Kristoffer Hemm

IPOs, regime switching and optimal portfolio allocation

The information contained in the first day returns of initial public offerings.

Master's thesis spring 2020
Oslo Business School
Oslo Metropolitan University
Master's program in business administration

Preface

This thesis marks the final chapter of the master's degree program in Business Administration with specialization in Finance and Economic Analysis at Oslo Metropolitan University. It was written during the spring semester of 2020.

First, I would like to thank OsloMet, and the administration at the Business School for their handling of the difficulties associated with the coronavirus lockdown of 2020. I would also like to thank my supervisor Johann Reindl for his good guidance and supplementation of useful material.

Oslo, June 2020

Kristoffer Hemm

Abstract

This thesis studied whether including the information contained in the first day IPO returns can be used to improve forecasts of the performance of financial markets. The S&P 500 index was used as a proxy for the financial market. The thesis utilized Hamilton's (1989) regime switching model in identifying the regimes in the training data. All the models contained two regimes, representing bull and bear markets. The first model was a regime switching (RS) model on the S&P 500 log returns, in the second model the VIX and the TED spread were added as regressors, in the third model the first day IPO returns was also added as a regressor. A GARCH(1,1) model was also be fitted to the S&P 500 log returns. The RS models and the GARCH model's performance was compared on pseudo out-of-sample performance during the first quarter of 2020. In addition, each RS model were back tested using Markowitz (1952) portfolio optimization theory, and compared to a portfolio created by an AR(1) model on the S&P 500 log returns that did not allow for regime switching. The back testing was done in the time period extending from January 1990 until December 2019. The results of the study showed that IPO returns had a non-statistically significant positive correlation with the S&P 500 log returns. In the negative returns regime, the IPO returns showed the highest level of significance, with a p-value of 0.1915. The results of the pseudo out-of-sample forecasts showed that the model that included the IPO returns performed the worst when forecasting the S&P 500 log returns during the first quarter of 2020. The best performing model in forecasting was the RS model that included the VIX and the TED spread as regressors, which even outperformed the GARCH model. The GARCH model outperformed the other RS models in the pseudo out-of-forecast. In back testing, there was a different story. Here, the best performing model was the RS model that included the IPO returns, which marginally outperformed the S&P 500 RS model. All the RS models outperformed the model that did not allow for regime switching. The RS model that included IPO returns had a holding period return that was 0.3% higher than the S&P 500 RS model. This comparison was done without considering the costs associated with trading. The S&P 500 RS model required rebalancing 9 times during the period, whereas the RS model that included the IPO returns required rebalancing 48 times in the same time period. These results indicate that adding IPO returns to the models did not add value to either the pseudo out-of-sample forecasts or in portfolio optimization.

Index

List of tables

List of figures

1. Introduction

The aim of this thesis is to explore if the information contained in the performance of first day returns of Initial Public Offerings (IPOs) can be used to improve forecasting of the performance of the financial market. The study will explore a regime switching model (Hamilton, 1989) on the S&P 500 log returns, and see how including traditional risk measures, and first day returns of IPOs, will affect the model. The reasoning behind this is that 2019 was a year where high-profile IPOs performed poorly. These include the IPOs of Lyft, Uber and Peloton, in addition to the IPO of WeWork which did not go through as planned (Bloomberg, 2019). Using the result of the forecast we will make suggestions for portfolio management using Markowitz portfolio theory (Markowitz, 1952).

To evaluate IPO returns as a factor, a regime switching model is employed to estimate returns and volatility in the U.S. stock market within two regimes: bull and bear markets. The data used was collected using the Thomson Reuters Eikon database, Yahoo Finance and The Federal Bank of St. Louis. Using the regime switching model, the study will try to predict when regime switch will occur with better accuracy by only using established indicators like the TED spread and the VIX. Using the results from the regime switching models, portfolio strategies that gives the highest risk adjusted return will be created for the different forecasted regimes. Each model will be back tested using the holding period return for 1990-2020. The models will also be compared using pseudo out-of-sample performance for the first quarter of 2020. The regime switching models will also be compared to the pseudo out-of-sample performance of a GARCH model (Bollerslev, 1986) of the S&P 500 log returns.

In economic theory, it is an established fact that the economy moves through cycles of booms and recessions (Diebold & Rudebusch, 1996; Harding & Pagan, 2002). 2019 marks the 10-year anniversary of one of the longest bull markets in history. The start of the bull market is marked by the lowest close of the S&P 500 index, which closed with a value of 676.53 on March 9th, 2009 (Macrotrends, 2019). In 2019, stock indices around the world were hovering around all time high levels. The S&P 500 hit an all-time closing high of 3025.86 on July 26th, 2019, after which it fell by a small amount, until it started increasing to new all-time highs in the last quarter of 2019. The performance of the S&P 500 during the last 30 years is shown in figure 1.1.

