



A thesis submitted in partial fulfilment of the requirements for the degree of
MASTER IN ECONOMETRICS & MANAGEMENT SCIENCE
SPECIALIZATION QUANTITATIVE FINANCE

Using Extreme Value Theory depending on multiple covariates and Machine Learning algorithms to model market risk

Author:

STEPHY RUAN

Student number:

432296

Supervisor:

PROF. DR. C. ZHOU

Co-reader:

DR. P. WAN

December 4, 2019

Abstract

This study develops a model that incorporates covariates to the parameters describing the tail of the distribution of asset loss returns. The informative covariates are selected by Machine Learning algorithms. We derive Value-at-Risk depending on covariates and compare the predictive ability with classical risk models such as Historical Simulation and GARCH(1,1). The methodology is applied on two major world stock indices: S&P 500 and FTSE 100. We find that Dow Jones, NASDAQ and VIX are the most influential covariates to describe the left tail behavior for S&P 500. Furthermore, CAC40, AEX and DAX are the top three important covariates to explain the extreme negative returns of FTSE 100. Finally, our proposed models generate low VaR estimates without having more violations than expected.

Keywords— Extreme Value Theory, Machine Learning, Value-at-Risk, Market Risk

Contents

| | | |
|----------|--|-----------|
| 1 | Introduction | 3 |
| 2 | Data | 7 |
| 3 | Methodology | 10 |
| 3.1 | Peak-over-Threshold approach with covariates | 10 |
| 3.1.1 | Model setup | 10 |
| 3.1.2 | Parameters depends on covariates | 11 |
| 3.2 | Variable selection based on Machine Learning algorithms and parameters estimation | 14 |
| 3.2.1 | Random Forests | 14 |
| 3.2.2 | Support Vector Machines | 17 |
| 3.3 | Model comparion | 20 |
| 3.3.1 | Alternative models | 20 |
| 3.3.2 | Performance evaluation | 21 |
| 4 | Empirical Results | 24 |
| 4.1 | Variable selection | 24 |
| 4.1.1 | Parameter estimates | 24 |
| 4.1.2 | Variable importance | 32 |
| 4.1.3 | Individual covariate | 35 |
| 4.2 | Choice of threshold u | 38 |
| 4.3 | VaR Forecasting | 39 |
| 5 | Conclusion | 43 |
| | References | 45 |
| | Appendix A Tables | 48 |
| A.1 | Summary statistics of covariates | 48 |
| A.2 | Top three covariates VaR Forecasting for S&P 500 | 54 |
| A.3 | Top three covariates VaR Forecasting for FTSE 100 | 55 |

| | |
|---|-----------|
| Appendix B Graphs | 56 |
| B.1 Variable importance of $g(x)$ estimation for S&P 500 | 56 |
| B.2 Variable importance of $g(x)$ estimation for FTSE 100 | 59 |
| B.3 Variable importance of $\xi(x)$ estimation for S&P 500 | 62 |
| B.4 Variable importance of $\xi(x)$ estimation for FTSE 100 | 65 |
| B.5 VaR Forecasting | 68 |

1 Introduction

Tail risk refers to the risk of having an extreme event which occurs with low probability, but once occurs has high (adverse) impact. It can be measured by famous risk measures such as the Value-at-Risk (VaR). The Basel Committee on Banking Supervision (BCBS) requires banks to calculate the VaR for market risk in order to meet the minimum capital requirement. Extreme Value Theory (EVT) motivates statistical methods for estimating risk measures. For example, [Gilli and K ellezi \(2006\)](#) measures the financial risk with several major stock market indices using EVT model. [Dowd and Blake \(2006\)](#) applies EVT for insurance risk. In our research, we focus on the modeling of market risk using EVT. More specifically, we model the extreme negative returns or the left tail of stock returns distribution. Generally, there are two different ways to identify the extremes. First, the Block Maxima approach divides the observations into blocks. Then the maximum of each block represents the Block Maxima and these observations are considered as extremes which are usually modeled by Generalized Extreme Value (GEV) distribution. Second, the Peak-over-Threshold (POT) approach focuses on the exceedances beyond a given threshold. [Balkema and de Haan \(1974\)](#) theorem states that the excesses follow approximately the Generalized Pareto Distribution (GPD) when threshold is high. In this paper, we focus on the second approach because this technique obtains extremes more efficiently.

The assumptions in conventional EVT is that the observations are independent and identically distributed (i.i.d.). In practice, however, the observations often violate the identically distributed condition. Consequently, for modeling the excesses above a high threshold, the parameters of GPD may also depend on a number of covariates. In our research, we focus on the market risk by modeling the extreme losses in major stock indices. Because the financial returns might depend on many explanatory variables, we have to work with a high dimensional covariates. [Rigby and Stasinopoulos \(2005\)](#) consider a covariate model for the GPD in which the number of covariates are limited, for example, no more than four. To handle high dimensional covariates, Machine Learning (ML) algorithm is an efficient technique for dimensional reduction and it allows capturing the relation between variables in a non-parametric way. To the best of my

knowledge, combining Machine Learning techniques with Extreme Value Analysis has not been explored. This research aims at implementing the ML techniques into parameters estimation procedure in Extreme Value Analysis. In this way, we are able to examine how the covariates are related with the tail distribution and possibly improve the estimation of risk measures such as VaR. Additionally, the selection of threshold in the POT approach is a typical issue in Extreme Value Analysis. A high threshold decreases the bias in the estimation but increases the variance, while a low threshold increases the estimation bias. Therefore, as an extension for our research, we study the impact of different threshold on model estimation. In this paper, we consider the following research questions when using ML techniques to select covariates:

- Which covariates should be included to explain the left tail?
- Can the Value-at-Risk estimates be improved compared to the classical risk models?
- How to choose threshold in the Peak-over-Threshold approach?

We tackle these research questions in four steps. First, we follow the methodology of the basic POT approach with covariates as in [Chavez-Demoulin et al. \(2015\)](#). In this step, we introduce two group parameters in the model: the intensity λ which counts the loss frequency and the GPD parameters ξ and β that explain the loss severity. Then, we let the parameters be a function of some covariates and derive the estimation of VaR. Hereafter, we use Q-Q plots to check whether the observations in the tail follow approximately the correct probability distribution. Second, we explain how to estimate parameters depends on covariates with two ML techniques Random Forests (RF) and Support Vector Machines (SVM). This techniques allow us to select the informative covariates only. In the third step, we study the impact of the procedure with different thresholds. Finally, the VaRs are calculated. We evaluate the VaR estimation using a binomial test and an Interval forecasts evaluation test established by [Christoffersen \(1998\)](#). Also, we compare the performance with other models such as Historical Simulation (HS) and Dynamic HS GARCH(1,1).

We apply our methodology on a real dataset and study the market risk in the U.S. and U.K. stock markets. We use daily returns of two stock indices S&P 500 and FTSE 100.

Since we study the left tail of stock returns, we look at the negative returns. We use covariates to explain the tail of the stock index returns and it contains 63 potential covariates.

Literature Review

The idea of letting parameters of the Generalized Extreme Value (GEV) distribution depends on covariates in a parametric way is first developed by [Coles et al. \(2001\)](#). Later, [Chavez-Demoulin et al. \(2015\)](#) model operational risk depending on two pre-selected covariates in a semi-parametric way. There are two sets of parameters introduced. First, the intensity λ follows a non-homogeneous Poisson process with a rate function. The rate function is related to the covariates by a standard Generalized Additive Model (GAD). The parameter can be estimated via a penalized maximum likelihood estimator with a smoothing spline. Second, the loss severity follows a GPD with shape parameter ξ and scaling β . Since these parameters have to be orthogonal with respect to the Fisher information metric, reparametrize technique is used. The reparameterized parameters are related to the covariates via some smooth functions. They develop a backfitting algorithm and maximize the penalized likelihood function to estimate the parameters in the model. Then, the corresponding risk measures such as VaR and ES can be computed by simulation. [Embrechts et al. \(2018\)](#) apply the same method on modeling operational losses, but include firm-specific covariates associated with Internal Control Weaknesses (ICWs). By contrast, we study the market risk from the financial markets using EVT and implement a covariates selection procedure based on two ML algorithms, namely Random Forests and Support Vector Machines. Very few available literature pay attention to high dimensional covariates when fitting the tail distribution. [Hambuckers et al. \(2018\)](#) add a LASSO based L_1 -penalty term on the log-likelihood function when estimate the parameters of the tail distribution. This approach allows one to select informative covariates from a big dataset, based on Generalized Pareto (GP) Regression concept. Moreover, [Mayr et al. \(2012\)](#) establish a boosting algorithm to choose informative covariates from high dimensional covariates. Their method considers more as a regression model with multiple covariates in the gradient boosting framework. Beyond the regression models, Machine Learning algorithms allow capturing the relation between variables in a non-parametric way. These techniques are advantageous when one works

with high dimensional data and it is almost impossible with the traditional regression models because the number of covariates included in the model are generally limited. To our knowledge, combining ML techniques and EVT has never been explored before.

The structure of the paper is as follows. Section 2 introduces the data and provides potential covariates. Section 3.1 gives the model setup and derive the parameters depend on covariates. Section 3.2 discusses the Machine Learning algorithms and parameters estimation procedure. Section 3.3 shows the methods for evaluating VaR forecasting. Section 4 presents the empirical results and we conclude the research with Section 5.

2 Data

In this research, we study the market risk in the U.S. and U.K. stock markets. We use daily returns of S&P 500 and FTSE 100 indices. The sample period starts from August 2001 to March 2019 which includes 4518 trading days. Our sample period covers several extreme events: stock market downturn of 2002, August 2011 stock markets fall and the Global financial crisis of 2007-2008.

Due to the dynamic characters of equity market, there are plenty of potential variables related to the left tail of stock returns. Most of the covariates is pre-selected based on the available literature which is related to the (negative) extreme equity returns and others are created based on our own intuitions. For example, [Hilliard \(1979\)](#) researches the relation between the 10 major world exchanges during the global financial crisis. He investigates that the intra-continental equity indices move simultaneously in hourly fluctuation. Equivalently, [Zhong and Enke \(2017\)](#) uses 7 major world indices to forecast the stock returns. Combing these literature and the availability of data, we obtain 9 world major indices in our research covariates. We exclude the Australian S&P\ASX 200 and Italian FTSE MIB indices since they are not ranked at the top 10 Global Stock Exchanges by market capitalization as of November 30, 2018, while these stock indices are studied by [Hilliard \(1979\)](#).

Also, many studies focus on the effect of financial variables on the stock market. For instance, [Sadorsky \(1999\)](#) examines the impact of oil price shocks on equity market, [Baur and McDermott \(2010\)](#) shows that the gold price is negatively related to the share price in the developed market and [Flannery and James \(1984\)](#) proves that the interest rate changes are correlated with the common stock returns in the financial institutions. These covariates are selected as our research covariates, we measure the interest rate changes with Moody's Seasoned Corporate Bond Yield and London Interbank Offered Rate (LIBOR). Additionally, we include the silver, platinum and CBOE Volatility Index (VIX) as our potential covariates. For example, VIX is a sensitive indicator to the (extreme) equity market movements. Trading volume is earlier used by [Zhong and Enke \(2017\)](#) to forecast stock returns. We include 12 financial variables as our potential covariates.

Furthermore, [Campbell \(1987\)](#) examines the relationship between stock returns and term structure. [Fama \(1990\)](#), [Domian and Reichenstein \(1998\)](#) and [Gupta et al. \(2019\)](#) show the role of term and default spreads in stock returns prediction. [Ma and Kao \(1990\)](#) and [Ajayi et al. \(1998\)](#) identify the impact of currency changes on the stock market. [Brock et al. \(1992\)](#) and [Campbell et al. \(1993\)](#) analyze the correlation between financial returns and technical indicators such as moving averages and lagged variables. Based on the literature review, we obtain 7 T-bill rates of different maturities, 9 term spreads, 8 default spreads, 8 macroeconomics variables and 10 technical indicators as our candidate covariates. Hereby we create two dummy covariates which measures the period of expansion or recession and the presence of an inverted yield curve.

Finally, we obtain 63 potential covariates to explain the left tail of the daily returns. All the covariates are given in daily frequency. A list of 63 covariates candidates and the summary statistics are shown in [Appendix A.1](#).

We use Jarque-Bera to test the normality of all variables used and the results are shown in the last column. All covariates are rejected for normality. The world major indices have negative skwness and excess kurtosis, which indicates a higher probability to obtain a negative daily returns and extreme events than the normal distribution. This reflects the fat-tailedness in the left tail. We plot the daily prices and losses returns for S&P 500 and FTSE 100 in [Figure 1](#) and [2](#), respectively.

We deal with missing data in our dataset as follows. The world ten major indices contain the most missing values, since each Stock Exchange has their own opening hours and public holiday days. To solve this problem, we use linear interpolation technique to fill in the missing values. We replace the missing data by taking the mean of the last non-missing and the next non-missing value within the variable, one of the disadvantage is that the observations become less extreme.

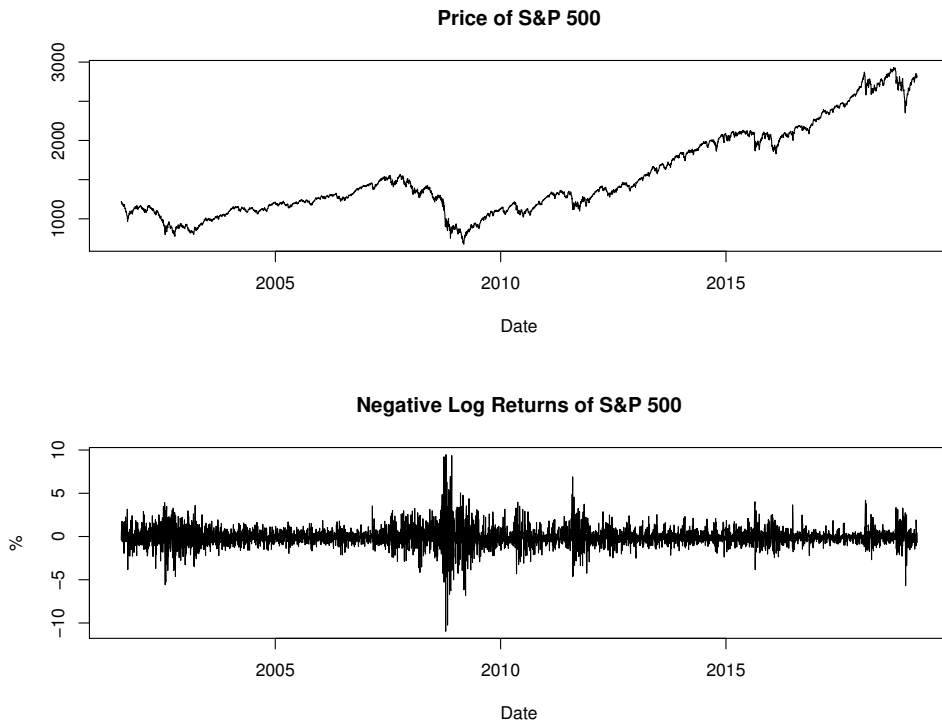


Figure 1: The price and negative log returns of S&P 500 over sample period 2001:08 - 2019:03.

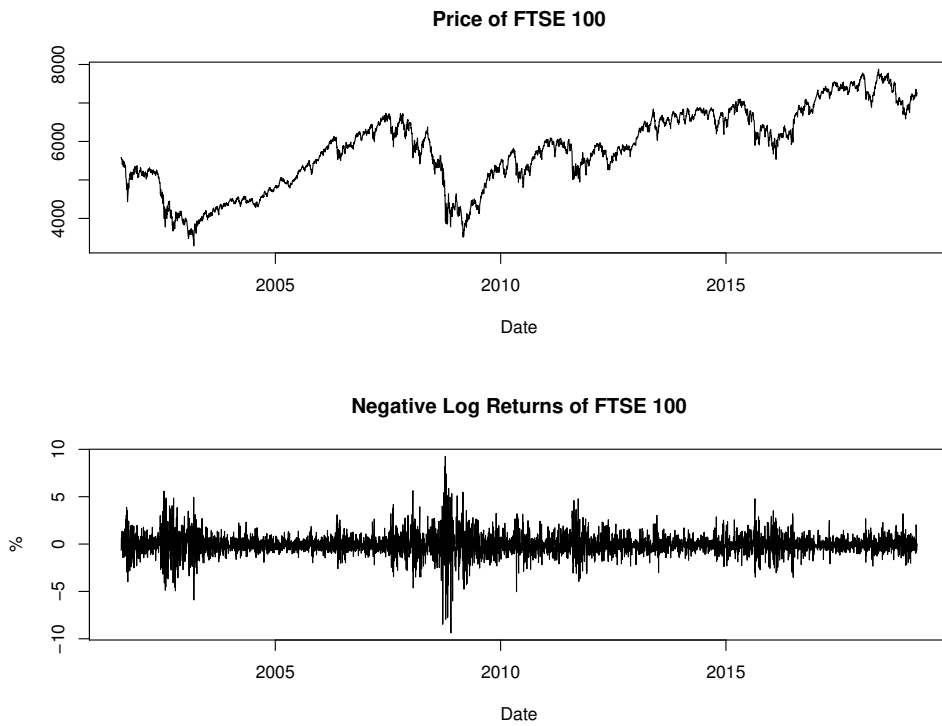


Figure 2: The price and negative log returns of FTSE 100 over sample period 2001:08 - 2019:03.

3 Methodology

In this section, we first introduce the basic Peak-over-Threshold model in Section 3.1.1. Then, we let the conditional probability of exceedances over threshold depends on some covariates and express it as a function $g(x)$ of covariates x . Afterwards, we derive the estimation of shape parameters $\xi(x)$ by transforming distributions. Also, we compute the risk measure VaR. Subsequently, we check whether the exceedances indeed follows the correct probability distribution with Q-Q plots. This is done in Section 3.1.2. Secondly, we describe the procedure of parameters estimation using Random Forests and Support Vector Machines in Section 3.2.1 and 3.2.2, respectively. Thirdly, we compare the VaR forecasting depends on covariates with other classical risk models and evaluate the model performance in Section 3.3.

3.1 Peak-over-Threshold approach with covariates

3.1.1 Model setup

We introduce the Peak-over-Threshold (POT) approach following the step in Embrechts et al. (1997). The key idea of this approach is to approximate the distribution of exceedances over a certain threshold. Let Z be a sequence of financial losses with marginal distribution F . Given a high threshold u , denote the threshold exceedances $Y = Z - u$. The excess distribution of Z is called *conditional excess distribution function* and given by

$$F_u(y) = P(Z - u \leq y | Z > u), \quad 0 \leq y \leq z_F - u$$

where $z_F \leq \infty$ is the right endpoint of F .

Under the i.i.d. assumption, the following two properties hold

1. S counts the number of exceedances over threshold u for a given period and it follows a non-homogeneous Poisson process with intensity λ . In the figure below, observation z_2 , z_3 , z_5 , z_6 , z_8 and z_9 are considered as exceedances. In this particular case, the number of exceedances S is 6.

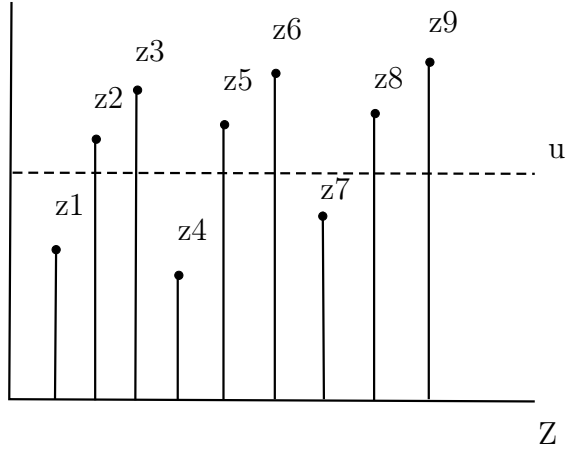


Figure 3: An example of POT approach

2. The excess over a high threshold u is approximated by the Generalized Pareto Distribution (GPD):

$$F_u(y) \approx G_{\xi, \beta}(y), \quad \text{as } u \rightarrow \infty,$$

where $\beta = \beta(u)$ depends on the threshold u , and

$$G_{\xi, \beta}(y) = \begin{cases} 1 - (1 + \frac{\xi}{\beta}y)^{-1/\xi} & \text{if } \xi \neq 0 \\ 1 - \exp(-\frac{y}{\beta}) & \text{if } \xi = 0 \end{cases},$$

for $y \in D(\xi, \beta)$ where

$$D(\xi, \beta) = \begin{cases} [0, +\infty) & \text{if } \xi \geq 0 \\ [0, -\frac{\beta}{\xi}] & \text{if } \xi < 0 \end{cases},$$

3.1.2 Parameters depends on covariates

Since the aim of our research is to model the tail distribution depending on multiple covariates, we let the number of excessdances S with intensity function λ and the GPD parameters depends on covariates x . First, the number of exceedances S with intensity function $\lambda(x)$ does not follow a homogeneous Poisson point process anymore. A non-homogenous Poisson process is defined as a Poisson process with a time-varying arrival

rate $\lambda(t)$. In our case, the intensity function is not time-varying, but it consists of 63 different covariates. Instead of model the number of excessdances S , we calculate the conditional probability of Z excess threshold u given covariates x :

$$g(x) = \mathbb{P}(Z > u | X = x) = \mathbb{E}(I_{Z>u} | X = x) \quad (1)$$

where $I[\cdot]$ is an indicator function that equals to 1 when $Z > u$ and otherwise 0. In the last term we use the fact that the probability of an event is the expectation of an indicator function for an event. Because the dependent variable $I_{Z>u}$ has a binary outcome and we can compute the conditional expectation using classification. Therefore, we can estimate the function $g(x)$ for any set of covarites x with Machine Learning techniques.

Second, we let the GPD parameters depend on covariates x , denote as $G_{\xi(x),\beta(x)}(y)$. Since we study the financial losses from the stock market, assume $\xi(x) > 0$ (See [Chavez-Demoulin et al. \(2015\)](#)). Then the scaling parameter $\beta(x) = \xi(x)u$. In other words, we do not need to estimate $\beta(x)$ anymore. It follows that $\frac{Z}{u} \Big| Z > u, X = x$ is approximately Pareto distributed with a shape parameter $\frac{1}{\xi(x)}$:

$$\begin{aligned} \frac{Z}{u} \Big| Z > u, X = x &\stackrel{d}{\approx} PD\left(\frac{1}{\xi(x)}\right) \\ \mathbb{P}\left(\frac{Z}{u} \leq y \Big| Z > u, X = x\right) &\rightarrow 1 - y^{-1/\xi(x)} \end{aligned}$$

Take the log-transformation and define $L = \log\left(\frac{Z}{u} \Big| Z > u\right)$,

$$L \Big| X = x \stackrel{d}{\approx} \text{Exp}\left(\frac{1}{\xi(x)}\right), \quad (2)$$

and it follows that

$$\mathbb{E}(L \Big| X = x) = \mathbb{E}\left(\log\left(\frac{Z}{u} \Big| Z > u, X = x\right)\right) \approx \xi(x) \quad (3)$$

In the last equation, $\xi(x)$ is written as a function of covariates x and it is approximately equal to the conditional expectation of L given x are known. In other words, we can estimate $\xi(x)$ for any given covariates set by a regression analysis. For example, we can

estimate the parameter $\xi(x)$ using Random Forests regression tree and Support Vector Machines.

We compute the risk measure VaR depends on covariates as follows. Define $VaR_\alpha(Z|X = x) = f(x)$, which satisfies $\mathbb{P}(Z \leq f(x)|X = x) = \alpha$. Approximate

$$\begin{aligned} \frac{\mathbb{P}(Z > f(x)|X = x)}{\mathbb{P}(Z > u|X = x)} &\approx \left(\frac{f(x)}{u}\right)^{-1/\xi(x)} \\ \frac{1 - \alpha}{g(x)} &\approx \left(\frac{f(x)}{u}\right)^{-1/\xi(x)} \end{aligned}$$

Finally, the estimation of VaR depends on covariates is

$$\widehat{VaR}_\alpha(Z|X = x) = \hat{f}(x) = u \left(\frac{1 - \alpha}{\hat{g}(x)}\right)^{-\hat{\xi}(x)} \quad (4)$$

After estimation of the parameters, we check with Q-Q plots whether the exceedances indeed follows $PD\left(\frac{1}{\xi(x)}\right)$ for $\xi(x) > 0$. This is done by examine whether $\frac{L}{\xi(x)} \Big| X = x$ follows approximately the standard exponential distribution.

