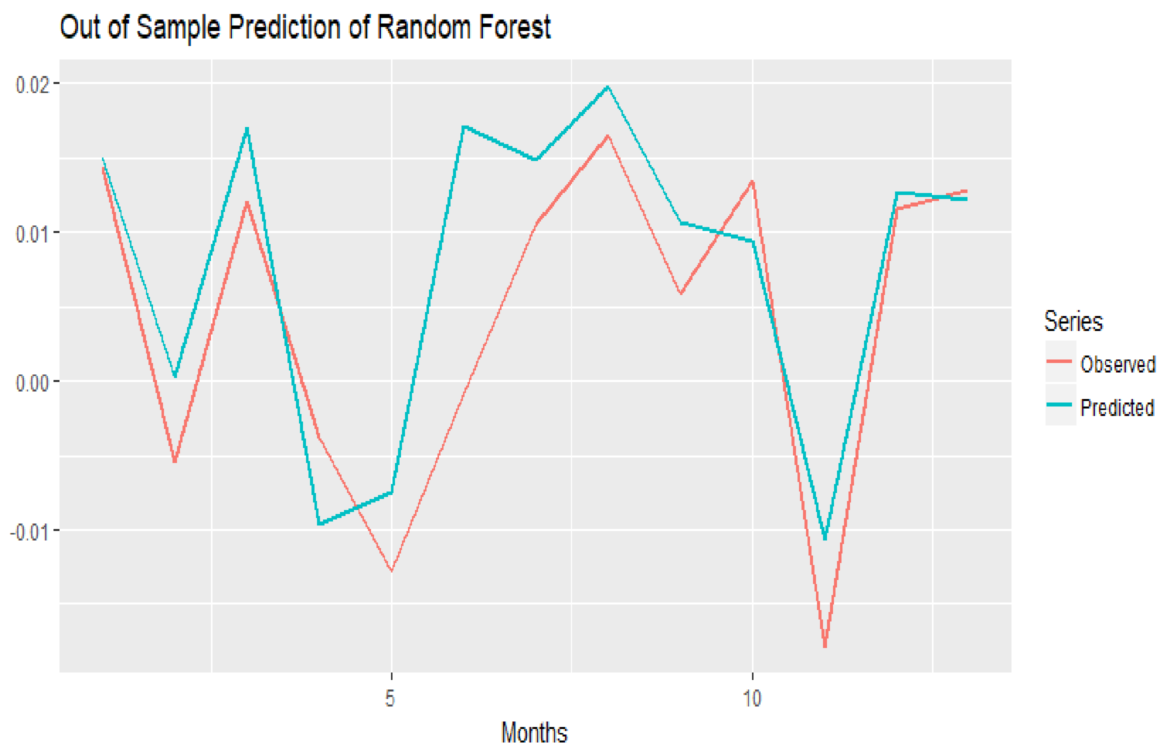

Figure 20 out-of- sample prediction of Random Forest

Checking the ACF and PACF of the model we observe that there is a structure in the squared and absolute residuals so we can use linear and nonlinear GARCH models for analysing the variance.

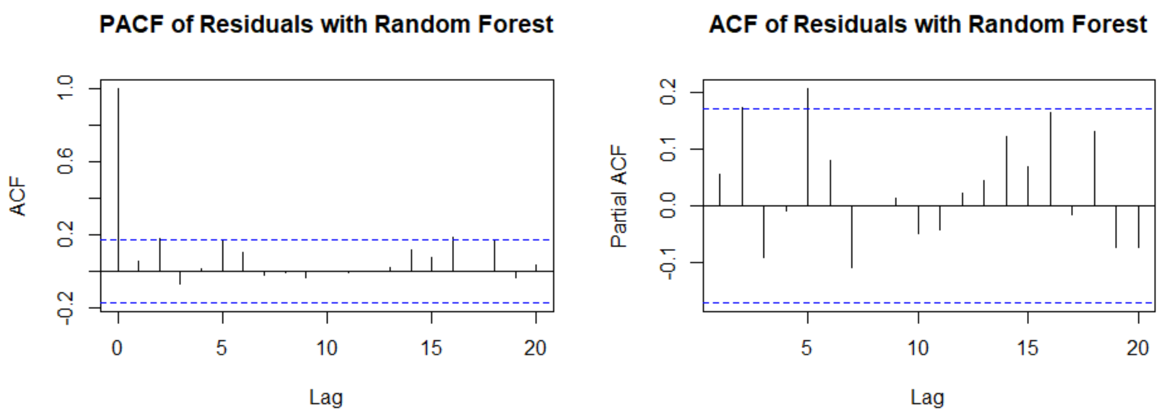

Figure 21 PACF and ACF of Squared Residuals from Random Forest

https://doi.org/10.26219/heal.aueb.6490

Using the linear and nonlinear GARCH models we have the following estimates for the coefficients