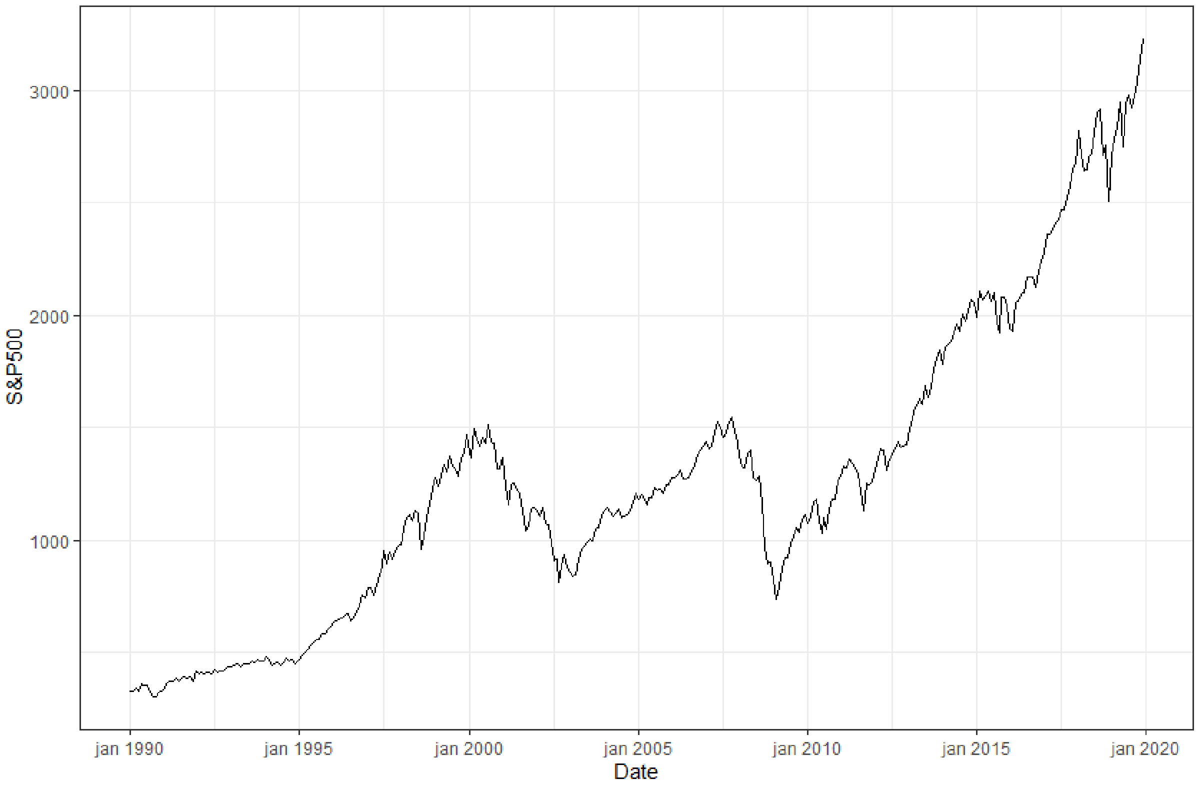

Figure 1.1. The S&P 500 during the last 30 years.

Despite indices such as the S&P 500 hovering around all time high levels in late 2019, indicators that suggest that the market is in the later stages of the business cycle started to emerge in 2018 and 2019. This include indicators such as the VIX, the TED spread and the slope of the zero curve, which have been showed to be correlated with the performance of the market indices such as the S&P 500 (Hull, 2012). During 2018 and 2019 the VIX showed an increase in the expected volatility in the markets, and the slope of the zero-curve inverted from positive to negative.

The argument for using IPO returns as an indicator for the state of the market is that historically IPOs tend to deliver results, that on average, are worse than a small cap index (Siegel, 2005). However, the distribution of IPO returns has a positive skew and fat tails, meaning that if an investor buys the right issue, he will achieve returns that are far above average (Ibbotson, 1975). Investing in IPOs can therefore be regarded as a quite speculative investing strategy. The argument for using IPO performance as an indicator of the market cycle is that when there is expected turbulence in the markets, investors tend to move from risky asset classes toward less risky assets. When investors are no longer willing to participate in new IPOs, the argument is that they are acting more prudent due to negative expectations for the future.