3.2 Variable selection based on Machine Learning algorithms and parameters estimation

3.2.1 Random Forests

As follows we use Random Forests algorithms to estimate the parameter $g(x)$ by classification and $\xi(x)$ by regression. Random forests is a famous machine learning technique and the idea is to build an ensemble of decision trees. Each decision tree is build by drawing a bootstrap sample from the original training data. Subsequently, select randomly a subset of all covariates and pick the best covariate to split at each node until the maximum number of nodes to split is reached. Then, a prediction is made for each decision tree. The final predicted response variable is calculated by taking the class with the most votes in case of classification and averaging the predicted values over all decision trees in case of regression.

The advantage of random forests is easy to implement and can handle the high dimensional of predictor variables. Additionally, random forests technique reduces the variance of prediction without increasing the bias by drawing randomly different sample from the original data, which improves the accuracy of predictions. The disadvantage is the complexity of random forests which means that the results are difficult to interpret. Since the popularity of random forests algorithm, there are many open source implementations available in different programming languages. In this paper, we use the **randomForest** package in R and Algorithm 1 shows the estimation procedure on next page.

Algorithm 1.

1. We choose a confidence level α , threshold u , then we split the data into training data of size M and test data of size N . Each parameter is computed based on an estimation window of size M .
2. For each estimation window, we do the following steps:
 - (a) For classification, create a dummy variable D which takes $D = 1$ when $Z > u$ and $D = 0$ when $Z \leq u$. We train the RF model with dummy variable D and covariates x as follows.
 - i. Draw n_{tree} bootstrap sample from the original training sample of size M , where n_{tree} defines as the number of trees in the random forests. [Oshiro et al. \(2012\)](#) show that the model performance will not increase after 128 trees and they suggest that the optimal number of trees in the random forests lies between 64-128. Therefore, we set $n_{tree} = 128$.
 - ii. For each decision tree in the random forests, select randomly \sqrt{p} predictors of all covariates, where p is the number of covariates in the dataset and this number is rounded down. Then, we choose the best covariate to split at each node until the maximum number of nodes is reached. Afterwards, a prediction is made for each tree in the random forests and the outcome is 0 or 1.
 - iii. Use the trained RF model to make prediction for new covariates in the testing sample. For a new set of covariates x , estimate $g(x)$ in Equation (1) by counting the number of trees in the random forests which predicts 1 and then dividing by $n_{tree} = 128$. The estimation of $g(x)$ is the predicted probability of Z excess threshold u with random forests classifier.
 - (b) For regression, create a new variable $L = \log\left(\frac{Z}{u} \mid Z > u\right)$. This variable is observed only for $Z > u$, thus results in a smaller size of training sample. We fit the model with generated variable L and covariates x as follows.
 - i. Draw $n_{tree} = 128$ bootstrap sample from the training sample of size M as for classification.

- ii. *Select randomly $p/3$ predictors of all covariates, where p is the number of covariates and this number is rounded down. Subsequently, we pick the best covariate to split at each node until the maximum number of nodes is reached in the same manner as for classification. Afterwards, a prediction is made for each tree in the random forests and the outcome is a numeric value.*
 - iii. *Use the trained RF model to make prediction for new covariates in the testing sample. For a new group of covariates, estimate $\xi(x)$ in Equation (3) by averaging the prediction over $n_{tree} = 128$ decision trees in the random forests. The shape parameter depending on multiple covariates is approximately equal to the predicted mean of L , thus $\xi(x)$ is automatically estimated by RF algorithm.*
- (c) *When we obtain the estimation for $g(x)$ and $\xi(x)$ in step 2.a and 2.b, respectively. We compute the VaR depends on covariates stated in Equation (4).*

3.2.2 Support Vector Machines

The second Machine Learning algorithm we use is Support Vector Machines (SVM). Similar to Random Forests, we estimate $g(x)$ using classification and $\xi(x)$ using regression technique.

SVM is widely used for classification, the goal is to find an optimal hyperplane which can separate the two classes correctly. In general, we deal with three issues. First, many hyperplanes can be found. In order to choose which hyperplane to use, we select hyperplane based on Maximal Margin Classifier. Margin is the distance between the nearest data point of each class to the hyperplane and these data points are also called support vectors. Maximal Margin Classifier is to find a hyperplane which maximizes the margin. Second, when a hyperplane separates the classes correctly without any errors or outliers, we call it hard margin classifier. However, in practice this is not always the case. For this reason, we need to use a soft margin classifier. Soft margin classifier includes some misclassified observations, make a trade-off between maximizing the margin and minimizing the misclassified data points. Third, in general case the data are not linearly separable and more complex functions are needed. Therefore, we use different kernel functions. A kernel function maps the data into high dimensional space in such a way that the data points are more (linearly) separable, which means that the data can be better classified. In this paper, we handle three different kernel functions which are linear, polynomial and radial. SVM is considered as an optimization problem and the optimal hyperplane can be found by Lagrangian multipliers.

SVM can also be used for regression, the idea is similar but the goals are different. Rather than find a hyperplane which separates the classes correctly, the goal for regression is to find an optimal hyperplane which minimizes the distance of the data points to it. The cost function for regression is based on epsilon-insensitive loss function, the idea is that the small amounts of errors are not count and errors greater than the epsilon range are needed to be minimize. As for classification, we use three different kernel functions to separate the non-linear data, which are linear, polynomial and radial. Again, we use an open source statistical package **e1071** in R and a roadmap of parameter estimation procedure for SVM is given in Algorithm 2.

Algorithm 2.

1. We choose a kernel function K , confidence level α , threshold u and split the data into training sample with size M and test sample with size N . Again each parameter is computed based on an estimation window of size M .
2. For each estimation window, we do the following steps:
 - (a) For classification, we create a dummy variable D in the same way as for RF. We train the model with dummy variable D and covariates x as follows.
 - i. Find an optimal hyperplane based on Maximal Margin Classifier, which means that we maximize the distance between the nearest observation of each class to the hyperplane. This is an standard optimization problem and can be solved by Lagrangian multiplier.
 - ii. In case that some data points are classified incorrectly in the separate hyperplane, we use SVM with soft margin technique. Hereby we introduce a cost parameter in the objective function of SVM which makes the trade-off between maximizing margin and minimizing misclassified data points. Until now, this is an optimization problem with a linear kernel.
 - iii. When the data points are not separable with linear classifier, we use two different non-linear kernel functions to make the data linearly separable. Beforehand, we need to decide the hyper-parameters. We use a grid search to find the following optimal hype-parameters for two non-linear kernel function:
SVM Polynomial: $\text{gamma}=0.01, \text{cost}=0.01, \text{degree} = 3, \text{coef0} = 1$
SVM Radial: $\text{cost}=0.01$
 - iv. Next, we use the hype-parameters identified in the previous step and non-linear kernel function to find the optimal hyperplane.
 - v. Finally, use the trained SVM to make prediction for new covariates in the testing sample. For a new set of covariates x , estimate $g(x)$ in Equation (1) by predicting the mean of D .
 - (b) For regression, create a new variable $L = \log\left(\frac{Z}{u} \mid Z > u\right)$ in the same manner as for RF. Notice that the number of observations in the training sample

becomes smaller. We fit the SVM with generated variable L and covariates x as follows. The steps are similar as for classification.

- i. Find an optimal hyperplane based on Maximal Margin Classifier by solving optimization problem.*
 - ii. Introduce a SVM with soft margin when some data points are classified incorrectly in the separate hyperplane. Again, this is considered as optimization problem with linear kernel.*
 - iii. Use non-linear kernel functions to make the data linearly separable. The optimal hyper-parameters for polynomial and radial kernels are the same as for classification.*
 - iv. Finally, use the trained SVM to make prediction for new covariates in the testing sample. For a new group of covariates, use the fitted SVM to estimate $\xi(x)$ in Equation (3) by predicting the mean of L . Again, the shape parameter depends on covariates is automatically estimated by SVM algorithm.*
- (c) *When we obtain the estimation for $g(x)$ and $\xi(x)$ in step 2.a and 2.b, respectively. We compute the VaR depends on covariates stated in Equation (4).*

3.3 Model comparion

In this section, we discuss how to forecast the Value-at-Risk depending on covariates and compare the model performance with other classical risk models such as Historical Simulation (HS) and dynamic HS with GARCH(1,1) model. To evaluate the model performance, we employ an unconditional binomial test mentioned by [Jorion et al. \(2010\)](#) and a conditional Interval forecasts evaluation test introduced by [Christoffersen \(1998\)](#).

3.3.1 Alternative models

Historical Simulation

The Historical Simulation is a classic and easy to implement method for estimating VaR. The idea is based on empirical distribution function and no assumptions are made on the parameters distribution. Consider a training sample with size M and testing sample with size N , again each prediction in the testing sample is calculated based on a moving window of M trading days. Consider the negative log-returns Z and take the order statistics such that $Z_{(1)} \leq Z_{(2)} \leq \dots \leq Z_{(M)}$. The estimation of VaR based on a moving window of size M is given by

$$\widehat{VaR}_\alpha^{HS} = Z_{(\lceil M\alpha \rceil)}$$

where M is number of observations in the training sample and α is the confidence level.

Dynamic HS with GARCH(1,1)

The second approach is called dynamic HS with GARCH(1,1) model. The difference with the first approach is that we obtain a GARCH(1,1) model as an immediate step to forecast volatility, afterwards we use it to compute the VaR. An advantageous of this method is that we can capture the volatility clustering in the time series since almost all stock returns presents autocorrelations. The data is again divided into training sample of size M and testing sample of size N . Consider $t = 1, \dots, M$ and the GARCH(1,1) model defines as

$$Z_t - \mu = \sigma_t \epsilon_t, \quad \sigma_t^2 = \alpha_0 + \alpha_1 Z_{t-1}^2 + \beta_1 \sigma_{t-1}^2$$

where Z_t is the negative log-return as mentioned before, μ is the estimates of the mean

of $Z_t, \epsilon_t \sim N(0, 1)$, $\alpha_0 > 0$, $\alpha_1, \beta_1 \geq 0$ and $\alpha_1 + \beta_1 < 1$ for stationary. The coefficients in the model can be estimated with Quasi Maximum Likelihood and we obtain $\hat{\alpha}_0, \hat{\alpha}_1, \hat{\beta}_1$ and the mean $\hat{\mu}$. Then, we use a start value σ_1 and estimate the in-sample volatility $\hat{\sigma}_t$ for $t = 2, \dots, M$. Subsequently, we forecast the volatility at $M + 1$,

$$\hat{\sigma}_{M+1}^2 = \hat{\alpha}_0 + \hat{\alpha}_1 Z_M^2 + \hat{\beta}_1 \hat{\sigma}_M^2$$

and the VaR estimate is given by

$$\widehat{VaR}_\alpha^{HS-GARCH} = \hat{\mu} + \hat{\sigma}_{M+1} \widehat{VaR}_\alpha^{FHS}$$

where $\widehat{VaR}_\alpha^{FHS}$ is the VaR estimation of filtered ϵ_t using HS simulation and filtered ϵ_t is computed by $Z_t/\hat{\sigma}_t$.

3.3.2 Performance evaluation

Binomial Test

The binomial test is standard backtesting method for VaR estimates. The concept is based on counting the number of violations for a testing period of N and $V = \sum_{i=1}^N I_i(\alpha)$, where

$$I_i(\alpha) = \begin{cases} 1 & \text{if } Z_i > \widehat{VaR}_\alpha \\ 0 & \text{if } Z_i \leq \widehat{VaR}_\alpha \end{cases}$$

If the number of violations V follow a binomial distribution $B(N, 1 - \alpha)$, the test statistics is given by

$$\frac{V - N(1 - \alpha)}{\sqrt{N\alpha(1 - \alpha)}}$$

which follows approximately the standard normal distribution as $N \rightarrow \infty$.

Interval forecasts evaluation

[Christoffersen \(1998\)](#) introduces an Interval forecasts evaluation method to test the model performance on VaR estimates. [Christoffersen \(1998\)](#) tests the following two hypotheses:

1. The unconditional coverage hypothesis states that the probability of a violation must equal to the coverage rate, in this case is $1 - \alpha$. Denote $\pi = \mathbb{P}[I_i(\alpha) = 1]$ and we

test the null hypothesis $H_0 : \pi = 1 - \alpha$. The likelihood function of all $I_i(\alpha)$ is given by:

$$L = \prod_{i=1}^N \pi^{I_i(\alpha)} (1 - \pi)^{1 - I_i(\alpha)} = \pi^V (1 - \pi)^{N - V}$$

where α is the confidence level, N is the testing period and V is the number of violations as defined before. Since the maximum likelihood estimate of π is $\hat{\pi} = \frac{V}{N}$, the unconditional coverage likelihood ratio statistic equals to:

$$LR_{UC} = -2 \log \left(\frac{(1 - \alpha)^V \alpha^{N - V}}{\left(1 - \frac{V}{N}\right)^{N - V} \left(\frac{V}{N}\right)^V} \right) \sim \chi^2(1)$$

2. The independence hypothesis states that the VaR violations are independent from each others. The idea is inspired by the property of Markov chain, denote a matrix of transition probabilities:

$$\Pi = \begin{pmatrix} \pi_{00} & \pi_{01} \\ \pi_{10} & \pi_{11} \end{pmatrix}$$

where $\pi_{j,k} = \mathbb{P}[I_i(\alpha) = k | I_{i-1}(\alpha) = j]$, 0 represents no violation and 1 represents a violation occurred. Denote $n_{j,k}$ is a random variable which counts the number of times event j followed by event k and an outcome table is illustrated below:

Table 1: Outcome of Christoffersen's test

| | $I_{i-1}(\alpha) = 0$ | $I_{i-1}(\alpha) = 1$ | |
|-------------------|-----------------------|-----------------------|-------------------|
| $I_i(\alpha) = 0$ | n_{00} | n_{10} | $n_{00} + n_{10}$ |
| $I_i(\alpha) = 1$ | n_{01} | n_{11} | $n_{01} + n_{11}$ |
| | $n_{00} + n_{01}$ | $n_{10} + n_{11}$ | N |

π_{01} and π_{11} are random variables. The maximum likelihood estimates are equals to:

$$\hat{\pi}_{01} = \frac{n_{01}}{n_{00} + n_{01}}$$

$$\hat{\pi}_{11} = \frac{n_{11}}{n_{10} + n_{11}}$$

The likelihood ratio test statistics of independence is given by:

$$LR_{IND} = -2\log \left(\frac{\left(1 - \frac{V}{N}\right)^{N-V} \left(\frac{V}{N}\right)^V}{(1 - \hat{\pi}_{01})^{n_{00}} (\hat{\pi}_{01})^{n_{01}} (1 - \hat{\pi}_{11})^{n_{10}} (\hat{\pi}_{11})^{n_{11}}} \right) \sim \chi^2(1)$$

Since [Christoffersen \(1998\)](#)'s method is a jointly test of unconditional coverage and independence, the likelihood ratio statistics of Conditional Coverage Independence (CCI) test is computed by:

$$LR_{CCI} = LR_{UC} + LR_{IND} = -2\log \left(\frac{(1 - \alpha)^V \alpha^{N-V}}{(1 - \hat{\pi}_{01})^{n_{00}} (\hat{\pi}_{01})^{n_{01}} (1 - \hat{\pi}_{11})^{n_{10}} (\hat{\pi}_{11})^{n_{11}}} \right) \sim \chi^2(2)$$

4 Empirical Results

The empirical results consist of three parts. First, we perform a covariate selection analysis in Section 4.1. Second, Section 4.2 tests the sensitivity of threshold u on the model performance. Finally, we compare the VaR forecasting with 6 different models in Section 4.3.

4.1 Variable selection

In this subsection, we perform a variable selection analysis and attempt to answer our main research question - Which covariates should be included to explain the left tail? We apply the methodology on dataset S&P 500 and FTSE 100, the aim is to find economically meaningful covariates to explain the tail behavior. We work out in the following three steps. First, we present the results of parameter estimates with RF and SVM, namely the estimation of $g(x)$ and $\xi(x)$. Afterwards, we investigate whether the exceedances indeed follow approximately Pareto distribution using Q-Q plots. Second, we rank the variable importance and detect which covariates give the most predictive power on estimating parameters of the left tail distribution. Lastly, we examine the relation between an individual covariate and predicted parameters with scatter plots. In other words, we research how does an individual covariate affects tail behavior and give economic interpretation.

4.1.1 Parameter estimates

In this paragraph, we follow the parameter estimation procedure described in Algorithm 1 and 2. First, we examine the predictive ability of $g(x)$ which estimates the probability of exceedances above threshold u in terms of Area Under the Receiver Operating Characteristic curve (AUROC). Receiver Operating Characteristic curves are a useful tool to classifies binary outcomes, it plots the True Positive Rate (Sensitivity) on the y-axis and $1 - \text{False Positive Rate}$ (Specificity) on the x-axis for a number of different thresholds. In our case, True Positive Rate calculates the percentage of correct predictions for $g(x)$ when in fact there is exceedances above threshold u . False Positive Rate computes the percentage of incorrect prediction for $g(x)$ when there is no exceedances above threshold in reality. Therefore, a high True Positive Rate and low False Positive Rate are desired.

Which means we prefer high values of Sensitivity and Specificity. An example of the ROC curve is illustrated in Figure 4.

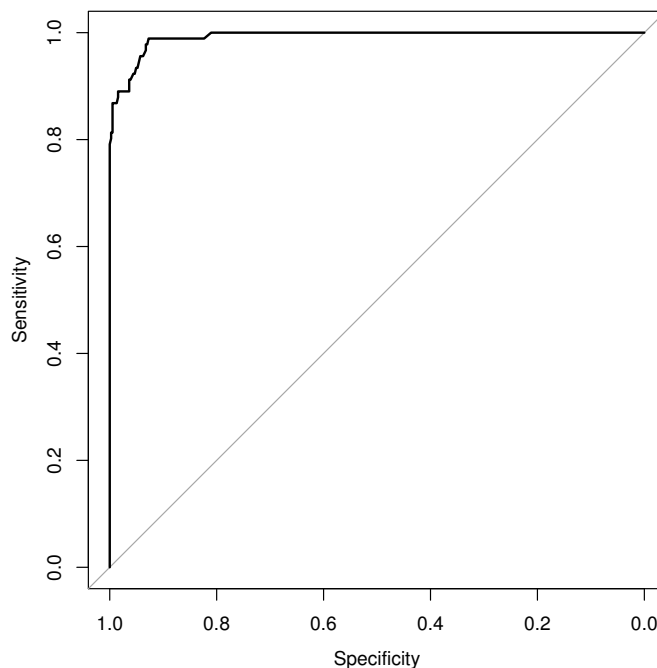


Figure 4: An example of ROC curve. The 45-degree straight line has an AUROC of 0.5 which represents random guess on the binary outcome. The black line has an AUROC of 0.993 which is calculated by the RF model and indicates an excellent classifier.

As the name itself says, AUROC computes the area under the ROC curve which ranges between 0 and 1. An AUROC of score 0.5 indicates that the model is equivalent to random guess and 1 denotes a perfect classifier. We calculate the AUROC with datasets S&P 500 and FTSE 100 using RF, SVM Linear, SVM Polynomial and SVM Radial. The results are shown in Table 2.

Table 2: AUROC Comparison

| | RF | SVM Linear | SVM Polynomial | SVM Radial |
|----------|-------|------------|----------------|------------|
| S&P 500 | 0.993 | 0.991 | 0.957 | 0.975 |
| FTSE 100 | 0.919 | 0.916 | 0.897 | 0.911 |

In general, we find that both 4 models are an excellent classifier with $AUROC > 0.90$. The predictive power has the following ranking: $RF > SVM\ Linear > SVM\ Radial > SVM\ Polynomial$. One explanation of high score of AUROC is that we apply a moving window to train the ML models each time, which allows the models learn the most

recent information till day $t - 1$ to predict day t . Another explanation is that we use large amounts of data which includes 63 covariates. Last but not least, we illustrate the estimation of $g(x)$ with RF using indices S&P 500 and FTSE 100 in Figure 5. From the graph, we observe that the predicted probabilities $g(x)$ are high when there is exceedances which meets our expectations.

Second, we estimate the parameter $\xi(x)$ which is the shape parameter of the Pareto distribution and it is an important parameter since it determines the tail behavior. As the shape parameter measures the heaviness of the tail, we expect a higher value of shape parameter when the losses are large. We identify the relation between these variables with scatter plots. Figure 6 and 7 display the scatter plots using RF, SVM Linear, SVM Polynomial, SVM Radial and applied on datasets S&P 500 and FTSE 100, respectively. Overall, we examine a positive relation between the estimate $\xi(x)$ and the losses. For example, we detect a strong positive linear relationship between the predicted shape parameter and S&P 500 losses when using RF model. Also, SVM Linear and SVM Polynomial show a similar pattern in the scatter plots. However, SVM Radial leads to a non-linear relation between $\hat{\xi}(x)$ and the losses.

Finally, Figure 8 illustrates the Q-Q plots for S&P 500 using 4 ML algorithms, these results are created with a fixed threshold of 0.7-quantile. Overall, RF shows a bad model fitting and SVM Radial is more appropriate fitted. In Section 4.2, we perform a threshold sensitivity test and investigate whether the model with a different threshold is better fitted. Figure 9 displays the Q-Q plots for FTSE 100 using 4 models and this set of results are based on a 0.75-quantile. In general, we observe a better fitted quantile line in all methods.

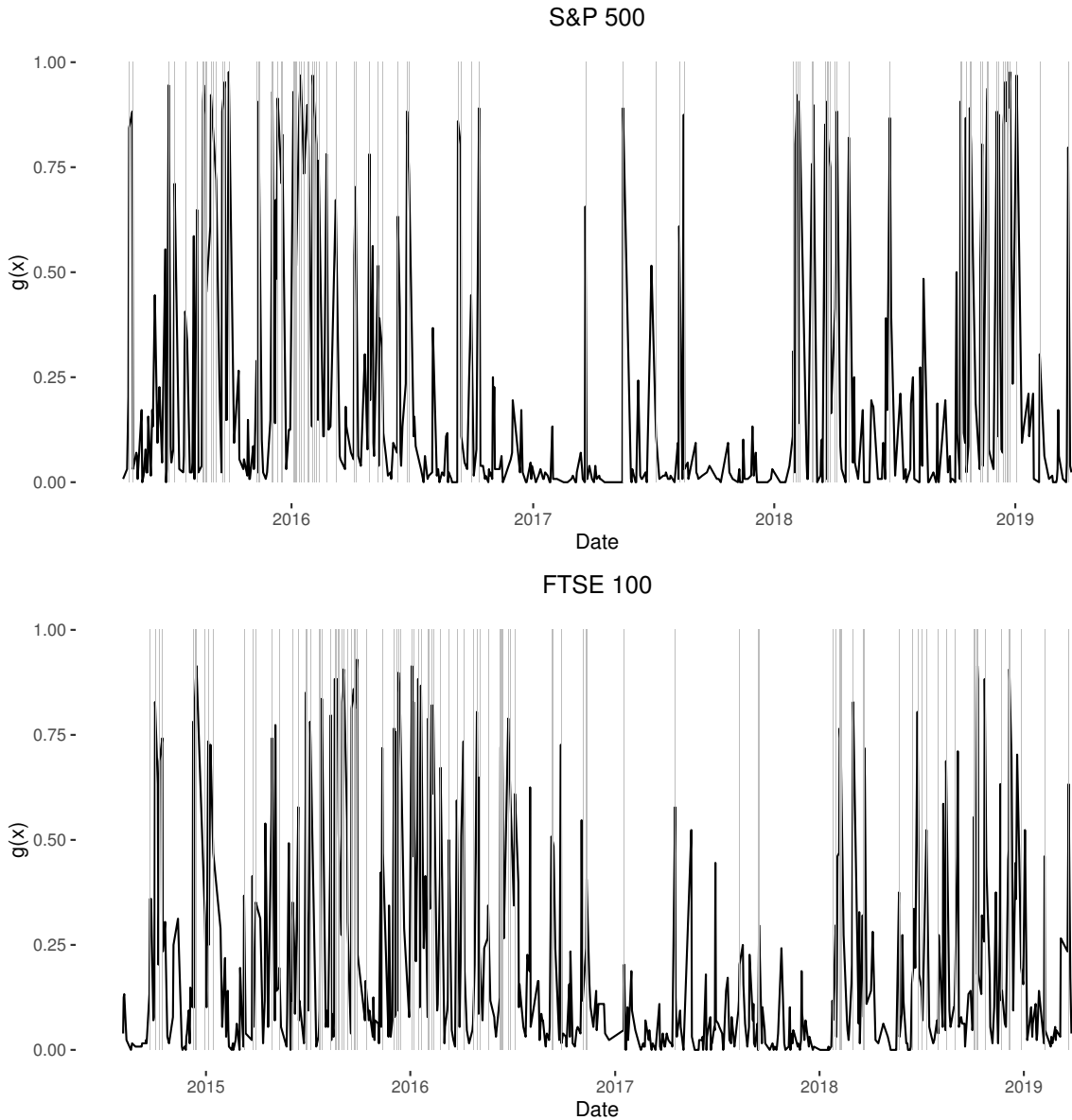


Figure 5: The predicted probabilities of $g(x)$ over time using Random Forests for S&P 500 in the upper graph and FTSE 100 in the bottom graph. The black line presents the predicted $g(x)$ over time which lies between 0 and 1. The background is shaded with grey when there is exceed over threshold u .