$$\text{GARCH:}\ \alpha_0 = 1 \times 10^{-8}\ \ \alpha_1 = 0.07\ \ b_1 = 0.90$$

$$\text{GARCH-GJR:}\ \omega = 8 \times 10^{-7}\ \ \alpha_1 = 0.05\ \gamma_1 = 0.09\ \ \beta_1 = 0.90$$

$$\text{VS-GARCH:}\ \omega = 1 \times 10^{-9}\ \ \zeta = 1 \times 10^{-8}\ \ \alpha_1 = 0.04\ \ \gamma_1 = 0.21\ \ \beta_1 = 0.98\ \ \delta_1 = 0.83$$

We have to mention that the coefficients of VS-GARCH is not to be trusted unfortunately. The starting values can determine the coefficients of VS-GARCH and if changed will be different. So for that reason we can't trust this model for this problem. Analysing the first two models we observe that the mean of coefficients of the errors in GARCH-GJR gives us the Coefficient of the error term in GARCH. GARCH-GJR also point us that there is a difference between the negative and positive news but the difference is not as large as we expected to be. That is probably because the Hedge Funds has less volatility in news than the rest of the market.

 https://doi.org/10.26219/heal.aueb.6490

4. Conclusion

Using the linear regression, the regression trees and the random forest we conclude that the most important variables in the dataset is the RUS, MEM , LHY, VIX, HML which are included to all three models. All the other variables are either important for each model uniquely or not at all. For the one step ahead out of sample prediction we saw that again the random forest is the best model using three different statistics which is the RMSE, MAE and MDAE with results 0.006, 0.005, 0.004 respectively. From the best model we saw that the volatility models are indeed different and specifically we see that the mean of the GARCH-GJR's coefficient where they express the leverage effect are half of the GARCH coefficient of the error. We can conclude that indeed positive news have larger volatility than negative which is strange because we assume that larger volatility comes from negative news. After looking closely the data we observe that positive returns are consist the 62% of the data so it is not that strange that positive coefficient is bigger than the negative. Of course we could assume that if we had a larger dataset the negative coefficient would be larger than the positive one. Also the Switching Volatility GARCH is the same for random forest and slightly different in linear regression and mean model. Unfortunately this model is not to be trusted because we don't know exactly its theoretical properties and also the coefficients were usually at the border of the constrains.

For later research we propose theoretical analysis for machine learning algorithms with GARCH models because this combination of models can make the predictions in financial industry more reliable to the users and for the market. This methods has not be introduced in the analysis and for financial data heteroscedasticity is very common and can be easily modelled with GARCH. Another research is for the hedge funds data to use different nonlinear GARCH models like EGARCH, ALST-GARCH and many more in order to extract information about the characteristics of volatility and for better and more reliably prediction. Finally we propose TAR and SETAR models to be used for the hedge fund data in order to find how many and what characteristic each regime has and extensions for the residuals using nonlinear GARCH models.

https://doi.org/10.26219/heal.aueb.6490

Bibliography

[1] **Franses H. P. and Van Dijk D. (2000).** *Non-linear time series models in empirical finance.* Cambridge University Press, UK

[2] **Tong H. (1978).** On a threshold model. *Pattern Recognition and Signal Processing*, Sijthoff &Noordhoff ,Amsterdam, 104-141

[3] **Tong H. (1990)**. *Non-linear time series: A Dynamical System Approach*, Oxford University Press, Oxford

[4] **Tong H. and Lim K.S. (1980)**. Threshold autoregressions, limit cycles, and data, *Journal of the Royal Statistical Society*,42,245-292

[5] **Hansen B. E. (2011)**. Threshold autoregression in economics, *STATISTICS AND ITS INTERFACE*, 4, 123-127

[6] **Hansen B. E. (1997)**. Inference in TAR models, *Studies in Nonlinear Dynamics and Econometrics*, 2, 1-14

[7] **Gospodinov N. (2005**). Testing For Threshold Nonlinearity in Short-Term Interest Rates, *Journal of Financial Econometrics*, Volume 3, Issue 3, Pages 344–371

[8] **Terasvirta T. (1994)**. Specification, Estimation, and Evaluation of Smooth Transition Autoregressive Models, *Journal of the American Statistical Association*,425,208-218

[9] **Umer U., Sevil T. and Sevil G. (2018)**. Forecasting performance of smooth transition autoregressive (STAR) model on travel and leisure stock index, *The Journal of Financial and Data Science*,4,90-100