Another argument for the use of IPO returns is the information asymmetry associated with IPOs (Eckbo, 2007). The investment banks and issuing firm have a much higher degree of information about the prospect than potential investors buying the IPO, which hinders investors in making the same degree of informed investment decisions as compared to the more thoroughly analyzed open market. One of the reasons corporations have for going public is the current need for capital (Ibbotson, Sindelar & Ritter, 1994). To raise as much capital as possible the underwriting investment bank of the IPO tries to find the highest offer price where demand is still met. The valuation of the issuing company can be challenging for the investment bank. Often the companies going public are characterized by high growth but does not have much of a positive cash flow to show for. The investment bank must often base the valuation on growth projections for the issuing company, rather than the value of existing cash flows. In booming financial markets, investors tend to assign more value to a company's projected growth possibilities than in the company's actual earnings, as seen in the dotcom bubble of the late 1990's. If the offer price is set too high by the investment bank in the late stages of a booming market, and the investors are not willing to accept this price, this can be an indication that the psychology dominating the market is changing, and that the market's sentiment is moving away from projected growth and toward fundamental value. The argument is that this is a shift that will occur in the later stages of the boom cycle, and that poorer performance of IPOs is a sign of this shift.

The remainder of this paper is organized as follows. Section 2 will consist of a review of literature covering business cycles, IPOs and portfolio theory. Section 3 will describe the methodology of the thesis, the regime switching model and the GARCH model. This section will also present the data used in this thesis. In section 4, the results of the regime switching model and the portfolios created by each model will be presented, in addition to the results of the GARCH model. This is followed by a discussion of the results. Section 5 consists of the conclusion.

2. Previous research and theory

2.1 Business cycles

Financial markets can change their behavior quickly, going from boom to bust. Many studies have been conducted trying to identify macroeconomic factors that can be used in forecasting the variance seen in the return of financial assets. This part will describe the stages of a financial crisis and factors that can be used to forecast an increase in financial volatility.

The typical financial crisis consists of two distinct phases (Mishkin, 2019). The first phase consists of a boom and bust in the credit market and in asset prices. The credit boom and bust stage is characterized by irresponsible lending and introduction of new speculative financial products. Restrictions and regulation of the markets tend to become more relaxed. Financial institutions tend to go on a lending spree, the credit boom. During this boom, the lenders may not have the ability, or incentive, to properly manage the credit risk they take on. Eventually the less qualified borrowers start to default on their loans, driving down the capital of the financial institutions. With less capital available the financial institutions are forced to cut back on lending. The reduced availability of capital increases the risk profile on the financial institutions, causing lenders to pull their funds out of the financial institutions. With lesser funds available, the financial institutions are not able to make loans and the credit market can freeze up, leading to a lending crash.

The boom and bust stage in assets is characterized by pricing of the assets that are more driven by the psychology of the market, rather than the fundamental economic value of the assets, causing an asset-price bubble. The asset-price bubble is often driven by a credit boom. When the bubble bursts, the asset-prices tend to fall until a balance with the assets fundamental value is achieved. The fall in asset prices will reduce the net worth of companies, and financial institutions will be more restrictive in their lending to companies, and it will be harder for the companies to get funding for their projects. Due the fall in asset-prices, the net worth of the financial institutions themselves will also decline, and lead to a decline in their balance sheet. They will have to lower their lending activity, which will lead to a decline in the overall economic activity.

The second phase of a financial crisis is characterized by a crisis in the banking sector. When the balance sheet of the financial institutions declines, their funds are may dry up, leading to insolvency. If some banks are not able to meet their obligations and go out of business, a run on other banks may follow, forcing the banks to sell off their assets quickly in order to raise the funds necessary. This can leave these banks insolvent, and the panic in the sector will increase. This can trigger the authorities to bail out or shut down the insolvent firms and sell them off or liquidate them, to reduce the uncertainty in the financial markets. If successful, the economy can start recovering.

When signs of a financial downturn start to appear, it is seldom one specific event that triggers a financial crisis. Usually, there are several events that happen over a short period of time that cause the volatility in the stock market to rise. Since such event tend to happen in such close proximity, it has been termed volatility clustering. Volatility, due to its behavior, has become an important input in asset pricing, and portfolio- and risk management (Schwert, 1989 a). The observed volatility has been showed to have an impact in predicting business cycles (Schwert, 1989 b; Paye, 2012; Chauvet, Senyuz, Yoldas, 2015). Schwert (1989 b) and Paye (2012) found that the stock market volatility then to move countercyclically. When the volatility in stock returns increases, the GDP growth rate tend to fall. The relationship implied is that the preceding volatility can be used as an indicator for the state of the economy.
From a portfolio- and risk management perspective, trying to forecast the volatility of stock return is of great concern, and a lot of work has gone into to identifying macroeconomic factors that can be used in forecasting the performance of stocks. Using a simple linear forecasting model, Paye (2012) showed that factors such as commercial paper-to-Treasury spread, default return, default spread and investment-to-capital ratio can be used to forecast the onset of recessions, which in turn will drive volatility and expected stock returns.