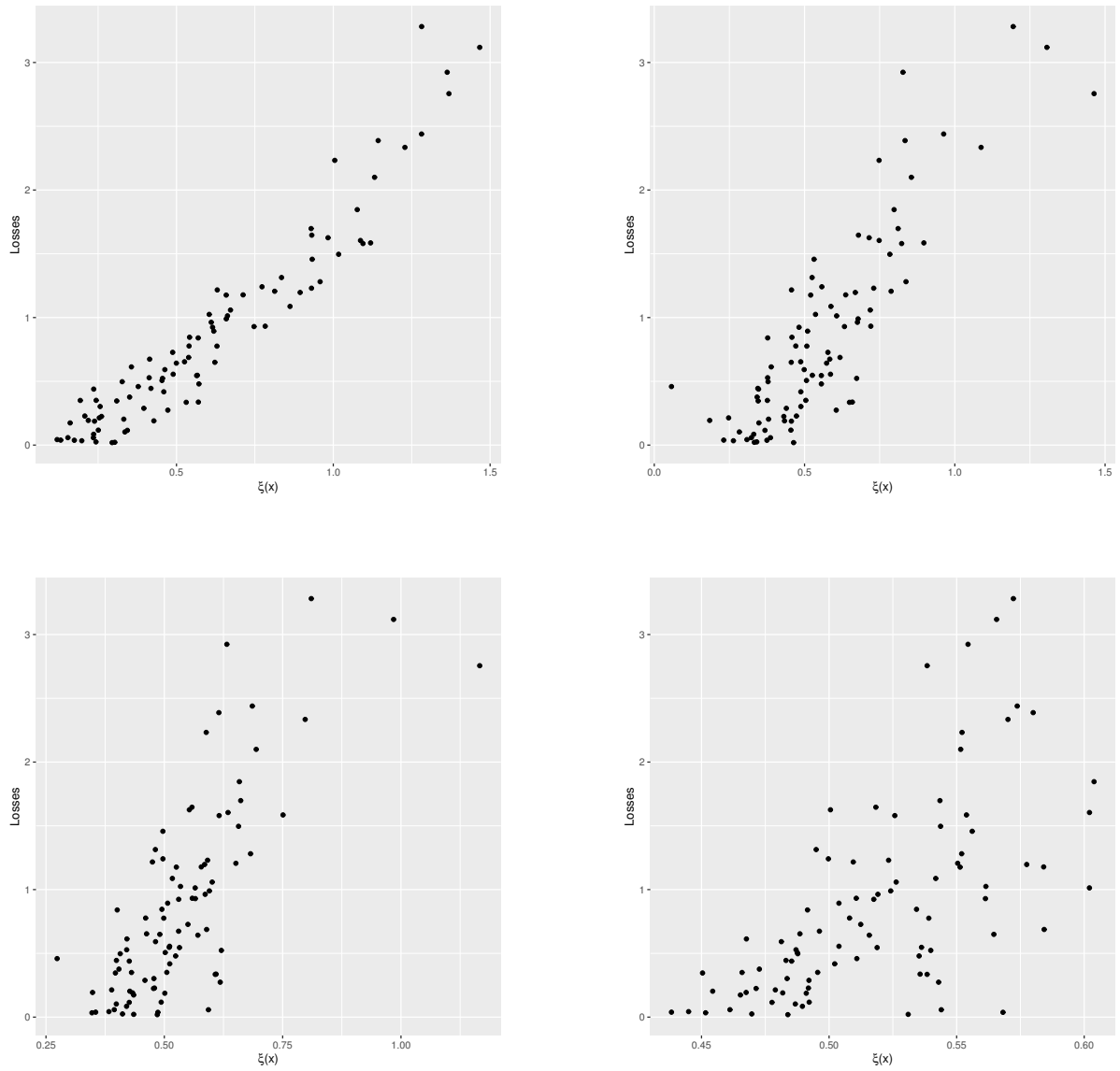


Figure 6: Scatter plots which shows the relationship between the predicted shape parameter and the S&P 500 losses. The x-axis represents the predicted $\xi(x)$ depends on multiple covariates and the y-axis shows the losses. The results for four models are displayed. From top to bottom and left to right: Random Forest, SVM Linear, SVM Polynomial and SVM Radial basis.

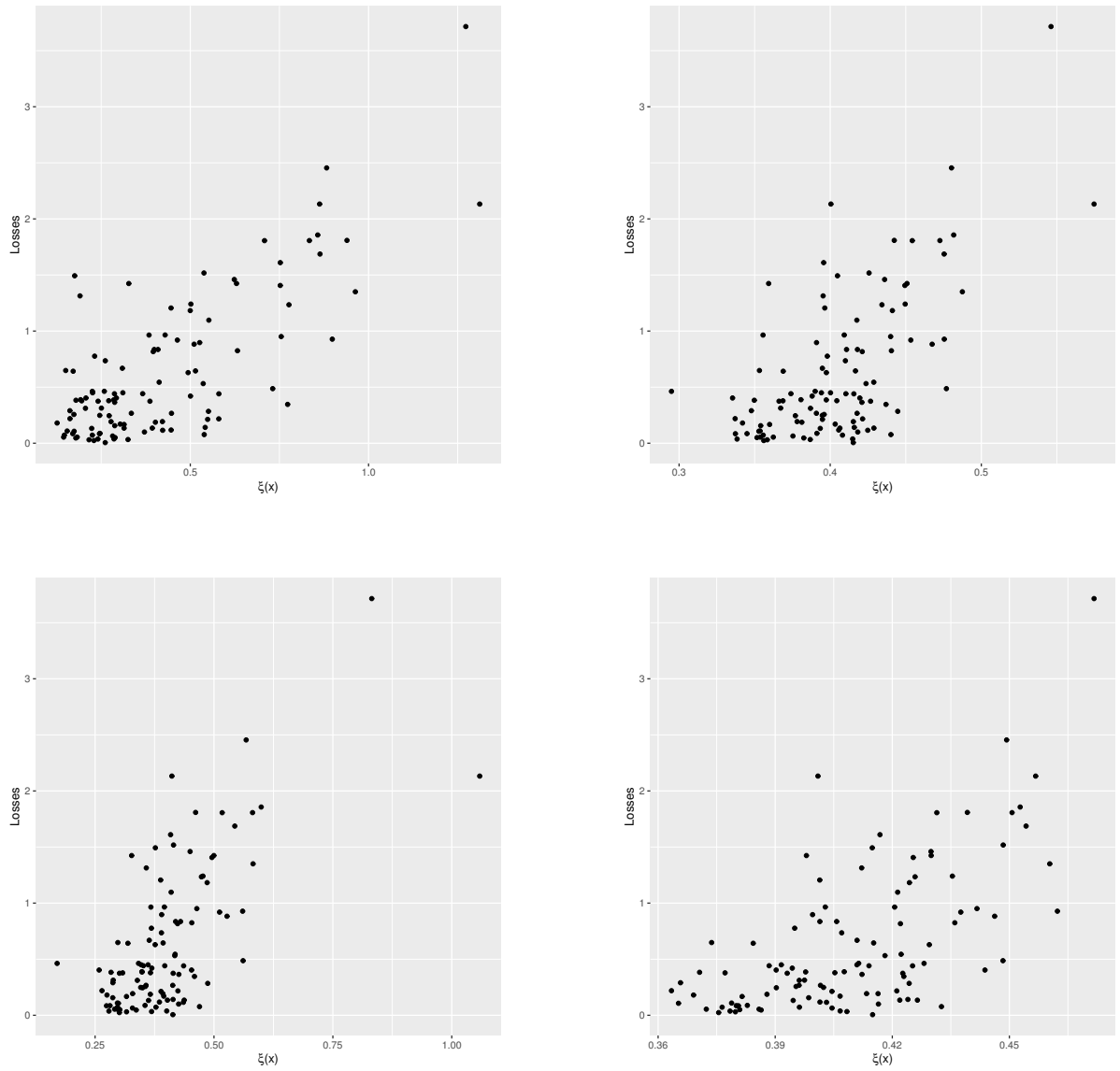
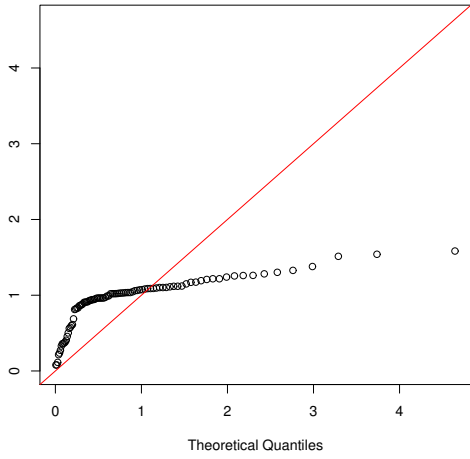
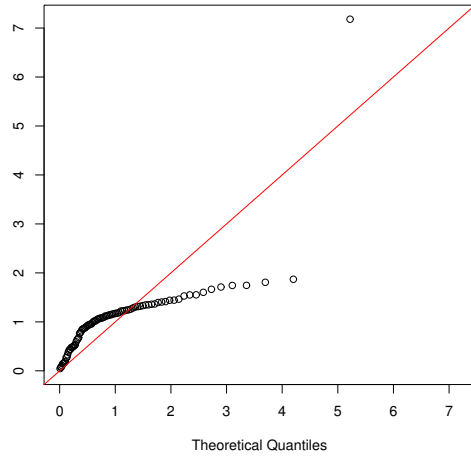


Figure 7: Scatter plots which shows the relationship between the predicted shape parameter and the FTSE 100 losses. The x-axis represents the predicted $\xi(x)$ depends on multiple covariates and the y-axis shows the losses. The results are shown for four models. From top to bottom and left to right: Random Forest, SVM Linear, SVM Polynomial and SVM Radial basis.

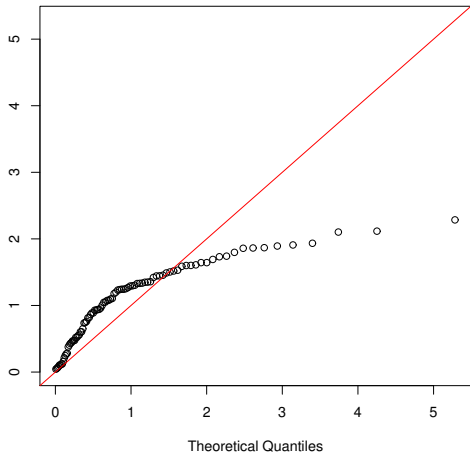
Q-Q Plot of sample data against Exponential Distribution



Q-Q Plot of sample data against Exponential Distribution



Q-Q Plot of sample data against Exponential Distribution



Q-Q Plot of sample data against Exponential Distribution

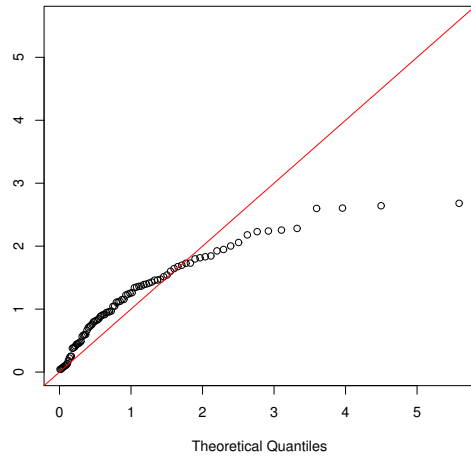


Figure 8: Q-Q Plots to check whether large S&P 500 losses are approximately Pareto distributed. The x-axis plots the quantiles of the standard exponential distribution and the y-axis plots the quantiles of the generated variable L defined in Equation (2). Note that when the models are correctly fitted, a 45-degree straight line should be present in the Q-Q Plots. The results are shown for four models. From top to bottom and left to right: Random Forest, SVM Linear, SVM Polynomial and SVM Radial basis.

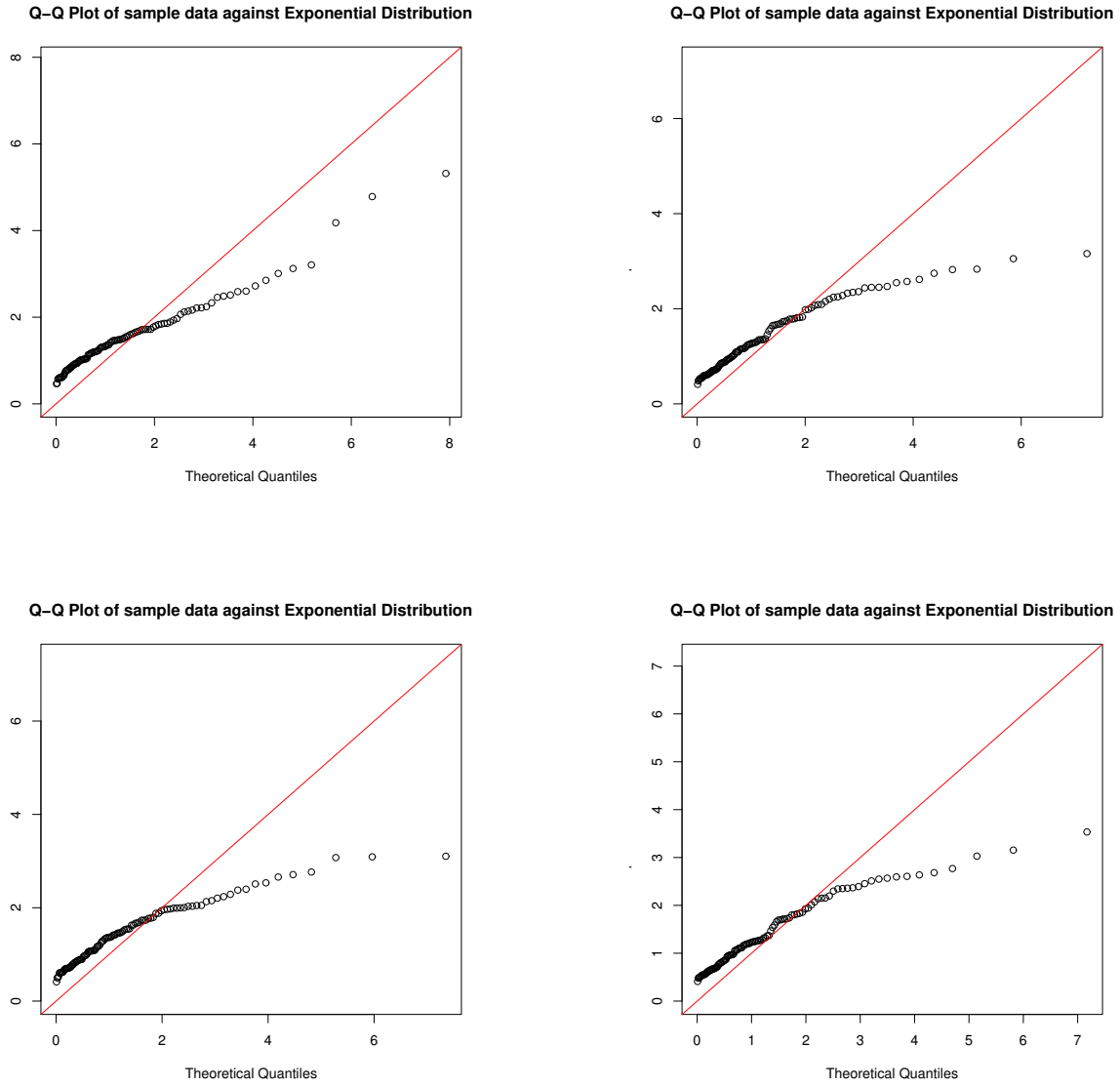


Figure 9: Q-Q Plots to check whether large FTSE 100 losses are approximately Pareto distributed. The x-axis plots the quantiles of the standard exponential distribution and the y-axis plots the quantiles of the generated variable L defined in Equation (2). Note that when the models are correctly fitted, a 45-degree straight line should be present in the Q-Q Plots. The results are shown for four models. From top to bottom and left to right: Random Forest, SVM Linear, SVM Polynomial and SVM Radial basis.

4.1.2 Variable importance

In this subsection, we rank the variable importance to explain the left tail behavior for S&P 500 and FTSE 100 using ML algorithms. The calculation methods to determine the variable importance are different for RF and SVM. First, we explain these techniques separately. Second, we detect the most informative covariates for the estimation on $g(x)$ and $\xi(x)$ with both algorithms.

The variable importance using RF algorithm can be determined in two ways: importance based on accuracy or importance based on impurity. First, variable importance based on accuracy measures how much the accuracy of the model decreases when a variable is excluded. In case of classification, variable importance is measured by calculating the percentage of observations that are incorrectly classified in % when excluding a given variable in the model. In case of regression, variable importance measures the increases of Mean Squared Errors (MSE) in % when a variable is excluded from the model. Therefore, a higher score indicates a more important variable. Second, variable importance based on impurity measures the homogeneity of the dataset by node splitting of a variable. In case of classification, variable importance measures the homogeneity with decrease of Gini Impurity in % when splitting node of a variable. Gini Impurity defines as the probability of labeling an observation incorrectly. In case of regression, variable importance measures the homogeneity with increase of Sum of Squared Errors (SSE) in % when splitting node of a variable. Therefore, a higher value indicates a more influential variable. The variable importance using both methods with RF can be automatically computed by function `varImpPlot()` in R.

The variable importance using SVM algorithm is calculated in a different way. [Guyon and Elisseeff \(2003\)](#) suggest that the weight of each variable in the SVM model can be used to interpret the variable importance. We determine the importance in the following two steps. First, we calculate the weighted sum of all variables by multiplying the estimated coefficients and the support vectors computed by SVM. These results are the output value of the SVM algorithm described in Algorithm 2. Second, we take the absolute value of the weighted sum vectors of all variables in the previous step and rank them in descending order.

Notice that the above-mentioned method is only applied for one model of unique datasets. Since we use a moving window of 1500 days and each forecast is based on a different sample data, we need to compare the covariates importance with same dependent variable but different datasets. However, we find that the variation of covariates importance across different sample period is not much and the top three variable importance is stable over different runs.

Now we can determine the variable importance for the estimation on $g(x)$ and $\xi(x)$ with both RF and SVM algorithms.

The function $g(x)$ predicts the probability of excess over a fixed threshold u given a sets of covariates are known. Appendix B.1 plots a list of variable importance for S&P 500 using RF, SVM Linear, SVM Polynomial and SVM Radial. Equivalently, Appendix B.2 illustrates a list of variable importance for FTSE 100 using RF, SVM Linear, SVM Polynomial and SVM Radial. For S&P 500, DJA, NASDAQ and VIX are the top three predictive variables for the estimation of $g(x)$ in most models except for SVM Linear. We find that the covariate VIX is the 10th important variable with SVM Linear model and the 3rd informative covariate is 12-month LIBOR in US. Dollar. This result is consistent with our expectations. As mentioned before in the data section, the major world indices move simultaneously with each other, specially in an extreme case. Furthermore, Giot (2005) reveals a negative relationship between the implied volatility index VIX and returns of S&P 500. For FTSE 100, CAC40, AEX and DAX are the top three influential variables on the $g(x)$ estimates in most methods except for SVM Linear. The 10-day moving average of FTSE 100 also predicts the excess probability when using SVM with linear kernel.

The shape parameter $\xi(x)$ plays a dominant role in the tail-fitting procedure and they determine mostly the shape of the tail distribution. Appendix B.3 represents the variable importance plot for S&P 500 using RF, SVM Linear, SVM Polynomial and SVM Radial. We find that DJA, NASDAQ and VIX are the most three informative covariates for almost all methods. AEX is the 3th significant important variable using RF measured

with increase of SSE in % by node splitting of a variable, the covariate VIX is the 6th influential covariate. Note that the top three important covariates which predicts $g(x)$ and $\xi(x)$ are the exactly the same. We conclude that they explain the tail behavior for S&P 500. Similarly, Appendix B.4 shows the variable importance plot for FTSE 100 using RF, SVM Linear, SVM Polynomial and SVM Radial. CAC40, AEX and DAX are the three significant covariates to describe the shape of tail distribution for FTSE 100. Again, these covariates are the same as for estimating $g(x)$.

4.1.3 Individual covariate

From the previous section, we discover the covariates with most predictive ability for $g(x)$ and $\xi(x)$. In this section, we examine the relation between the individual covariate and predicted parameters in the form of scatter plots. It provides us insight how an individual covariate affects the outcome of response variable in ML algorithms.

Figure 10 shows the scatter plots of the top three most informative covariates for S&P 500. Overall, we observe a positive relation between covariates Dow Jones, NASDAQ and predicted $g(x)$. For example, we examine a monotonically increasing $\hat{g}(x)$ when the negative log-returns of Dow Jones lie between $0.50\%(= \exp(-0.005) - 1)$ and $1.49\%(= \exp(-0.015) - 1)$, which indicates losses in Dow Jones Index leads to higher probability of S&P 500 crash. When the negative log-returns of Dow Jones become greater than 1.49% , the probabilities approach to 1. The relationship between individual covariate NASDAQ and predicted $g(x)$ can be interpreted in a similar way. Additionally, we find a linear positive relationship between covariates DJIA, NASDAQ and $\hat{\xi}(x)$. The predicted shape parameter increases when the negative log-returns of DJIA and NASDAQ increases, the function $\hat{\xi}(x)$ depends on covariates does not converge. For covariate VIX, we detect a negative relation with $\hat{g}(x)$ and $\hat{\xi}(x)$. If the positive log-returns of fear index VIX get larger, which means there is more risk exists and investors have less confidence on the financial market, the probabilities of exceedances over threshold raise up as well. However, we examine that there is no clear association between VIX and $\hat{\xi}(x)$.

Equivalently, Figure 11 displays the scatter plots of top three informative covariates for FTSE 100 versus $\hat{g}(x)$ and $\hat{\xi}(x)$. Overall, we observe a positive relation between covariates CAC40, AEX, DAX and predicted probability $\hat{g}(x)$. Additionally, there is a monotonic non-linear relationship among CAC40, AEX, DAX and the predicted shape parameter $\hat{\xi}(x)$.