[10] **Luukkonen R. , Saikkonen P. and Teräsvirta T. (1988)**. Testing Linearity Against Smooth Transition Autoregressive Models, *Biometrica*,75,491-499

[11] **Lanne M. and Saikkonen P. (2003)**. On Mixture Autoregressive Models, Statistics, *econometrics and society : essays in honour of Leif Nordberg*,69-89

[12] **Tsay R. S.(1998)**. Testing and Modeling Multivariate Threshold Models, *Journal of the American Statistical Association* , 443,1188-1202

[13] **Anderson H.M., Nam K. and Vahid F,(1999)**. Asymmetric nonlinear smooth transition GARCH models, *Nonlinear Time Series Analysis of Economic and Financial Data*, Boston: Kluwer, 191–207

[14] **Glosten L.R., Jagannathan R. and Runkle D.E., (1993)**. On the relation between the expected value and the volatility of the nominal excess return on stocks, *Journal of Finance*,48, 1779–1801

[15] **Fornari F. and Mele A. (1996)**. Modeling the changing asymmetry of conditional variances,*Economics Letters*,50, 197–203

https://doi.org/10.26219/heal.aueb.6490

[16] **Fornari F. and Mele A. (1997)**. Sign- and volatility-switching ARCH models: theory and applications to international stock markets, *Journal of Applied Econometrics*,12, 49–65

[17] **Audrino F. and Ьuhlmann P. (2000)**. Tree Structured GARCH Models, *Seminar fur Statistik Research Report*,91

[18] **Gokcan S. (2000)**. Forecasting volatility of emerging stock markets: linear versus non-linear GARCH models, *Journal of Forecasting*,19, 499-504

[19] **Wu J. (2011).**Threshold GARCH Model: Theory and Application. *The University of Western Ontario*, Canada

[20] **Gonźalez-Rivera G. (1998)**. Smooth-Transition GARCH Models, *Studies in Nonlinear Dynamics and Econometrics*, 3 , 61–78

[21] **Ramasamy R. and Munisamy S. (2012)**, Predictive Accuracy of GARCH, GJR and EGARCH Models Select Exchange Rates Application, *Global Journal of Management and Business Research*,12

[22] **Breiman L., Freidman J.H., Olshen R.A. and Stone, C.J. (1984)**. *Classification and regression trees*. Wadsworth, Belmont CA.

[23**] Dingli A. and Fournier K. S. (2017)**. FINANCIAL TIME SERIES FORECASTING – A MACHINE LEARNING APPROACH. *Machine Learning and Applications: An International Journal,* 4, 11-27

[24] **Abe M. and Nakayama H. (2018)**. Deep Learning for Forecasting Stock Returns in the Cross-Section, *PAKDD*,4,1-12

[25] **Chiang T., So M. and Chen C. (2005)**. Asymmetries in Return and Volatility and Composite Stock Return News–Evidence from Global Markets Based on a Bayesian Analysis

[26] **Giannikis D. and Vrontos I. (2010)**. A Bayesian approach to detecting nonlinear risk exposures in hedge fund strategies, *Journal of Banking & Finance*, 35 ,1399-1414

[27] **Vrontos I.(2015)**. Financial Econometrics, *Economical Athens University of Economics and Business*. Department of Statistics

[28] **Mandelbrot B. (1963)**.The variation of Certain Speculative Prices. *The Journal of Business*, 36, 4,394-419

[29] **Wikipedia** /Wikipedia Hedge funds https://en.wikipedia.org/wiki/Hedge_fund

[30] **Investopedia** https://www.investopedia.com/terms/h/hedgefund.asp

[31] **Agarwal, V. and Naik N.Y. (2004)**. Risks and portfolio decisions involving hedge funds. *Review of Financial Studies*, 17, 63-98.

[32] **Vrontos S., Vrontos I. and Giamouridis D. (2008)**. Hedge fund pricing and model uncertainty. *Journal of Banking & Finance*,32,741-753

[33] **Montgomery D. C., Peck E. A., Vining G. G. (2012).** *Introduction to Linear Regression Analysis*, Wiley

[34] **Akaike, H. (1973).** Information theory and an extension of the maximum likelihood principle. *In Second International Symposium on Information Theory*, Budapest: Akademiai Kiado, 267-281.

https://doi.org/10.26219/heal.aueb.6490

[35] **Hastie T., Tibshirani R. and Friedman J. (2009).** *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer second Edition

[36] **Kuhn M. and Johnson K. (2013).** *Applied Predictive Modeling* ,Springer

[37] **Bollerslev T. (1986).** A generalized autoregressive conditional heteroscedasticity. *Journal of Econometrics*,31, 307—327.