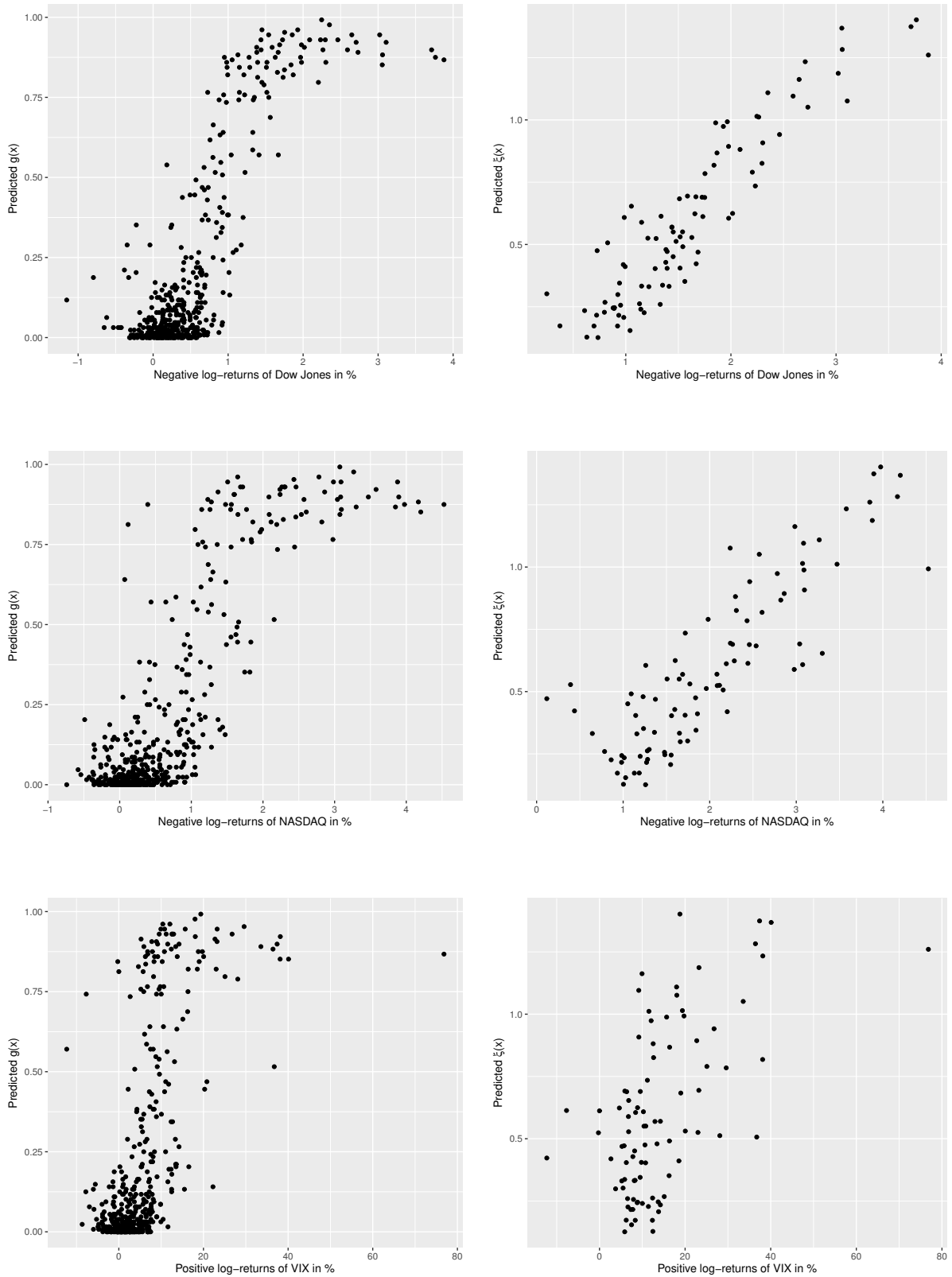


Figure 10: Scatter plots which shows the relationship between the top three covariates (DJA, NASDAQ, VIX) for S&P 500 and the predicted parameters. Left graphs display the scatter plots for $\hat{g}(x)$ and right graphs display the scatter plots for $\hat{\xi}(x)$. The x-axis represents the log-returns of an individual covariate in % and y-axis shows the predicted parameters.

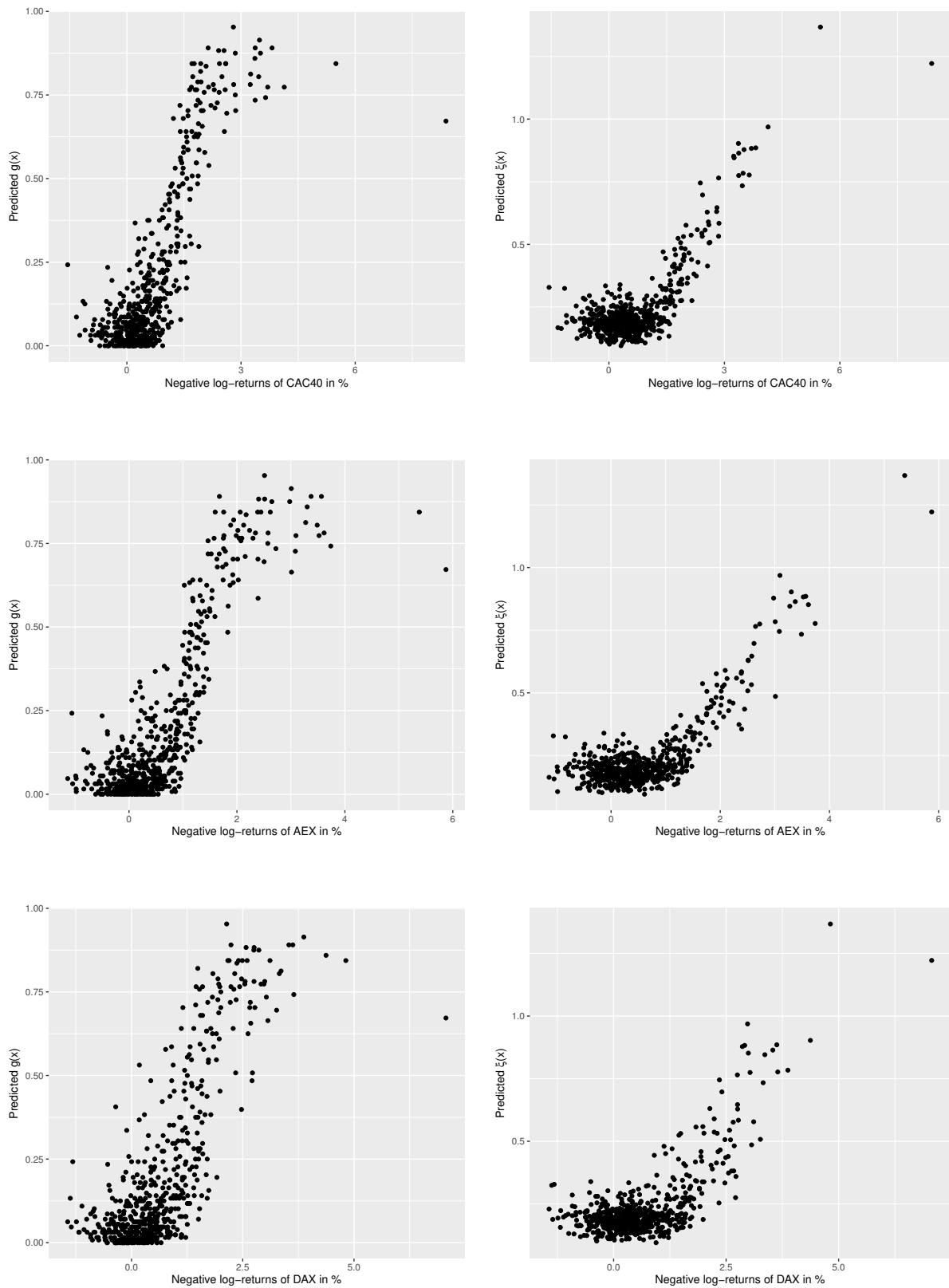


Figure 11: Scatter plots which shows the relationship between the top three covariates (CAC40, AEX, DAX) for FTSE 100 and the predicted parameters. Left graphs display the scatter plots for $\hat{g}(x)$ and right graphs display the scatter plots for $\hat{\xi}(x)$. The x-axis represents the log-returns of an individual covariate in % and y-axis shows the predicted parameters.

4.2 Choice of threshold u

In this section, we test a sets of threshold and examine the effect of model fitting with Q-Q plots. Based on the results for S&P 500 in Figure 8, we use a threshold of 0.65, 0.75-quantile of the sample data and compare the model performance with pre-selected threshold which is 0.70-quantile of the sample data. Since SVM Radial has the best model fitting in the Q-Q plot, we perform a threshold sensitivity test with this model. The results are illustrated in Figure 12. We find the difference is not much across various thresholds. For FTSE 100, we try a threshold of 0.70, 0.80-quantile using SVM Linear model and we compare the results with 0.75-quantile threshold. The Q-Q plots are shown in Figure 13. Overall, we find that we select an appropriate threshold and other threshold has rarely impact on the model performance. In our further analysis, we use a fixed threshold of 0.70-quantile for S&P 500 and 0.75-quantile for FTSE 100.

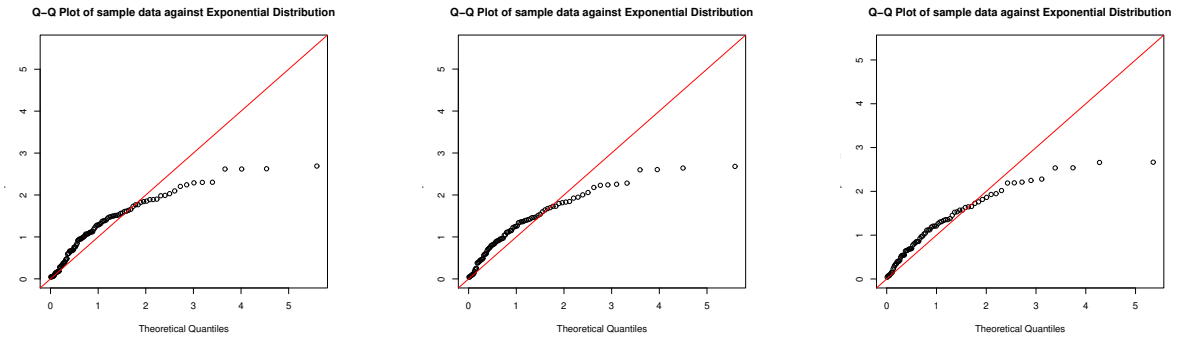


Figure 12: Q-Q Plots with 0.65, 0.70, 0.75-quantile threshold using SVM Radial for S&P 500

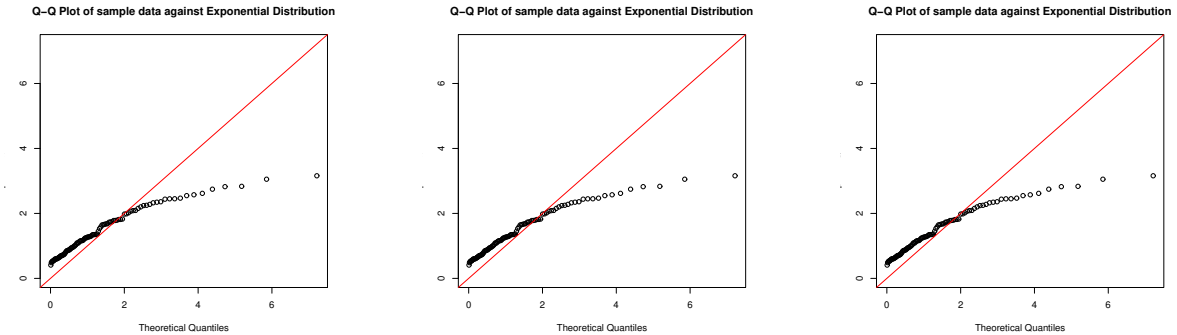


Figure 13: Q-Q Plots with 0.70, 0.75, 0.80-quantile threshold using SVM Linear for FTSE 100

4.3 VaR Forecasting

In this section, we execute an out-of-sample VaR forecasting with 4 proposed models depending on covariates and 2 classical risk models. For each model, we predict the VaR based on a moving window of 1500 days and use a confidence level of 0.90, 0.95, 0.99. The out-of-sample period for S&P 500 and FTSE 100 contains 476 and 571 trading days, respectively. We compare the performance of VaR Forecasting across different models in terms of Binomial test and Conditional Coverage Independence (CCI) test invent by [Christoffersen \(1998\)](#).

Table 3 shows the results of out-of-sample 90%, 95%, 99% VaR forecasting for S&P 500 using RF, SVM Linear, SVM Polynomial, SVM Radial, HS and GARCH(1.1) models. We use a significance level of 5% for both evaluation tests and reject the null hypothesis when the p-value is smaller than 5%. Overall, both the binomial test and CCI test by [Christoffersen \(1998\)](#) show a consistent conclusion. We reject all models for both 90% and 95% VaR levels. For 99% VaR level, we accept the SVM Linear, SVM Polynomial and GARCH(1.1) models. The models show poorer performance than expected. However, we examine that the observed number of violations is lower than the expected number of violations for all models except RF. For example, we expect VaR violations of 1%, 5%, 10% for 99%, 95%, 90% VaR levels, respectively. A lower percentage of VaR violations than expected indicates that the models overestimate risk. For the regulatory purpose, overestimating risk is better than underestimating risk since they are more conservative. Additionally, we observe that the average VaR for ML models are low without having more violation than expected. This results is desirable as banks reporting such VaR estimates would comply with regulation with having a lower capital requirements.

Table 4 shows the results of out-of-sample 90%, 95%, 99% VaR forecasting for FTSE 100 using RF, SVM Linear, SVM Polynomial, SVM Radial, HS and GARCH(1.1) models. Again, both evaluation test leads to the same conclusion. Besides, the models perform better than for S&P 500. For 90% VaR level, we accept RF, SVM Linear and SVM Radial models. For 95% VaR level, only SVM Linear and SVM Radial are accepted. For 99% VaR, we accept SVM Linear, SVM Polynomial and HS models. We observe a similar results as for S&P 500, the average VaR estimates are low while the observed number

of violations is smaller than expected. For illustration purpose, we plot the 90% VaR Forecasting with RF, SVM Linear, SVM Polynomial, SVM Radial, Historical Simulation and Dynamic HS GARCH(1,1) for S&P 500 and FTSE 100 in Appendix B.5.

Finally, as a robustness check, we perform an out-of-sample VaR forecasting with top three covariates which is found in Section 4.1.2. Appendix A.2 shows the results of out-of-sample VaR forecasting with top three covariates (DJA, NASDAQ, VIX) for S&P 500 using RF, SVM Linear, SVM Polynomial, SVM Radial. We accept SVM Radial for 90% VaR level and SVM Polynomial model for 95% VaR with both evaluation tests. For 99% VaR level, all models are rejected. When we compare the results with Table 3 which all 63 covariates are obtained for the VaR forecasting, the models using only top three covariates perform better with 90%, 95% VaR level, but worse with 99% VaR level.

Equivalently, Appendix A.3 shows the results of out-of-sample VaR forecasting with top three covariates (CAC40,AEX,DAX) for FTSE 100 using RF, SVM Linear, SVM Polynomial, SVM Radial. For 90% and 95% VaR level, we accept both SVM Linear and SVM Polynomial models for binomial and Christoffersen (1998)'s test. For 99% VaR level, only SVM Radial model is accepted. When we compare the results obtained from Table 4, we do not observe that the models using top three covariates perform better than using 63 covariates for VaR forecasting. Overall, we observe that the ML models with top three covariates only does not perform better for VaR forecasting than using 63 covariates.

Table 3: Performance Evaluation S&P 500

| a) $\alpha = 0.90$ | | | | | |
|-----------------------|-------------|------------------|--------------------|-----------------------|--------------------|
| Model | Average VaR | Binomial Test | | CCI by Christoffersen | |
| | | # Violations (%) | P-value (Decision) | Test statistics | P-value (Decision) |
| RF | 1.412 | 75 (15.756) | 0.000 (Reject) | 65.839 | 0.000 (Reject) |
| SVM Linear | 1.316 | 13 (2.731) | 0.000 (Reject) | 38.852 | 0.000 (Reject) |
| SVM Polynomial | 1.258 | 11 (2.311) | 0.000 (Reject) | 45.364 | 0.000 (Reject) |
| SVM Radial Basis | 1.100 | 32 (6.723) | 0.014 (Reject) | 7.845 | 0.020 (Reject) |
| Historial Simulation | 1.900 | 30 (6.303) | 0.006 (Reject) | 18.341 | 0.000 (Reject) |
| Dynamic HS GARCH(1,1) | 2.235 | 15 (3.151) | 0.000 (Reject) | 39.948 | 0.000 (Reject) |
| b) $\alpha = 0.95$ | | | | | |
| Model | Average VaR | Binomial Test | | CCI by Christoffersen | |
| | | # Violations (%) | P-value (Decision) | Test statistics | P-value (Decision) |
| RF | 2.267 | 72 (15.126) | 0.000 (Reject) | 116.932 | 0.000 (Reject) |
| SVM Linear | 1.913 | 2 (0.420) | 0.000 (Reject) | 34.745 | 0.000 (Reject) |
| SVM Polynomial | 1.734 | 2 (0.420) | 0.000 (Reject) | 34.745 | 0.000 (Reject) |
| SVM Radial Basis | 1.553 | 8 (1.681) | 0.000 (Reject) | 14.976 | 0.001 (Reject) |
| Historial Simulation | 2.606 | 11 (2.311) | 0.004 (Reject) | 10.331 | 0.006 (Reject) |
| Dynamic HS GARCH(1,1) | 2.694 | 9 (1.891) | 0.001 (Reject) | 14.616 | 0.001 (Reject) |
| c) $\alpha = 0.99$ | | | | | |
| Model | Average VaR | Binomial Test | | CCI by Christoffersen | |
| | | # Violations (%) | P-value (Decision) | Test statistics | P-value (Decision) |
| RF | 10.032 | 72 (15.126) | 0.000 (Reject) | 315.367 | 0.000 (Reject) |
| SVM Linear | 6.577 | 1 (0.210) | 0.101 (Accept) | 4.434 | 0.109 (Accept) |
| SVM Polynomial | 3.990 | 2 (0.420) | 0.348 (Accept) | 2.085 | 0.353 (Accept) |
| SVM Radial Basis | 3.469 | 0 (0.000) | 0.018 (Reject) | 9.568 | 0.008 (Reject) |
| Historial Simulation | 4.828 | 0 (0.000) | 0.018 (Reject) | 9.568 | 0.008 (Reject) |
| Dynamic HS GARCH(1,1) | 3.560 | 2 (0.420) | 0.348 (Accept) | 2.085 | 0.353 (Accept) |

Note: Out-of-sample VaR forecasting for S&P 500. This table shows from column left to right: the average VaR over a testing period, the number of violations and their percentage is expressed in parentheses, then a two-sided binomial test is executed with null hypothesis of correct number of expected violations and we calculate the p-value, followed by Christoffersen (1998)'s jointly test with null hypothesis of independent violations and the test statistics with p-value is computed. We reject the null hypothesis when the p-value is smaller than 5% significance level and the decision results are expressed in parentheses.

Table 4: Performance Evaluation FTSE 100

| a) $\alpha = 0.90$ | | | | | |
|-----------------------|-------------|------------------|--------------------|-----------------------|--------------------|
| Model | Average VaR | Binomial Test | | CCI by Christoffersen | |
| | | # Violations (%) | P-value (Decision) | Test statistics | P-value (Decision) |
| RF | 1.333 | 71 (12.434) | 0.059 (Accept) | 4.167 | 0.124 (Accept) |
| SVM Linear | 1.114 | 59 (10.333) | 0.780 (Accept) | 0.072 | 0.965 (Accept) |
| SVM Polynomial | 1.255 | 41 (7.180) | 0.025 (Reject) | 6.949 | 0.031 (Reject) |
| SVM Radial Basis | 1.184 | 59 (10.333) | 0.780 (Accept) | 0.226 | 0.893 (Accept) |
| Historial Simulation | 1.838 | 35 (6.130) | 0.001 (Reject) | 11.221 | 0.004 (Reject) |
| Dynamic HS GARCH(1,1) | 2.194 | 18 (3.152) | 0.000 (Reject) | 39.831 | 0.000 (Reject) |

| b) $\alpha = 0.95$ | | | | | |
|-----------------------|-------------|------------------|--------------------|-----------------------|--------------------|
| Model | Average VaR | Binomial Test | | CCI by Christoffersen | |
| | | # Violations (%) | P-value (Decision) | Test statistics | P-value (Decision) |
| RF | 1.772 | 58 (10.158) | 0.000 (Reject) | 25.799 | 0.000 (Reject) |
| SVM Linear | 1.452 | 19 (3.327) | 0.068 (Accept) | 5.871 | 0.053 (Accept) |
| SVM Polynomial | 1.611 | 16 (2.802) | 0.012 (Reject) | 14.049 | 0.001 (Reject) |
| SVM Radial Basis | 1.559 | 19 (3.327) | 0.068 (Accept) | 5.871 | 0.053 (Accept) |
| Historial Simulation | 2.516 | 14 (2.452) | 0.004 (Reject) | 13.688 | 0.001 (Reject) |
| Dynamic HS GARCH(1,1) | 2.612 | 12 (2.102) | 0.001 (Reject) | 14.153 | 0.001 (Reject) |

| c) $\alpha = 0.99$ | | | | | |
|-----------------------|-------------|------------------|--------------------|-----------------------|--------------------|
| Model | Average VaR | Binomial Test | | CCI by Christoffersen | |
| | | # Violations (%) | P-value (Decision) | Test statistics | P-value (Decision) |
| RF | 4.402 | 41 (7.180) | 0.000 (Reject) | 100.643 | 0.000 (Reject) |
| SVM Linear | 2.698 | 1 (0.175) | 0.053 (Accept) | 5.978 | 0.050 (Accept) |
| SVM Polynomial | 3.030 | 1 (0.175) | 0.053 (Accept) | 5.978 | 0.050 (Accept) |
| SVM Radial Basis | 2.952 | 0 (0.000) | 0.005 (Reject) | 11.477 | 0.003 (Reject) |
| Historial Simulation | 4.617 | 1 (0.175) | 0.053 (Accept) | 5.978 | 0.050 (Accept) |
| Dynamic HS GARCH(1,1) | 3.574 | 0 (0.000) | 0.005 (Reject) | 11.477 | 0.003 (Reject) |

Note: Out-of-sample VaR forecasting for FTSE 100. This table shows from column left to right: the average VaR over a testing period, the number of violations and their percentage is expressed in parentheses, then a two-sided binomial test is executed with null hypothesis of correct number of expected violations and we calculate the p-value, followed by Christoffersen (1998)'s jointly test with null hypothesis of independent violations and the test statistics with p-value is computed. We reject the null hypothesis when the p-value is smaller than 5% significance level and the decision results are expressed in parentheses.

5 Conclusion

In this study, we implement the Machine Learning techniques into parameters estimation procedure in Extreme Value Analysis. We derive parameters depend on multiple covariates to describe the tail of the Pareto distribution. The informative covariates to explain the left tail behavior are automatically chosen by Random Forests and Support Vector Machines algorithms. We apply the methodology on two empirical datasets: world major indices S&P 500 and FTSE 100. We conclude this paper by answering the three main research questions.

First, we identify covariates that explains two major parameters $g(x)$ and $\xi(x)$ describing the tail of a Pareto distribution. We find that DJA, NASDAQ and VIX are the most informative covariates to explain tail behavior for S&P 500. More specifically, DJA and NASDAQ have positive relation with the tails and fear index VIX is negatively correlated with the tails. Correspondingly, three European stock indices CAC40, AEX and DAX are the influential covariates for losses of FTSE100. Furthermore, the covariates are positively related with the financial losses of FTSE 100.

Second, we compare the out-of-sample VaR forecasting between our proposed models depends on covariates and two classical risk models. We find that our implemented models which combining ML and EVT perform better for 99% VaR forecasting than 90%, 95% VaR level. Due to the fact that EVT is superior only for the estimation of tail probabilities or high quantiles, the 99% VaR forecasting is accepted by the evaluation tests for both S&P 500 and FTSE 100 datasets but not for 90%, 95% VaR levels. Additionally, we find that our proposed models perform better when applied on empirical dataset FTSE 100 than S&P 500. A possible explanation is the choice of threshold for S&P 500. We observe before from the Q-Q plots that the S&P 500 losses are not very good approximated by the Pareto distribution with a threshold of 0.70-quantile, while the FTSE 100 losses are better fitted in the model. Therefore, the 90%, 95% VaR forecasting using ML models for S&P 500 are both rejected by the null hypothesis, while few ML models (RF, SVM Linear, SVM Radial) are accepted by the null hypothesis for FTSE 100. Overall, we do not find evidence to show that our proposed models outperform the Historical

Simulation and GARCH(1,1) risk models. However, the ML models produce low VaR estimates without having more violations than expected. This is favorable results since low VaR estimates can help banks to hold capital more efficiently.

Third, we find that varying from other candidate thresholds have rarely impact on the model performance.

For this research, we work out with two basic Random Forests and Support Vector Machines algorithms. One possible extension is to use Deep Learning techniques such as Artificial Neural Networks, which allows to model the complex relation between the covariates and parameter estimates. Additionally, research on a longer sample period or other major world indices could be interesting.

References

- Ajayi, R. A., Friedman, J., and Mehdian, S. M. (1998). On the relationship between stock returns and exchange rates: tests of granger causality. *Global finance journal*, 9(2):241–251.
- Balkema, A. A. and de Haan, L. (1974). Residual life time at great age. *Ann. Probab.*, 2(5):792–804.
- Baur, D. G. and McDermott, T. K. (2010). Is gold a safe haven? international evidence. *Journal of Banking & Finance*, 34(8):1886–1898.
- Brock, W., Lakonishok, J., and LeBaron, B. (1992). Simple technical trading rules and the stochastic properties of stock returns. *Journal of Finance*, 47(5):1731–1764.
- Campbell, J. Y. (1987). Stock returns and the term structure. *Journal of Financial Economics*, 18(2):373–399.
- Campbell, J. Y., Grossman, S. J., and Wang, J. (1993). Trading volume and serial correlation in stock returns. *Quarterly Journal of Economics*, 108(4):905–939.
- Chavez-Demoulin, V., Embrechts, P., and Hofert, M. (2015). An extreme value approach for modeling operational risk losses depending on covariates. *Journal of Risk and Insurance*, 83(3):735–776.
- Christoffersen, P. F. (1998). Evaluating interval forecasts. *International Economic Review*, 39(4):841–862.
- Coles, S., Bawa, J., Trenner, L., and Dorazio, P. (2001). *An introduction to statistical modeling of extreme values*. Springer.
- Domian, D. L. and Reichenstein, W. (1998). Term spreads and predictions of bond and stock excess returns. *Financial Services Review*, 7(1):25–44.
- Dowd, K. and Blake, D. (2006). After VaR: The Theory, Estimation, and Insurance Applications of Quantile-Based Risk Measures. *Journal of Risk and Insurance*, 73(2):193–229.
- Embrechts, P., Kluppelberg, C., and Mikosch, T. (1997). *Modelling Extremal Events- For Insurance and Finance*. Springer.
- Embrechts, P., Mizgier, K., and Chen, X. (2018). Modeling operational risk depending on covariates. an empirical investigation. *Journal of Operational Risk*, 13:17–46.
- Fama, E. F. (1990). Stock returns, expected returns, and real activity. *Journal of*

- Finance*, 45(4):1089–1108.
- Flannery, M. J. and James, C. M. (1984). The effect of interest rate changes on the common stock returns of financial institutions. *Journal of Finance*, 39(4):1141–1153.
- Gilli, M. and K ellezi, E. (2006). An application of extreme value theory for measuring financial risk. *Computational Economics*, 27:207–228.
- Giot, P. (2005). Relationships between implied volatility indexes and stock index returns. *Journal of Portfolio Management*, 31(3):92–100.
- Gupta, R., Risse, M., Volkman, D. A., and Wohar, M. E. (2019). The role of term spread and pattern changes in predicting stock returns and volatility of the united kingdom: Evidence from a nonparametric causality-in-quantiles test using over 250 years of data. *North American Journal of Economics and Finance*, 47:391–405.
- Guyon, I. and Elisseeff, A. (2003). An introduction to variable and feature selection. *Journal of Machine Learning Research*, 3(Mar):1157–1182.
- Hambuckers, J., Groll, A., and Kneib, T. (2018). Understanding the economic determinants of the severity of operational losses: A regularized generalized Pareto regression approach. *Journal of Applied Econometrics*, 33(6):898–935.
- Hilliard, J. E. (1979). The relationship between equity indices on world exchanges. *Journal of Finance*, 34(1):103–114.
- Jorion, P. et al. (2010). *Financial Risk Manager Handbook: FRM Part I*, volume 625. John Wiley & Sons.
- Ma, C. K. and Kao, G. W. (1990). On exchange rate changes and stock price reactions. *Journal of Business Finance & Accounting*, 17(3):441–449.
- Mayr, A., Fenske, N., Hofner, B., Kneib, T., and Schmid, M. (2012). Generalized additive models for location, scale and shape for high dimensional data—a flexible approach based on boosting. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 61(3):403–427.
- Oshiro, T. M., Perez, P. S., and Baranauskas, J. A. (2012). How many trees in a random forest? In *International workshop on machine learning and data mining in pattern recognition*, pages 154–168. Springer.
- Rigby, R. and Stasinopoulos, D. (2005). Generalized additive models for location, scale and shape (with discussion). *Journal of the Royal Statistical Society: Series C*

(Applied Statistics), 54:507 – 554.

Sadorsky, P. (1999). Oil price shocks and stock market activity. *Energy economics*, 21(5):449–469.

Zhong, X. and Enke, D. (2017). Forecasting daily stock market return using dimensionality reduction. *Expert Systems with Applications*, 67:126–139.

A Tables

A.1 Summary statistics of covariates

Table 5: Summary statistics of covariates

| Name | Description | Mean | Min. | Max. | Std. Dev. | Skewness | Kurtosis | Jarque-Bera(Pr.) |
|--|---|--------|---------|--------|-----------|----------|----------|------------------|
| <i>World major indices</i> | | | | | | | | |
| S&P500 (Excluded when consider the U.S. Markets) | Log returns of S&P 500 in % | 0.019 | -9.470 | 10.957 | 1.165 | -0.209 | 10.058 | 19095.721(0.000) |
| FTSE100 (Excluded when consider the UK Markets) | Log returns of FTSE 100 in % | 0.006 | -9.266 | 9.384 | 1.152 | -0.136 | 7.210 | 9809.590(0.000) |
| DJA | Log returns of Dow Jones Industrial Average in % | 0.022 | -8.695 | 10.089 | 1.109 | -0.163 | 7.843 | 11612.491(0.000) |
| NASDAQ | Log returns of NASDAQ in % | 0.029 | -9.588 | 11.159 | 1.345 | -0.095 | 5.406 | 5514.062(0.000) |
| AEX | Log returns of AEX in % | 0.000 | -9.590 | 10.028 | 1.412 | -0.110 | 6.901 | 8983.750(0.000) |
| DAX | Log returns of DAX in % | 0.015 | -7.442 | 10.797 | 1.460 | -0.048 | 4.883 | 4495.631(0.000) |
| CAC40 | Log returns of CAC40 in % | 0.001 | -9.472 | 10.595 | 1.423 | -0.013 | 5.587 | 5883.911(0.000) |
| NIKKEI225 | Log returns of Nikkei225 in % | 0.013 | -12.111 | 9.494 | 1.425 | -0.645 | 6.458 | 8172.529(0.000) |
| HANGSENG | Log returns of Hang Seng Composite Index in % | 0.019 | -13.582 | 13.407 | 1.388 | -0.013 | 9.750 | 17912.834(0.000) |
| SHANGHAI | Log returns of Shanghai Stock Exchange Composite Index in % | 0.010 | -9.256 | 9.401 | 1.552 | -0.323 | 5.279 | 5330.130 (0.000) |
| <i>Financial variables</i> | | | | | | | | |
| Crude oil | Log returns of crude oil in % | 0.018 | -17.092 | 19.144 | 2.365 | -0.015 | 5.439 | 5574.689(0.000) |
| Gold | Log returns of gold in % | 0.035 | -8.913 | 9.554 | 1.104 | -0.376 | 6.469 | 7992.935(0.000) |
| Silver | Log returns of silver in % | 0.028 | -21.955 | 10.642 | 1.877 | -1.207 | 11.494 | 25989.694(0.000) |
| Platinum | Log returns of platinum in % | 0.013 | -9.734 | 8.761 | 0.863 | -0.167 | 14.676 | 40600.048(0.000) |
| VIX | Log returns of CBOE volatiltiy index VIX in % | -0.009 | -35.059 | 76.825 | 6.856 | 0.949 | 7.387 | 10961.485(0.000) |
| DAAA | Log returns of Moody's Seasoned Aaa Corporate Bond Yield in % | 4.858 | 3.180 | 7.400 | 0.955 | 0.209 | -0.927 | 194.529(0.000) |

Table 5 – Continued on next page

Table 5 – Continued from previous page

| Name | Description | Mean | Min. | Max. | Std. Dev. | Skewness | Kurtosis | Jarque-Bera(Pr.) |
|---------------------------|---|--------|----------|---------|-----------|----------|----------|--------------------|
| DBAA | Log returns of Moody's Seasoned Baa Corporate Bond Yield in % | 5.931 | 4.150 | 9.540 | 1.126 | 0.490 | -0.429 | 215.276(0.000) |
| <u>U.S. Markets</u> | | | | | | | | |
| Trading Volume of S&P500 | Log returns of trading volume of S&P500 | 0.023 | -143.194 | 144.329 | 18.032 | 0.017 | 9.497 | 16995.243(0.000) |
| LIBOR 1MUSD | Log returns of 1-Month London Interbank Offered Rate (LIBOR), based on U.S. Dollar in % | 1.555 | 0.000 | 5.824 | 1.635 | 1.163 | 0.196 | 1025.639(0.000) |
| LIBOR 3MUSD | Log returns of 3-Month London Interbank Offered Rate (LIBOR), based on U.S. Dollar in % | 1.689 | 0.000 | 5.725 | 1.625 | 1.111 | 0.086 | 931.055(0.000) |
| LIBOR 6MUSD | Log returns of 6-Month London Interbank Offered Rate (LIBOR), based on U.S. Dollar in % | 1.846 | 0.000 | 5.640 | 1.580 | 1.061 | -0.002 | 847.315(0.000) |
| LIBOR 12MUSD | Log returns of 12-Month London Interbank Offered Rate (LIBOR), based on U.S. Dollar in % | 2.088 | 0.000 | 5.766 | 1.486 | 0.982 | -0.111 | 729.212(0.000) |
| <u>UK Markets</u> | | | | | | | | |
| Trading Volume of FTSE100 | Log returns of trading volume of FTSE100 | -0.005 | -910.982 | 988.034 | 35.616 | 0.855 | 227.489 | 9749347.787(0.000) |
| LIBOR 1MGBP | Log returns of 1-Month London Interbank Offered Rate (LIBOR), based on British Pound in % | 2.247 | 0.000 | 6.750 | 2.105 | 0.503 | -1.503 | 615.184(0.000) |
| LIBOR 3MGBP | Log returns of 3-Month London Interbank Offered Rate (LIBOR), based on British Pound in % | 2.368 | 0.000 | 6.904 | 2.105 | 0.502 | -1.449 | 585.157(0.000) |
| LIBOR 6MGBP | Log returns of 6-Month London Interbank Offered Rate (LIBOR), based on British Pound in % | 2.493 | 0.000 | 6.799 | 2.056 | 0.493 | -1.438 | 572.007(0.000) |
| LIBOR 12MGBP | Log returns of 12-Month London Interbank Offered Rate (LIBOR), based on British Pound in % | 2.721 | 0.000 | 6.654 | 1.968 | 0.466 | -1.426 | 545.760(0.000) |
| <u>T-bill rates</u> | | | | | | | | |
| TBILL 1M | Log returns of 1-month Treasury constant maturity rate in % | 1.266 | 0.000 | 5.270 | 1.506 | 1.238 | 0.483 | 1199.209(0.000) |
| TBILL 3M | Log returns of 3-month Treasury constant maturity rate in % | 1.325 | 0.000 | 5.190 | 1.531 | 1.194 | 0.343 | 1096.881(0.000) |
| TBILL 6M | Log returns of 6-month Treasury constant maturity rate in % | 1.443 | 0.020 | 5.330 | 1.561 | 1.136 | 0.181 | 978.850(0.000) |

Table 5 – Continued on next page

Table 5 – Continued from previous page

| Name | Description | Mean | Min. | Max. | Std. Dev. | Skewness | Kurtosis | Jarque-Bera(Pr.) |
|------------------------|--|-------|--------|-------|-----------|----------|----------|------------------|
| TBILL 1Y | Log returns of 1-year Treasury constant maturity rate in % | 1.563 | 0.080 | 5.300 | 1.526 | 1.000 | −0.110 | 755.394(0.000) |
| TBILL 5Y | Log returns of 5-year Treasury constant maturity rate in % | 2.565 | 0.560 | 5.230 | 1.237 | 0.364 | −0.996 | 286.353(0.000) |
| TBILL 10Y | Log returns of 10-year Treasury constant maturity rate in % | 3.288 | 1.370 | 5.420 | 1.062 | 0.099 | −1.266 | 308.702(0.000) |
| TBILL 20Y | Log returns of 20-year Treasury constant maturity rate in % | 3.864 | 1.690 | 6.050 | 1.093 | −0.033 | −1.369 | 353.128(0.000) |
| <i>Term spreads</i> | | | | | | | | |
| TS10Y1M | Log returns of Term spread between 10-year and 1- month treasury rate | 2.022 | −0.780 | 4.010 | 1.054 | −0.446 | −0.421 | 183.359(0.000) |
| TS10Y3M | Log returns of Term spread between 10-year and 3- month treasury rate | 1.963 | −0.640 | 3.850 | 1.068 | −0.429 | −0.594 | 204.742(0.000) |
| TS10Y6M | Log returns of Term spread between 10-year and 6- month treasury rate | 1.845 | −0.640 | 3.720 | 1.084 | −0.425 | −0.719 | 233.359(0.000) |
| TS20Y1M | Log returns of Term spread between 20-year and 1- month treasury rate | 2.598 | −0.570 | 4.730 | 1.272 | −0.458 | −0.691 | 247.689(0.000) |
| TS20Y3M | Log returns of Term spread between 20-year and 3- month treasury rate | 2.539 | −0.410 | 4.610 | 1.291 | −0.446 | −0.808 | 272.817(0.000) |
| TS20Y6M | Log returns of Term spread between 20-year and 6- month treasury rate | 2.421 | −0.440 | 4.510 | 1.310 | −0.438 | −0.897 | 295.654(0.000) |
| TS5Y1M | Log returns of Term spread between 5-year and 1- month treasury rate | 1.299 | −0.820 | 3.150 | 0.766 | −0.184 | −0.145 | 29.415(0.000) |
| TS6M1M | Log returns of Term spread between 6-month and 1- month treasury rate | 0.177 | −0.430 | 1.700 | 0.197 | 1.517 | 4.246 | 5130.974(0.000) |
| TS3M1M | Log returns of Term spread between 3-month and 1- month treasury rate | 0.059 | −0.520 | 0.780 | 0.110 | 1.139 | 5.778 | 7270.569(0.000) |
| <i>Default spreads</i> | | | | | | | | |
| DSBAAAAA | Log returns of Default spread between Moody's sea- soned Aaa and Baa corporate bond yield | 1.073 | 0.530 | 3.500 | 0.443 | 2.912 | 10.587 | 27506.566(0.000) |

Table 5 – Continued on next page

Table 5 – Continued from previous page

| Name | Description | Mean | Min. | Max. | Std. Dev. | Skewness | Kurtosis | Jarque-Bera(Pr.) |
|---------------------------------|---|--------|--------|-------|-----------|----------|----------|-------------------|
| DSBAA20Y | Log returns of Default spread between Moody's seasoned Baa and 20-year treasury yield | 2.067 | 1.110 | 5.380 | 0.657 | 2.166 | 6.860 | 12400.213(0.000) |
| DSBAA10Y | Log returns of Default spread between Moody's seasoned Baa and 10-year treasury yield | 2.643 | 1.480 | 6.160 | 0.786 | 1.561 | 4.259 | 5253.998(0.000) |
| DSBAA5Y | Log returns of Default spread between Moody's seasoned Baa and 5-year treasury yield | 3.366 | 1.540 | 7.200 | 1.072 | 0.517 | 0.909 | 357.236(0.000) |
| DSBAA1Y | Log returns of Default spread between Moody's seasoned Baa and 1-year treasury yield | 4.368 | 1.120 | 8.260 | 1.558 | -0.263 | -0.486 | 96.370(0.000) |
| DSBAA6M | Log returns of Default spread between Moody's seasoned Baa and 6-month treasury yield | 4.488 | 0.990 | 8.670 | 1.617 | -0.295 | -0.406 | 96.560(0.000) |
| DSBAA3M | Log returns of Default spread between Moody's seasoned Baa and 3-month treasury yield | 4.606 | 0.950 | 9.160 | 1.605 | -0.225 | -0.185 | 44.504(0.000) |
| DSBAA1M | Log returns of Default spread between Moody's seasoned Baa and 1-month treasury yield | 4.665 | 0.860 | 9.420 | 1.584 | -0.193 | 0.001 | 28.155(0.000) |
| <i>Macroeconomics variables</i> | | | | | | | | |
| Dummy expansion | A dummy variable that takes 1 to indicate expansion and 0 for recession | 0.103 | 0.000 | 1.000 | 0.305 | 2.604 | 4.784 | 9424.095(0.000) |
| Dummy Inverted Yield Curve | A dummy variable that indicates 1 when it is an inverted yield curve, otherwise 0 | 0.056 | 0.000 | 1.000 | 0.229 | 3.883 | 13.089 | 43637.892(0.000) |
| <u>U.S. Markets</u> | | | | | | | | |
| USDCAD | Log returns of exchange rate between U.S. Dollar and Canadian Dollar in % | -0.003 | -5.072 | 3.807 | 0.572 | -0.128 | 5.491 | 5695.325(0.000) |
| USDCNY | Log returns of exchange rate between U.S. Dollar and Chinese Yuan in % | -0.005 | -2.019 | 1.816 | 0.138 | -0.151 | 26.246 | 129792.585(0.000) |
| USDYEN | Log returns of exchange rate between U.S. Dollar and Japanese Yen in % | -0.003 | -5.216 | 3.343 | 0.620 | -0.330 | 4.599 | 4069.271(0.000) |
| USDEUR | Log returns of exchange rate between U.S. Dollar and Euro in % | -0.005 | -3.444 | 2.542 | 0.485 | -0.132 | 2.481 | 1174.048(0.000) |
| USDGBP | Log returns of exchange rate between U.S. Dollar and British Pound in % | 0.002 | -2.993 | 5.752 | 0.380 | 0.847 | 15.786 | 47491.890(0.000) |

Table 5 – Continued on next page

Table 5 – Continued from previous page

| Name | Description | Mean | Min. | Max. | Std. Dev. | Skewness | Kurtosis | Jarque-Bera(Pr.) |
|-----------------------------|---|--------|---------|--------|-----------|----------|----------|--------------------|
| USDCHF | Log returns of exchange rate between U.S. Dollar and Swiss Franc in % | -0.012 | -13.022 | 4.544 | 0.667 | -1.742 | 35.461 | 239187.985(0.000) |
| <u>UK Markets</u> | | | | | | | | |
| GBPCAD | Log returns of exchange rate between British Pound and Canadian Dollar in % | -0.005 | -7.498 | 2.997 | 0.608 | -0.546 | 6.779 | 8886.050(0.000) |
| GBPCNY | Log returns of exchange rate between British Pound and Chinese Yuan in % | -0.007 | -6.274 | 6.113 | 0.827 | -0.285 | 6.966 | 9205.399(0.000) |
| GBPYEN | Log returns of exchange rate between British Pound and Japanese Yen in % | -0.005 | -13.640 | 8.346 | 0.818 | -1.442 | 24.845 | 117855.539(0.000) |
| GBPEUR | Log returns of exchange rate between British Pound and Euro in % | -0.007 | -6.556 | 3.068 | 0.513 | -0.551 | 8.147 | 12734.845(0.000) |
| GBPUSD | Log returns of exchange rate between British Pound and U.S. Dollar in % | -0.002 | -9.505 | 3.130 | 0.594 | -1.136 | 16.573 | 52718.184(0.000) |
| GBPCHF | Log returns of exchange rate between British Pound and Swiss Franc in % | -0.014 | -17.375 | 8.075 | 0.679 | -3.837 | 107.081 | 2171133.731(0.000) |
| <i>Technical indicators</i> | | | | | | | | |
| <u>U.S. Markets</u> | | | | | | | | |
| SP500MA10 | The 10-day moving average of S&P 500 log returns in % | 0.019 | -2.995 | 1.959 | 0.323 | -1.256 | 7.756 | 12499.896(0.000) |
| SP500MA20 | The 20-day moving average of S&P 500 log returns in % | 0.019 | -1.654 | 1.055 | 0.224 | -1.369 | 6.281 | 8809.620(0.000) |
| SP500MA50 | The 50-day moving average of S&P 500 log returns in % | 0.020 | -1.014 | 0.571 | 0.139 | -1.615 | 6.538 | 9912.341(0.000) |
| SP500MA100 | The 100-day moving average of S&P 500 log returns in % | 0.020 | -0.518 | 0.377 | 0.101 | -1.701 | 5.202 | 7120.161(0.000) |
| SP500MA200 | The 200-day moving average of S&P 500 log returns in % | 0.021 | -0.360 | 0.242 | 0.075 | -1.649 | 3.823 | 4591.255(0.000) |
| SP500LAG1 | 1-day lag log returns of S&P 500 in % | 0.012 | -9.470 | 10.957 | 1.165 | -0.208 | 10.057 | 19086.838(0.000) |
| SP500LAG2 | 2-day lag log returns of S&P 500 in % | 0.019 | -9.470 | 10.957 | 1.165 | -0.208 | 10.055 | 19073.250(0.000) |
| SP500LAG3 | 3-day lag log returns of S&P 500 in % | 0.019 | -9.470 | 10.957 | 1.165 | -0.208 | 10.053 | 19062.204(0.000) |
| SP500LAG4 | 4-day lag log returns of S&P 500 in % | 0.018 | -9.470 | 10.957 | 1.165 | -0.208 | 10.052 | 19054.243(0.000) |

Table 5 – Continued on next page

Table 5 – Continued from previous page

| Name | Description | Mean | Min. | Max. | Std. Dev. | Skewness | Kurtosis | Jarque-Bera(Pr.) |
|-------------------|--|-------|--------|--------|-----------|----------|----------|------------------|
| SP500LAG5 | 5-day lag log returns of S&P 500 in % | 0.018 | −9.470 | 10.957 | 1.165 | −0.208 | 10.049 | 19039.322(0.000) |
| <u>UK Markets</u> | | | | | | | | |
| FTSE100MA10 | The 10-day moving average of FTSE 100 log returns in % | 0.006 | −2.578 | 1.432 | 0.326 | −0.995 | 4.724 | 4942.340(0.000) |
| FTSE100MA20 | The 20-day moving average of FTSE 100 log returns in % | 0.006 | −1.602 | 0.805 | 0.220 | −1.071 | 3.842 | 3630.986(0.000) |
| FTSE100MA50 | The 50-day moving average of FTSE 100 log returns in % | 0.007 | −0.719 | 0.478 | 0.130 | −1.077 | 3.291 | 2883.727(0.000) |
| FTSE100MA100 | The 100-day moving average of FTSE 100 log returns in % | 0.007 | −0.452 | 0.268 | 0.090 | −1.059 | 2.521 | 1999.016(0.000) |
| FTSE100MA200 | The 200-day moving average of FTSE 100 log returns in % | 0.008 | −0.283 | 0.205 | 0.068 | −1.064 | 1.765 | 1377.010(0.000) |
| FTSE100LAG1 | 1-day lag log returns of FTSE 100 in % | 0.006 | −9.266 | 9.383 | 1.152 | −0.135 | 7.209 | 9804.459(0.000) |
| FTSE100LAG2 | 2-day lag log returns of FTSE 100 in % | 0.006 | −9.266 | 9.384 | 1.152 | −0.135 | 7.207 | 9798.706(0.000) |
| FTSE100LAG3 | 3-day lag log returns of FTSE 100 in % | 0.006 | −9.266 | 9.384 | 1.153 | −0.135 | 7.205 | 9790.423(0.000) |
| FTSE100LAG4 | 4-day lag log returns of FTSE 100 in % | 0.006 | −9.266 | 9.384 | 1.153 | −0.135 | 7.203 | 9782.616(0.000) |
| FTSE100LAG5 | 5-day lag log returns of FTSE 100 in % | 0.006 | −9.266 | 9.384 | 1.153 | −0.135 | 7.202 | 9776.123(0.000) |

Note: This table shows the name, description, mean, minimum, maximum, skewness, kurtosis and Jarque-Bera test for normality with probability in brackets of 63 potential covariates. Some of the covariates are exclusive for U.S. or UK Markets and these variables are indicated separately, otherwise are the common variables for both markets.

A.2 Top three covariates VaR Forecasting for S&P 500

Table 6: Performance Evaluation with top three covariates for S&P 500

| a) $\alpha = 0.90$ | | | | | | |
|--------------------|-------------|------------------|--------------------|-----------------------|--------------------|--|
| Model | Average VaR | Binomial Test | | CCI by Christoffersen | | |
| | | # Violations (%) | P-value (Decision) | Test statistics | P-value (Decision) | |
| RF | 1.359 | 228 (47.899) | 0.000 (Reject) | 444.878 | 0.000 (Reject) | |
| SVM Linear | 1.244 | 8 (1.681) | 0.000 (Reject) | 54.458 | 0.000 (Reject) | |
| SVM Polynomial | 1.203 | 35 (7.353) | 0.056 (Accept) | 9.452 | 0.009 (Reject) | |
| SVM Radial Basis | 1.170 | 36 (7.563) | 0.079 (Accept) | 3.603 | 0.165 (Accept) | |

| b) $\alpha = 0.95$ | | | | | | |
|--------------------|-------------|------------------|--------------------|-----------------------|--------------------|--|
| Model | Average VaR | Binomial Test | | CCI by Christoffersen | | |
| | | # Violations (%) | P-value (Decision) | Test statistics | P-value (Decision) | |
| RF | 2.543 | 224 (47.059) | 0.000 (Reject) | 712.997 | 0.000 (Reject) | |
| SVM Linear | 1.882 | 2 (0.420) | 0.000 (Reject) | 34.737 | 0.000 (Reject) | |
| SVM Polynomial | 1.696 | 29 (6.092) | 0.291 (Accept) | 4.895 | 0.086 (Accept) | |
| SVM Radial Basis | 1.742 | 8 (1.681) | 0.000 (Reject) | 14.941 | 0.001 (Reject) | |

| c) $\alpha = 0.99$ | | | | | | |
|--------------------|-------------|------------------|--------------------|-----------------------|--------------------|--|
| Model | Average VaR | Binomial Test | | CCI by Christoffersen | | |
| | | # Violations (%) | P-value (Decision) | Test statistics | P-value (Decision) | |
| RF | 15.389 | 223 (46.849) | 0.000 (Reject) | Inf | 0.000 (Reject) | |
| SVM Linear | 6.199 | 0 (0.000) | 0.018 (Reject) | 9.568 | 0.008 (Reject) | |
| SVM Polynomial | 3.830 | 12 (2.521) | 0.004 (Reject) | 8.446 | 0.015 (Reject) | |
| SVM Radial Basis | 4.734 | 0 (0.000) | 0.018 (Reject) | 9.568 | 0.008 (Reject) | |

Note: Out-of-sample VaR forecasting with top three covariates (DJA,NASDAQ,VIX) for S&P 500. This table shows from column left to right: the average VaR over a testing period, the number of violations and their percentage is expressed in parentheses, then a two-sided binomial test is executed with null hypothesis of correct number of expected violations and we calculate the p-value, followed by [Christoffersen \(1998\)](#)'s jointly test with null hypothesis of independent violations and the test statistics with p-value is computed. We reject the null hypothesis when the p-value is smaller than 5% significance level and the decision results are expressed in parentheses.

A.3 Top three covariates VaR Forecasting for FTSE 100

Table 7: Performance Evaluation with top three covariates for FTSE 100

| a) $\alpha = 0.90$ | | | | | | |
|--------------------|-------------|------------------|--------------------|-----------------------|--------------------|--|
| Model | Average VaR | Binomial Test | | CCI by Christoffersen | | |
| | | # Violations (%) | P-value (Decision) | Test statistics | P-value (Decision) | |
| RF | 1.305 | 203 (35.552) | 0.000 (Reject) | 271.912 | 0.000 (Reject) | |
| SVM Linear | 1.155 | 62 (10.858) | 0.485 (Accept) | 0.738 | 0.692 (Accept) | |
| SVM Polynomial | 1.214 | 54 (9.457) | 0.727 (Accept) | 3.256 | 0.196 (Accept) | |
| SVM Radial Basis | 1.355 | 41 (7.180) | 0.025 (Reject) | 5.539 | 0.063 (Accept) | |

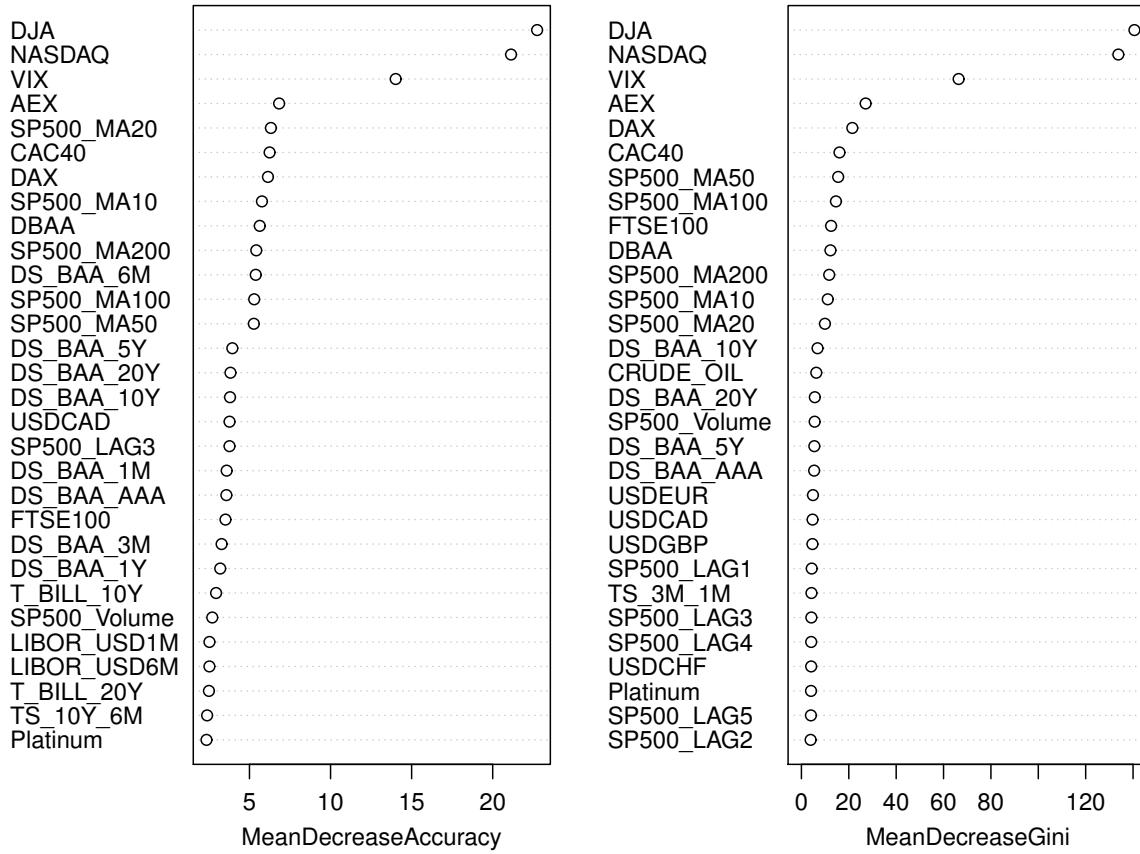
| b) $\alpha = 0.95$ | | | | | | |
|--------------------|-------------|------------------|--------------------|-----------------------|--------------------|--|
| Model | Average VaR | Binomial Test | | CCI by Christoffersen | | |
| | | # Violations (%) | P-value (Decision) | Test statistics | P-value (Decision) | |
| RF | 2.135 | 184 (32.224) | 0.000 (Reject) | 430.383 | 0.000 (Reject) | |
| SVM Linear | 1.529 | 26(4.553) | 0.701 (Accept) | 0.764 | 0.682 (Accept) | |
| SVM Polynomial | 1.587 | 24 (4.203) | 0.442 (Accept) | 3.717 | 0.156 (Accept) | |
| SVM Radial Basis | 1.785 | 17 (2.977) | 0.027 (Reject) | 6.120 | 0.047 (Reject) | |

| c) $\alpha = 0.99$ | | | | | | |
|--------------------|-------------|------------------|--------------------|-----------------------|--------------------|--|
| Model | Average VaR | Binomial Test | | CCI by Christoffersen | | |
| | | # Violations (%) | P-value (Decision) | Test statistics | P-value (Decision) | |
| RF | 14.764 | 167 (29.247) | 0.000 (Reject) | Inf | 0.000 (Reject) | |
| SVM Linear | 2.937 | 0 (0.000) | 0.005 (Reject) | 11.477 | 0.003 (Reject) | |
| SVM Polynomial | 2.979 | 0 (0.000) | 0.005 (Reject) | 11.477 | 0.003 (Reject) | |
| SVM Radial Basis | 3.477 | 1 (0.175) | 0.053 (Accept) | 5.978 | 0.050 (Accept) | |

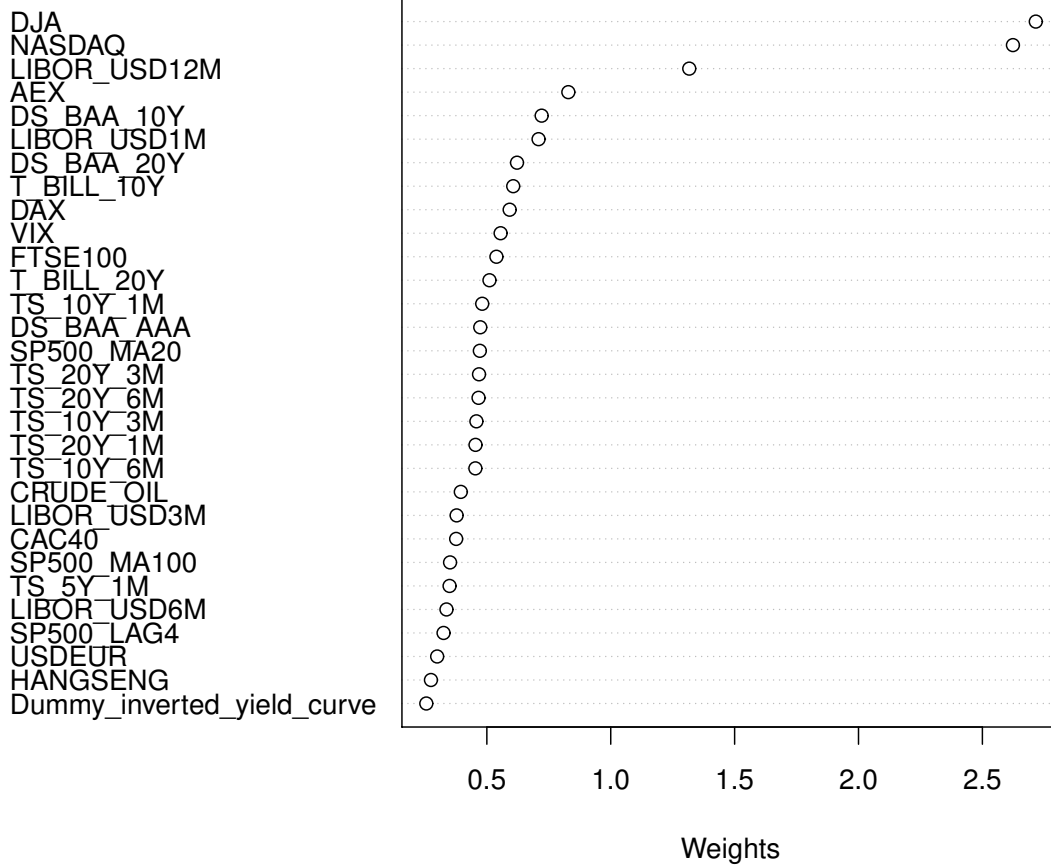
Note: Out-of-sample VaR forecasting with top three covariates (CAC40,AEX,DAX) for FTSE 100. This table shows from column left to right: the average VaR over a testing period, the number of violations and their percentage is expressed in parentheses, then a two-sided binomial test is executed with null hypothesis of correct number of expected violations and we calculate the p-value, followed by [Christoffersen \(1998\)](#)'s jointly test with null hypothesis of independent violations and the test statistics with p-value is computed. We reject the null hypothesis when the p-value is smaller than 5% significance level and the decision results are expressed in parentheses.

B Graphs

B.1 Variable importance of $g(x)$ estimation for S&P 500



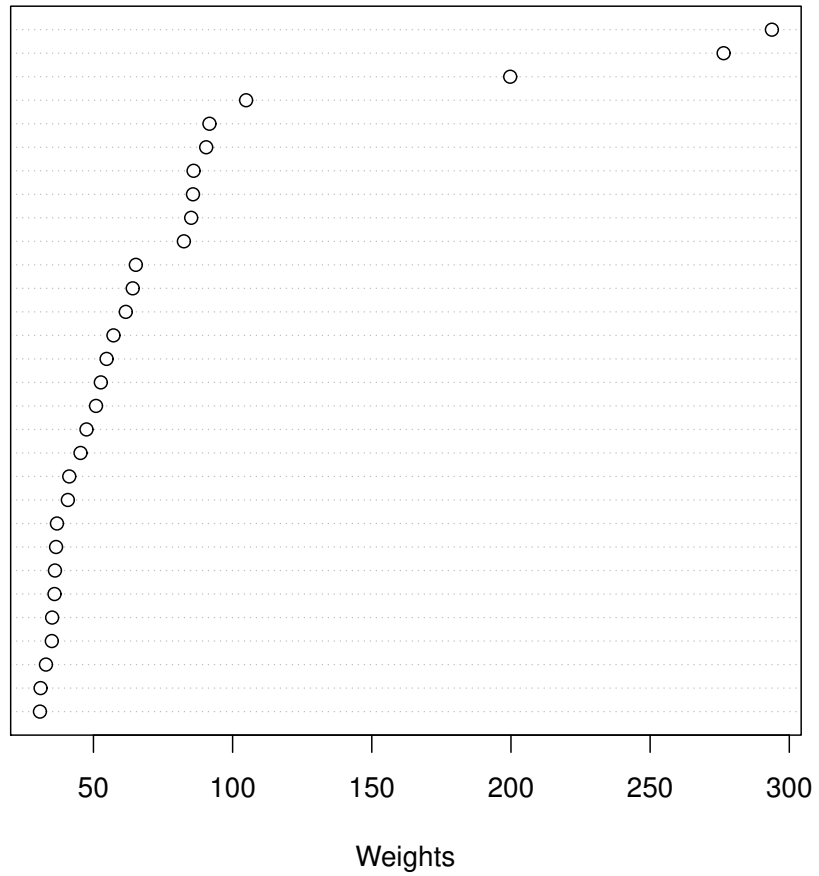
(a) Variable importance using RF



(b) Variable importance using SVM Linear

Figure 14: Variable importance of $g(x)$ estimation for S&P 500. Graph (a) displays the variable importance using RF. Left panel of graph (a) represents the percentage of observations that are incorrectly classified in % when excluding a variable in the model. Right panel of graph (a) represents the decrease of Gini Impurity in % when splitting node of a variable. A higher value indicates a more important variable. Graph (b) represents the importance measured by weights using SVM Linear. The covariates are sorted by their importance in descending order.

NASDAQ
 DJA
 VIX
 SP500_MA20
 DAX
 CAC40
 AEX
 SP500_MA10
 SP500_MA50
 FTSE100
 SP500_MA100
 DBAA
 SP500_MA200
 DS_BAA_AAA
 CRUDE_OIL
 T_BILL_20Y
 DAAA
 USDGBP
 Dummy_expansion
 T_BILL_10Y
 USDEUR
 Silver
 DS_BAA_10Y
 LIBOR_USD12M
 DS_BAA_5Y
 LIBOR_USD6M
 LIBOR_USD3M
 LIBOR_USD1M
 USDYEN
 SP500_LAG1



DJA
 NASDAQ
 VIX
 CAC40
 DAX
 AEX
 FTSE100
 SP500_MA100
 SP500_MA10
 SP500_MA50
 SP500_MA20
 DS_BAA_20Y
 DS_BAA_10Y
 SP500_MA200
 DS_BAA_AAA
 DS_BAA_5Y
 DS_BAA_3M
 DS_BAA_1M
 DS_BAA_6M
 DS_BAA_1Y
 CRUDE_OIL
 Dummy_expansion
 USDCAD
 USDYEN
 USDGBP
 DBAA
 TS_20Y_3M
 SP500_LAG4
 T_BILL_3M
 TS_20Y_1M

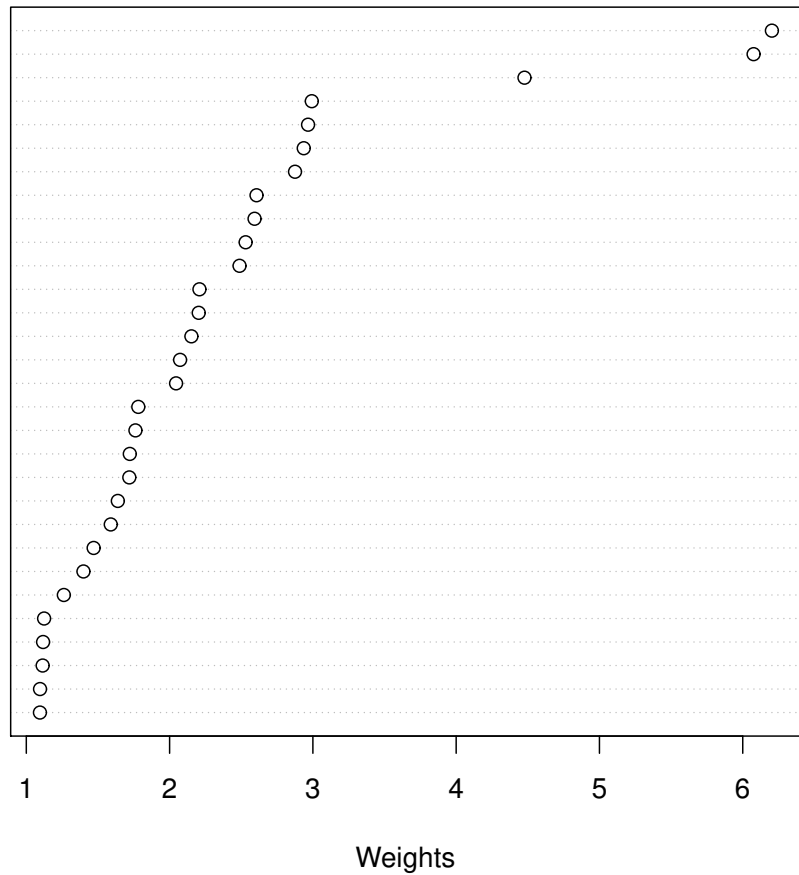
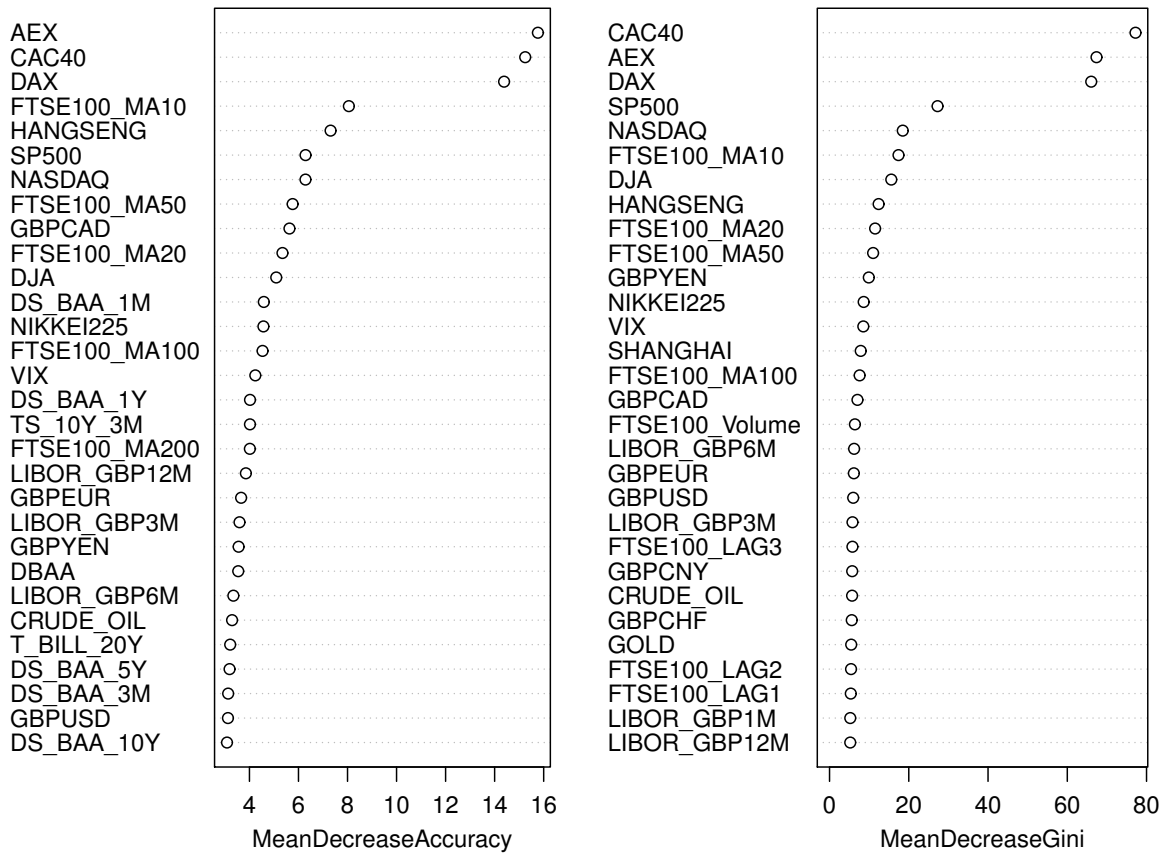
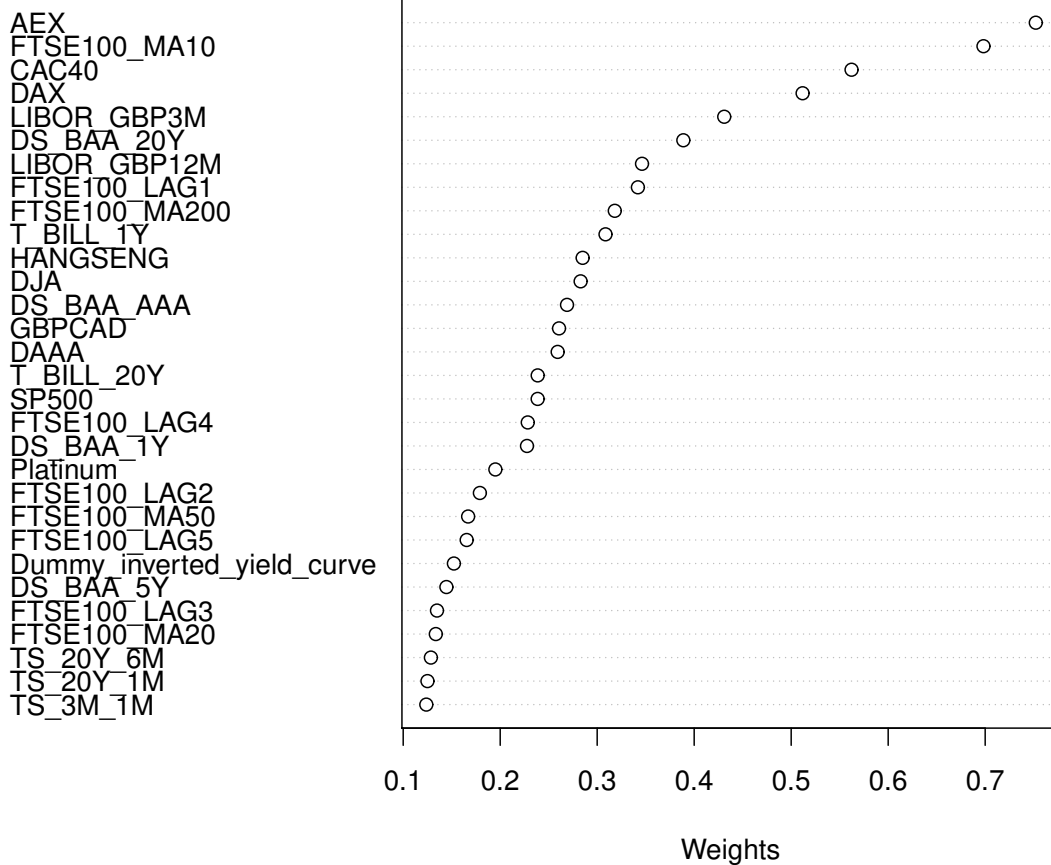


Figure 15: Variable importance of $g(x)$ estimation for S&P 500. Upper graph represents the importance measured by weights using SVM Polynomial. Bottom graph represents the importance measured by weights using SVM Radial. The covariates are sorted by their importance in descending order.

B.2 Variable importance of $g(x)$ estimation for FTSE 100



(a) Variable importance using RF



(b) Variable importance using SVM Linear

Figure 16: Variable importance of $g(x)$ estimation for FTSE 100. Graph (a) displays the variable importance using RF. Left panel of graph (a) represents the percentage of observations that are incorrectly classified in % when excluding a variable in the model. Right panel of graph (a) represents the decrease of Gini Impurity in % when splitting node of a variable. A higher value indicates a more important variable. Graph (b) represents the importance measured by weights using SVM Linear. The covariates are sorted by their importance in descending order.

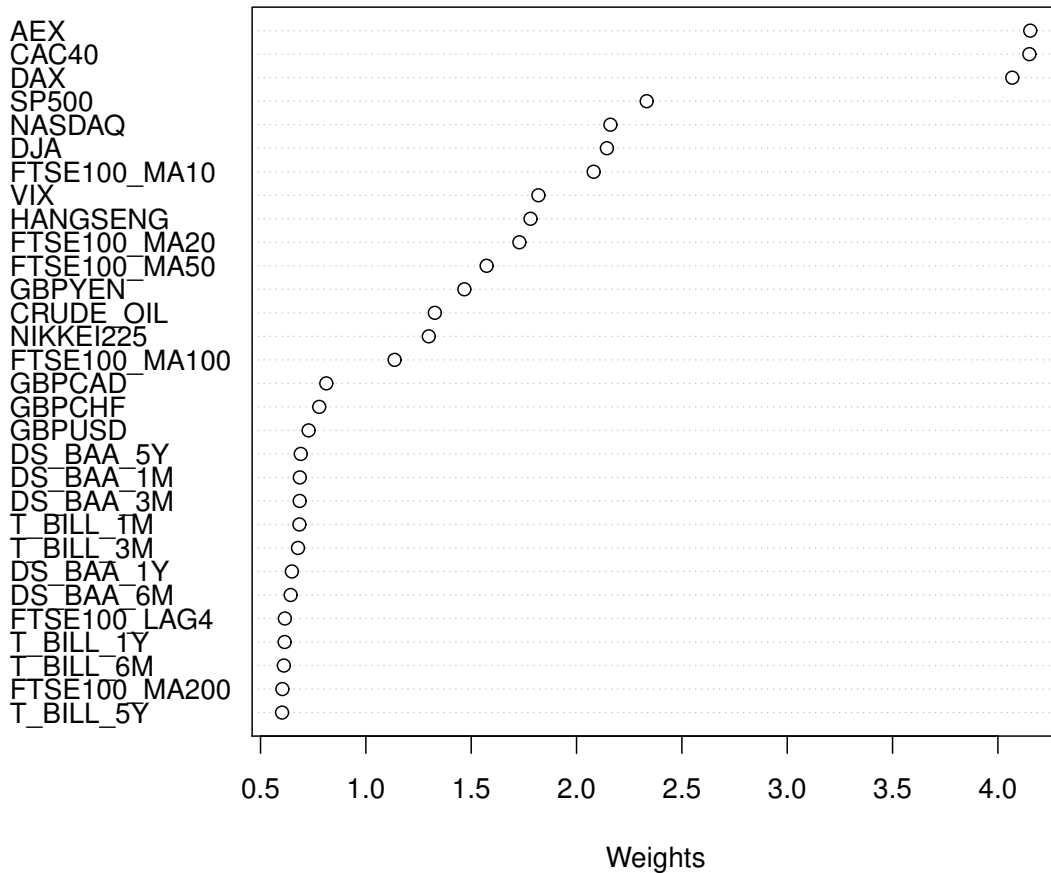
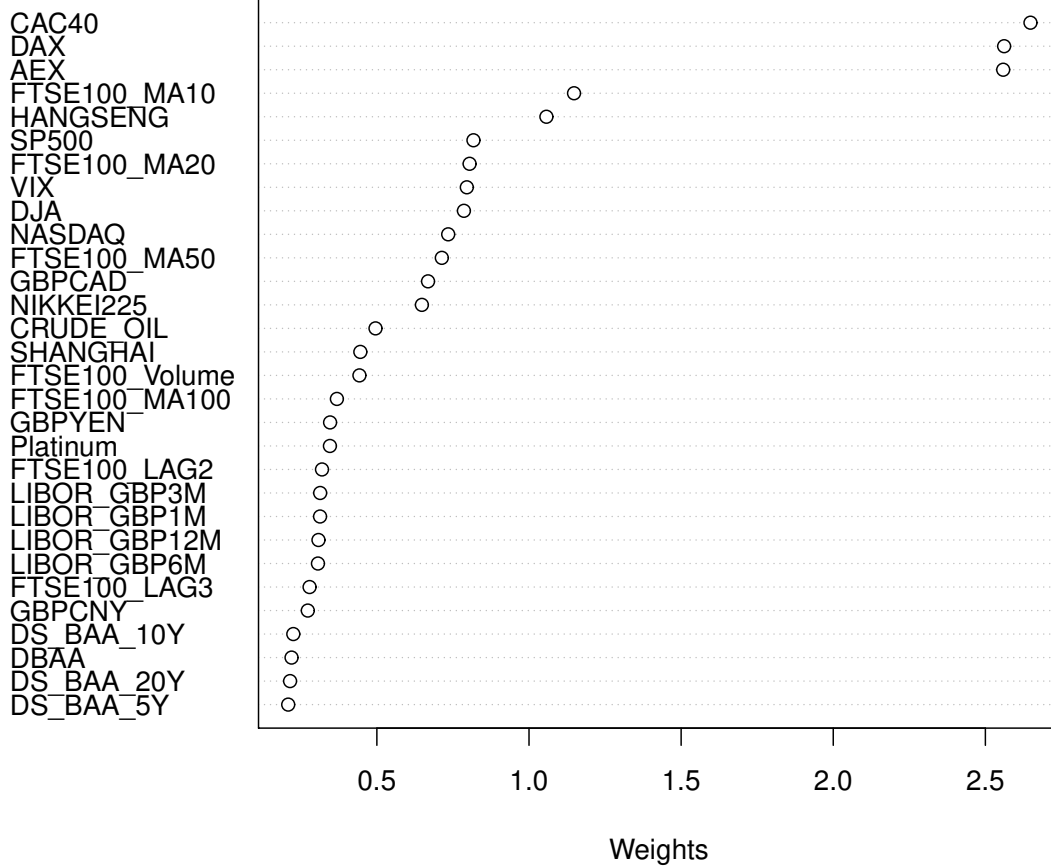
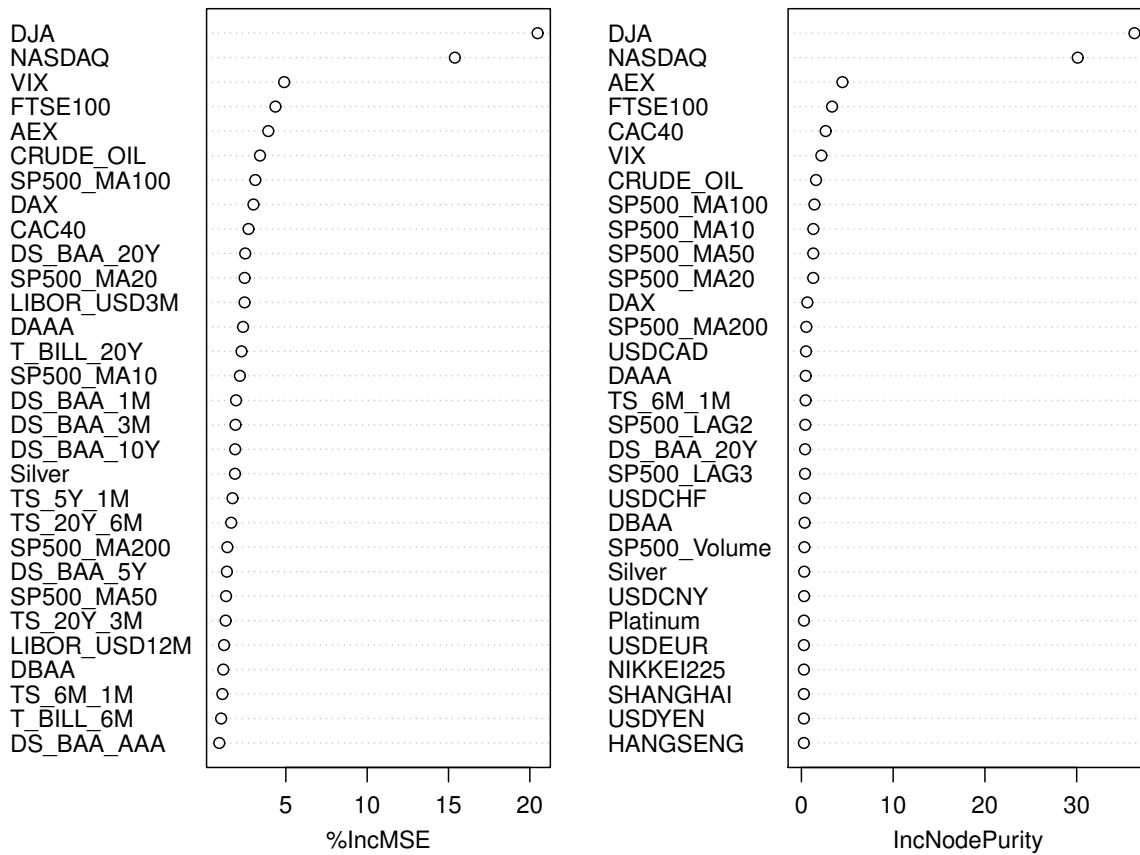
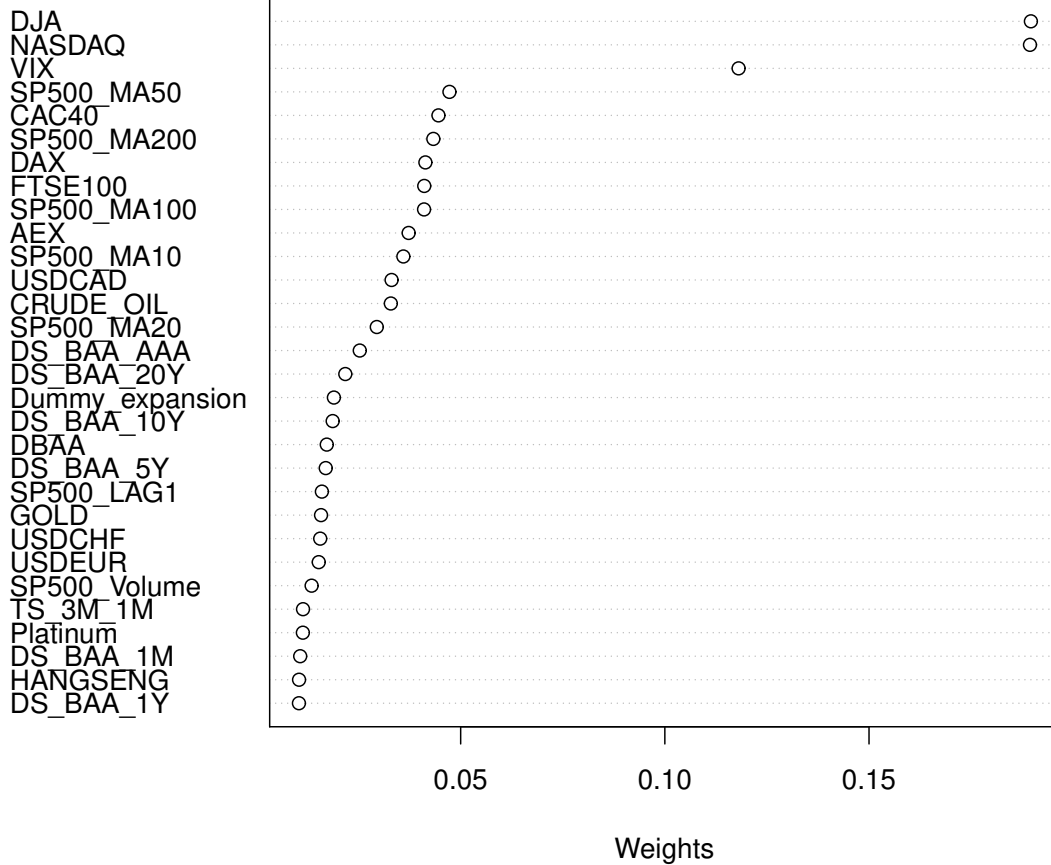


Figure 17: Variable importance of $g(x)$ estimation for FTSE 100. Upper graph represents the importance measured by weights using SVM Polynomial. Bottom graph represents the importance measured by weights using SVM Radial. The covariates are sorted by their importance in descending order.

B.3 Variable importance of $\xi(x)$ estimation for S&P 500



(a) Variable importance using RF



(b) Variable importance using SVM Linear

Figure 18: Variable importance of $\xi(x)$ estimation for S&P 500. Graph (a) displays the importance using RF. Left panel of graph (a) represents the increase of MSE in % when a variable is excluded from the model. Right panel of graph (a) represents the increase of SSE in % by node splitting of a variable. A higher value indicates a more important variable. Graph (b) represents the importance measured by weights using SVM Linear. The covariates are sorted by their importance in descending order.

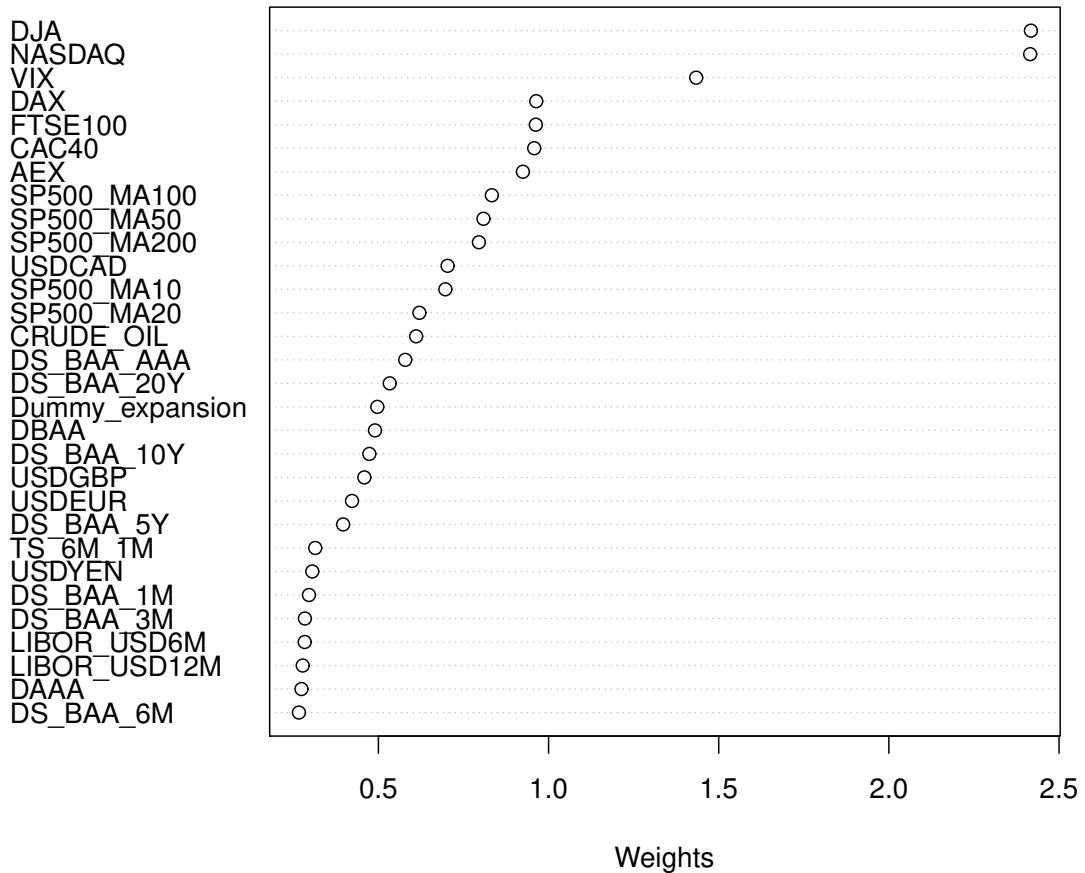
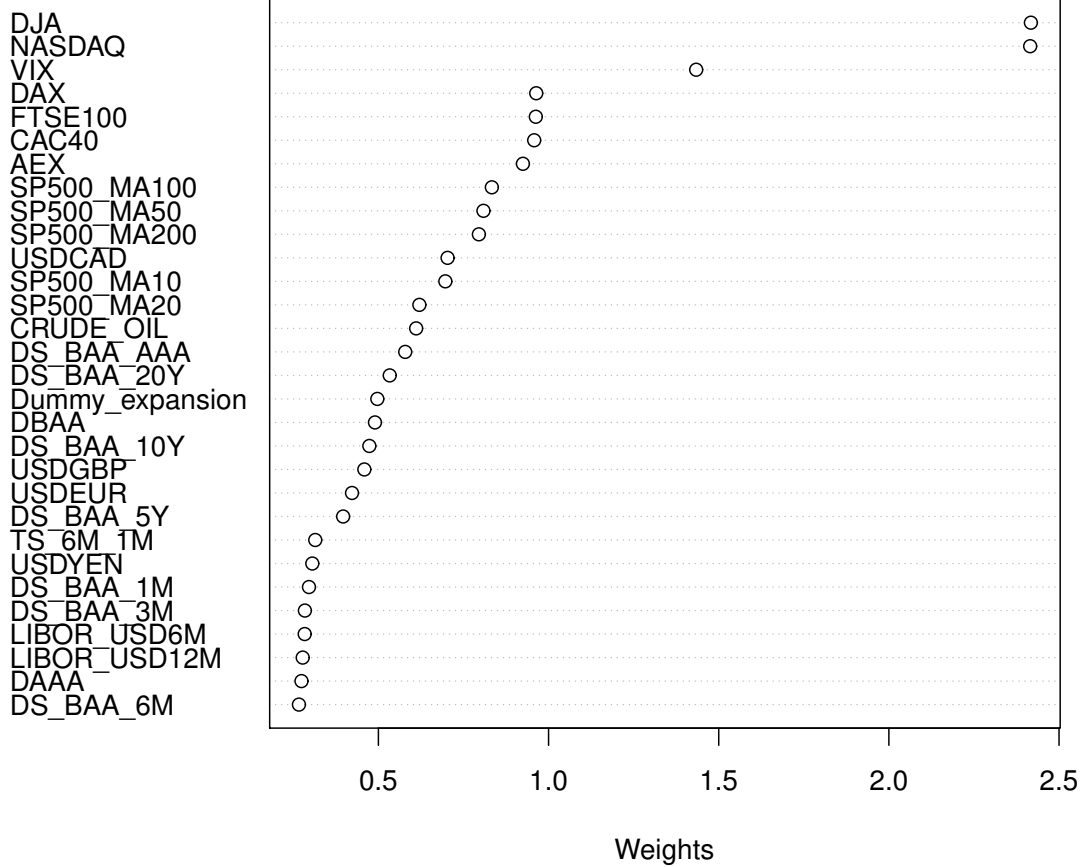
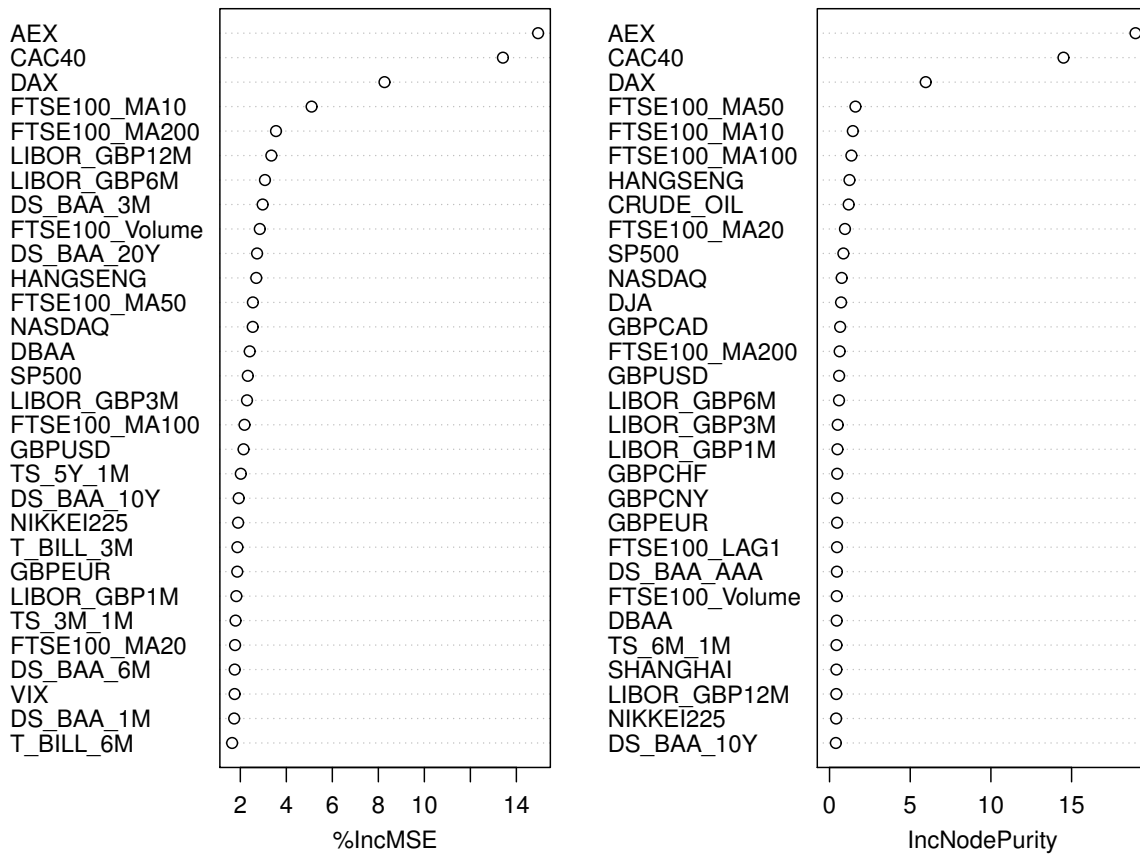
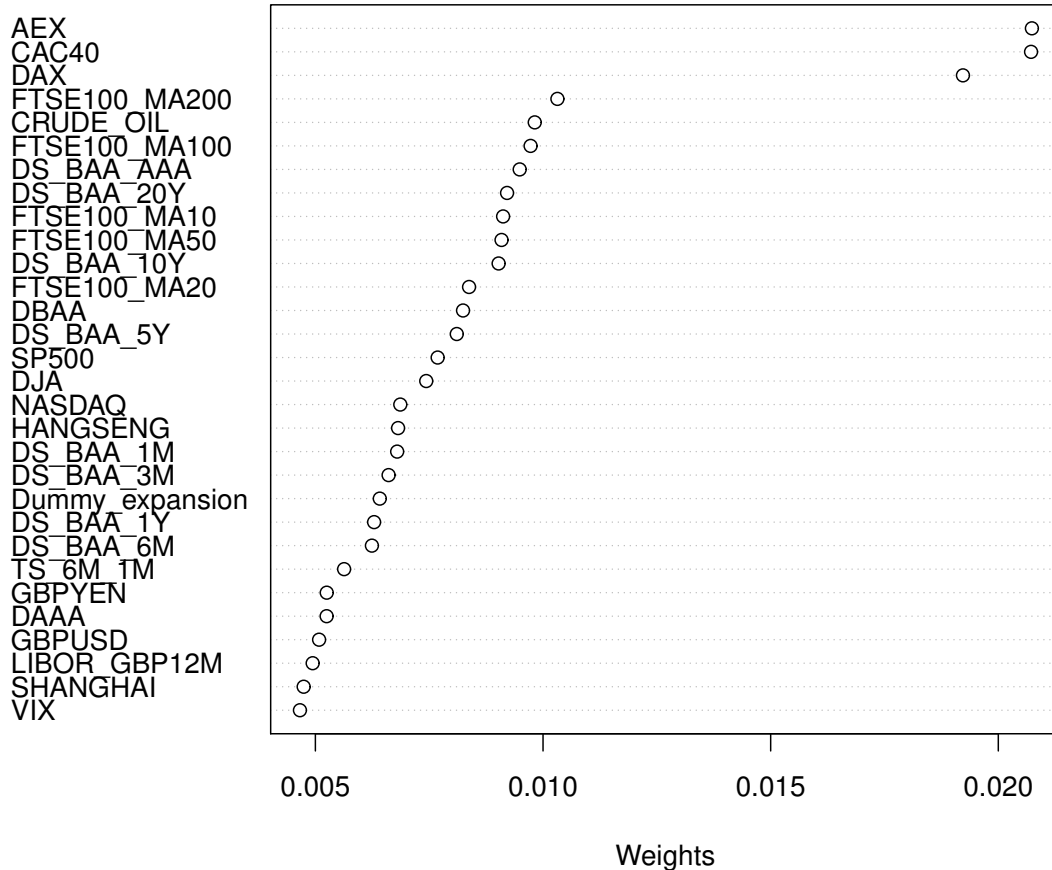


Figure 19: Variable importance of $\xi(x)$ estimation for S&P 500. Upper graph represents the importance measured by weights using SVM Polynomial. Bottom graph represents the importance measured by weights using SVM Radial. The covariates are sorted by their importance in descending order.

B.4 Variable importance of $\xi(x)$ estimation for FTSE 100



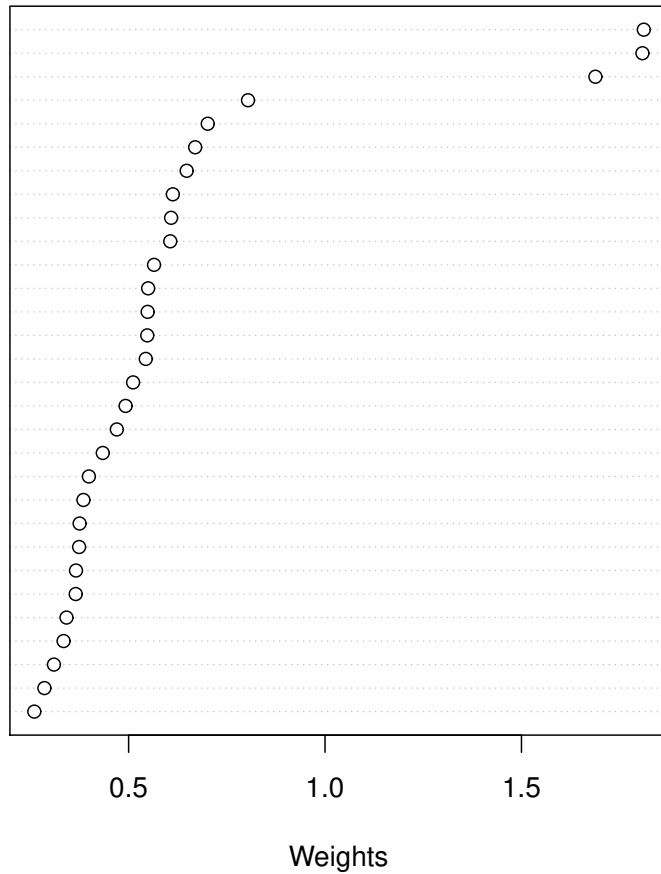
(a) Variable importance using RF



(b) Variable importance using SVM Linear

Figure 20: Variable importance of $\xi(x)$ estimation for FTSE 100. Graph (a) displays the importance using RF. Left panel of graph (a) represents the increase of MSE in % when a variable is excluded from the model. Right panel of graph (a) represents the increase of SSE in % by node splitting of a variable. A higher value indicates a more important variable. Graph (b) represents the importance measured by weights using SVM Linear. The covariates are sorted by their importance in descending order.

AEX
 CAC40
 DAX
 FTSE100_MA10
 CRUDE_OIL
 FTSE100_MA200
 FTSE100_MA20
 HANGSENG
 FTSE100_MA100
 FTSE100_MA50
 SP500
 DS_BAA_AAA
 DS_BAA_10Y
 DS_BAA_20Y
 DJA
 NASDAQ
 DS_BAA_5Y
 SHANGHAI
 DBAA
 VIX
 DS_BAA_1M
 DS_BAA_3M
 DS_BAA_1Y
 DS_BAA_6M
 FTSE100_Volume
 Dummy_expansion
 NIKKEI225
 GBP_YEN
 Dummy_inverted_yield_curve
 LIBOR_GBP12M



CAC40
 AEX
 DAX
 FTSE100_MA200
 DS_BAA_AAA
 CRUDE_OIL
 DS_BAA_10Y
 DS_BAA_20Y
 FTSE100_MA100
 DS_BAA_5Y
 FTSE100_MA50
 FTSE100_MA10
 DBAA
 SP500
 DJA
 FTSE100_MA20
 DS_BAA_1M
 DS_BAA_3M
 NASDAQ
 DS_BAA_1Y
 DS_BAA_6M
 Dummy_expansion
 HANGSENG
 TS_6M_1M
 GBP_YEN
 GBP_USD
 DAAA
 VIX
 FTSE100_Volume
 LIBOR_GBP12M

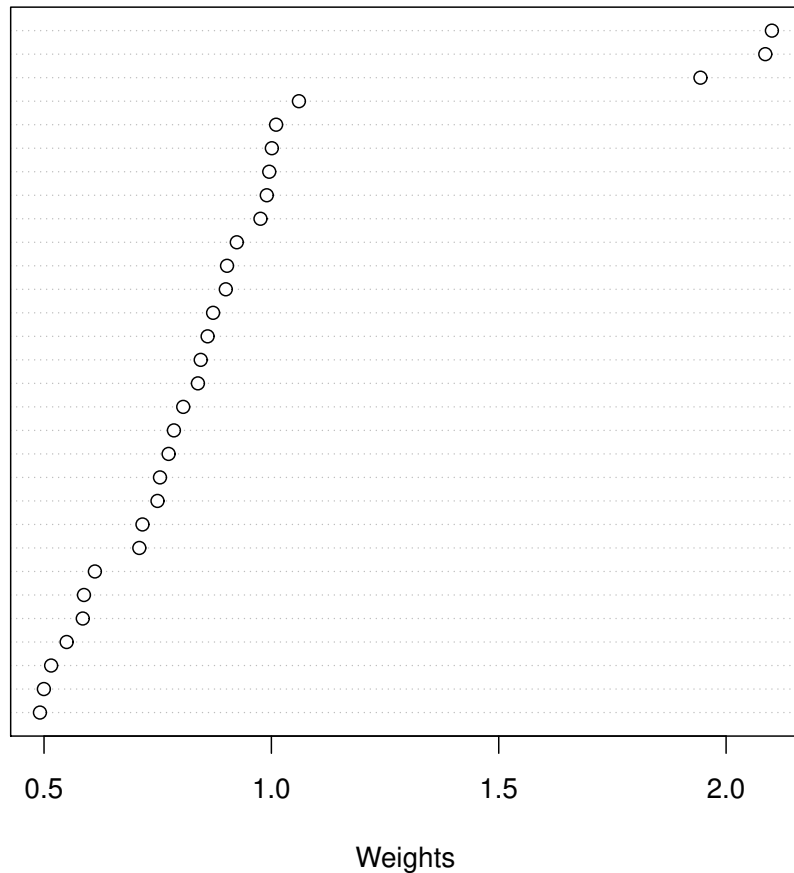
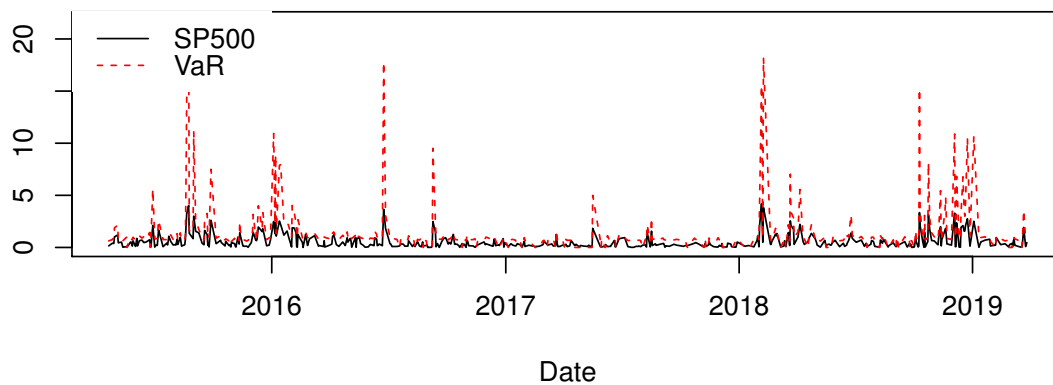


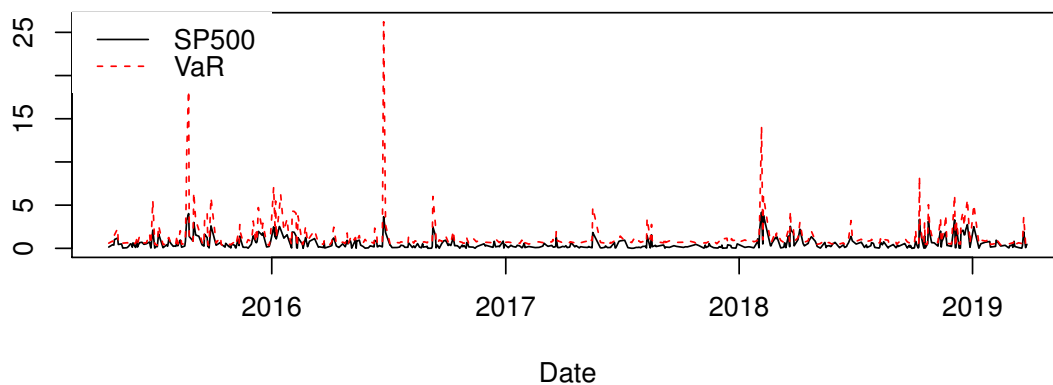
Figure 21: Variable importance of $\xi(x)$ estimation for FTSE 100. Upper graph represents the importance measured by weights using SVM Polynomial. Bottom graph represents the importance measured by weights using SVM Radial. The covariates are sorted by their importance in descending order.

B.5 VaR Forecasting

VaR Forecasting with RF



VaR Forecasting with SVM Linear



VaR Forecasting with SVM Polynomial

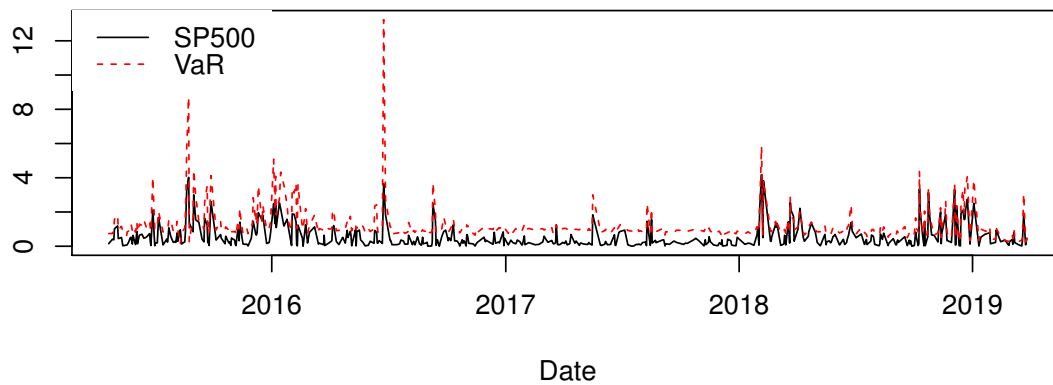
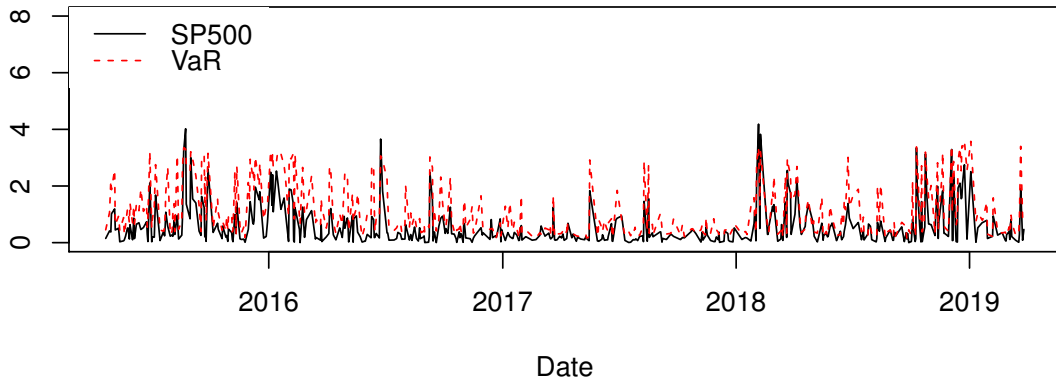
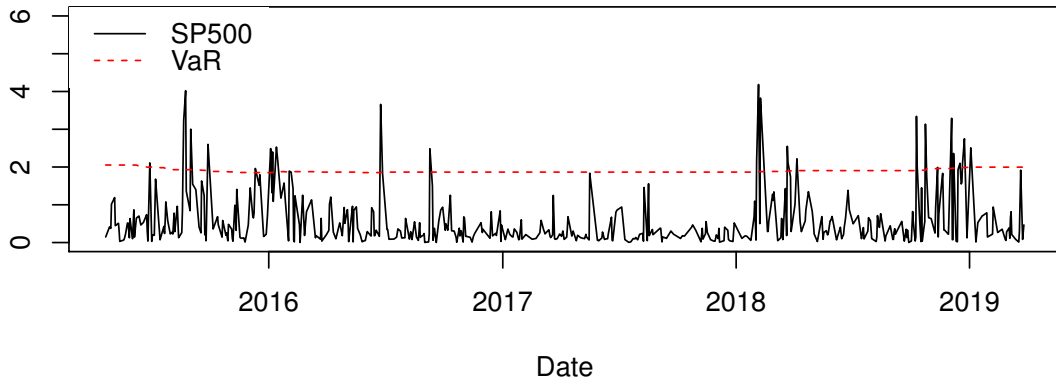


Figure 22: VaR Forecasting with 6 different models for S&P 500 and $\alpha = 90$

VaR Forecasting with SVM Radial



VaR Forecasting with HS



VaR Forecasting with GARCH

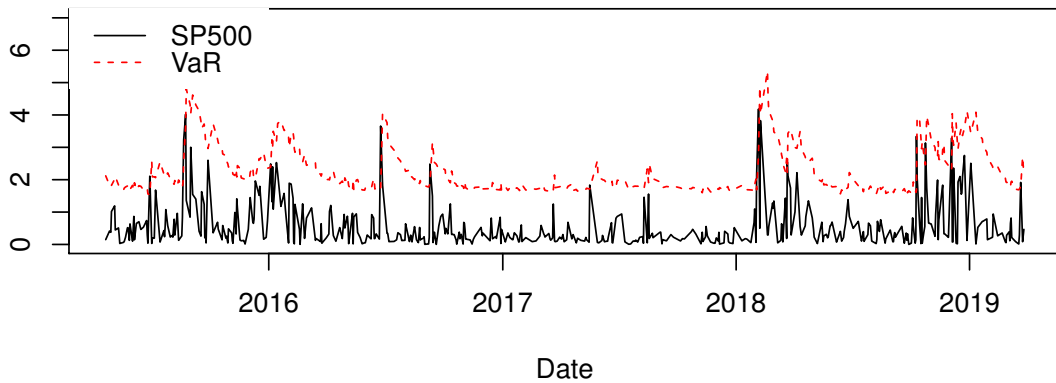
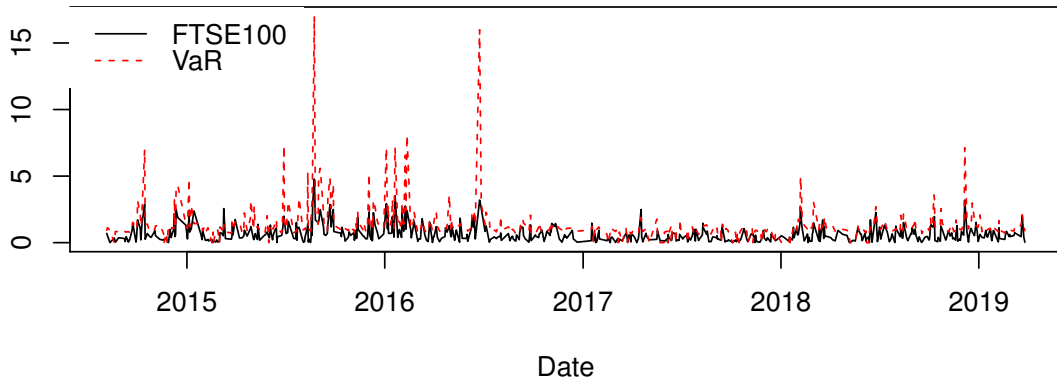
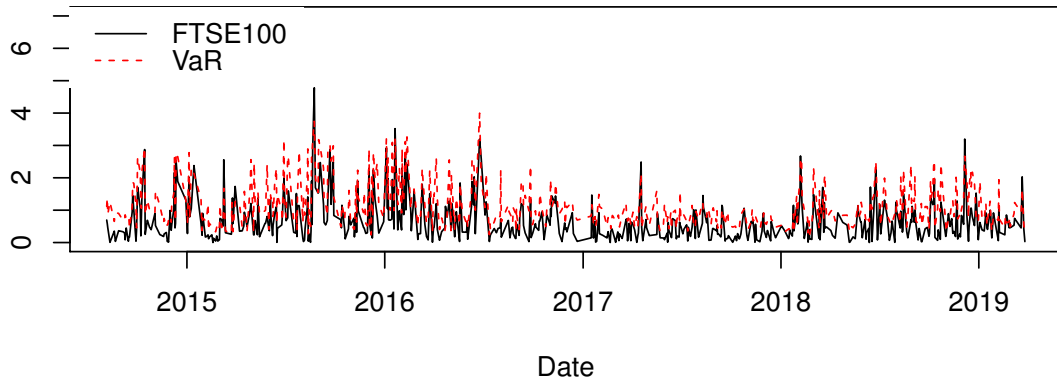


Figure 22: (Continued) VaR Forecasting with 6 different models for S&P 500 and $\alpha = 90$

VaR Forecasting with RF



VaR Forecasting with SVM Linear



VaR Forecasting with SVM Polynomial

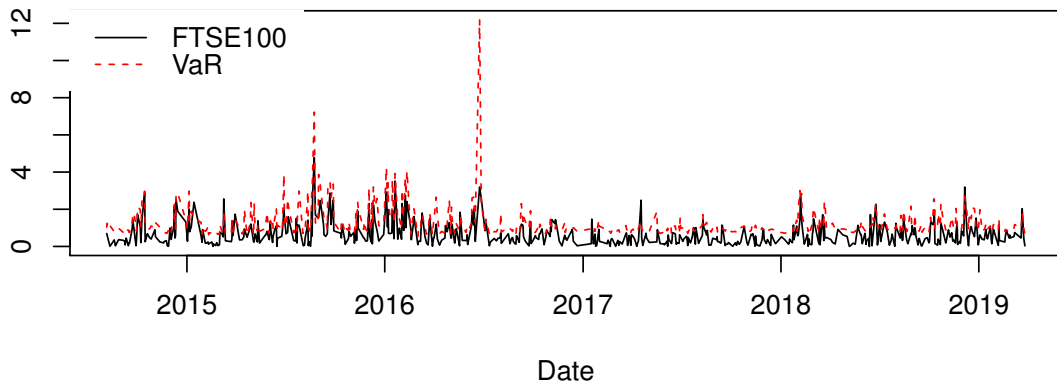
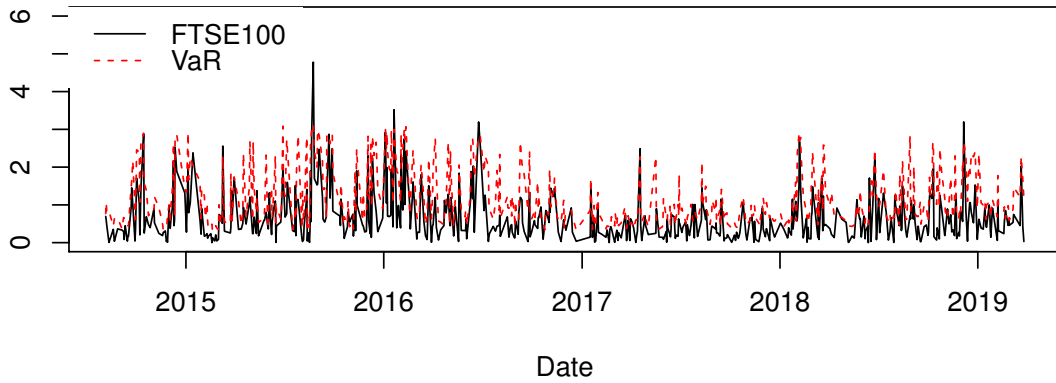
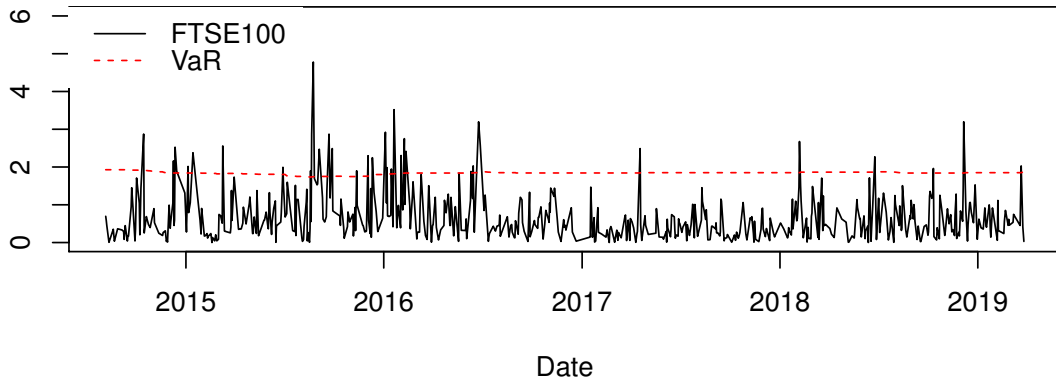


Figure 23: VaR Forecasting with 6 different models for FTSE100 and $\alpha = 90$

VaR Forecasting with SVM Radial



VaR Forecasting with HS



VaR Forecasting with GARCH

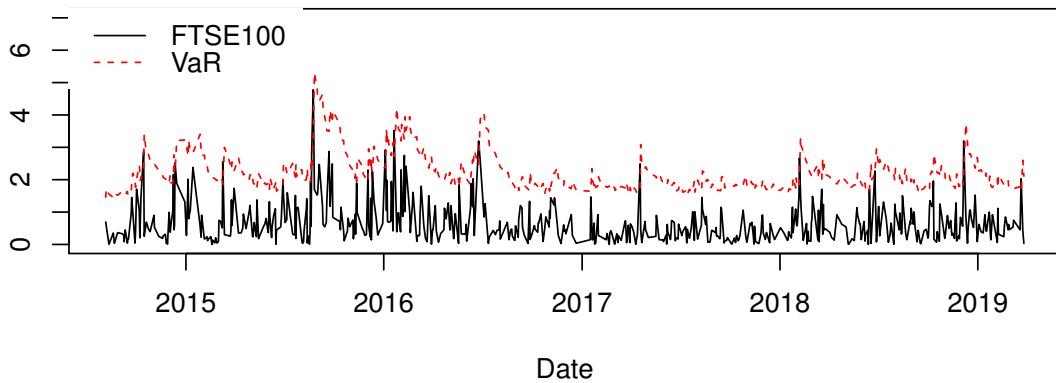


Figure 23: (Continued) VaR Forecasting with 6 different models for FTSE100 and $\alpha = 90$