Christiansen, Schmeling & Schrimpf (2012) developed a more extensive model than Paye when trying to identify macroeconomic factors that can be used in forecasting the performance of the equity market. First, they developed a model that included the volatility in the foreign exchange market, the bond market and in commodity markets, in addition to stock volatility as the explanatory variables. In addition to this model, they also explored a model that contained 38 explanatory variables. They were able to identify the following variables as statistically significant predictors for volatility observed in asset markets: valuation ratios, interest rate differentials in foreign exchange markets, and proxies for market liquidity and credit risk such as the TED spread. They found that other macroeconomic variables, and their proxies, had little or none significance in forecasting volatility of financial assets.

Wang, Wei, Wu & Yin (2018) found that the volatility of crude oil has a positive predictive relationship with stock volatility and can be used for forecasting the volatility of stock returns in the short term, up to 9 months. Crude oil is an important factor of production in modern industry, and earlier studies have shown that shocks in the oil price has impacted both the real economy (Hamilton, 1983) and future cash flows of companies (Jones & Kaul, 1996). Wang, et al. found that the forecasting power of crude oil was stronger using WTI oil rather than using Brent oil.

These studies only skim the water of the many attempts done in trying to identify factors that can help forecast the direction of the asset markets. In order to keep the models simple in this thesis, only traditional risk measures such as the VIX and the TED spread will be used in this study.

2.2 IPOs

Initial public offerings (IPOs), the process where a company offers shares to the public in order to raise capital, has been a topic of great interest for researchers within finance. The topic started attracting attention in the 1960's, and gained popularity in the 1970's with publications such as Louge (1973), Ibbotson (1975) and Ibbotson and Jaffe (1975). These early studies focused on the performance, both initially and in the aftermarket, of stocks issued to the public market during the 1960's. The results from these early studies showed that when a company go through with an IPO, the shares issued tend to be underpriced and that the investors who bought shares at the issuing offer price tend to make relatively large profits. The investors cannot be sure if any single investment in a new issue will give him positive returns, but the distribution of the initial returns shows a positive skew and fat tails. Thus, if the issue has a positive return, it is likely that the return will be much larger than the loss experienced in a negative return issue (Ibbotson, 1975).

The parties involved in an IPO are the issuing company, the investors and investment bankers (Louge, 1973). The role of the investment bankers is to be middlemen between the issuing company and the investors. They commonly buy the stocks from the issuing company, then resell them to investors in the public market. In addition, they assist the issuing company with procedural and financial advice. Because of the double role the investment bankers play, they must try to please both the issuing company and the investor who will buy the issue. If the issue is underpriced by too much, the investment bankers will reduce their own risk and gain favors with the investors but is likely to lose out on business from future issuers. If the issue is priced at equilibrium, or is overpriced, the investment bankers will take on unwanted risk, and lose favor with the investors. The investments bankers may also lose potential future issuers if they consistently show that they are not able to sell the issuing stocks. Due to these reasons, it is expected that new issues will be underpriced.

In the wake of these initial studies, a lot of the research regarding IPOs have tried to identify as to why IPOs tend to show this underpricing behavior. The theories regarding IPO underpricing have be divided into four main groups: asymmetric information, institutional reasons, control considerations and behavioral approaches (Eckbo, 2007). The theory that has received the most attention is the theory of asymmetric information, which states that the parties involved in the IPO: the issuing firm, the investment bank and the investors, have different levels of knowledge about the financial state of the issuing firm (Baron, 1982; Rock, 1986; Welch, 1989). This asymmetry in information is assumed to be the cause of the observed underpricing.

Baron’s (1982) model claims that the investment banker has more information about the market demand than the issuing company and is in a superior position to the issuing firm when the negotiations with the issuer is initiated. This will allow the investment banker to take on less risk and demand a larger compensation from the issuing firm, and the issuer will lose out on the contract. The problem of moral hazard arises.

Baron assumed that the issuer is in need to raise capital, and that the contract initially entered into with the investment banker will specify the number of shares to be issued. The only factor affecting the issuer’s proceeds will be the offering price and the terms of the contract with the investment banker. The proceeds, $x = x(p, e, \theta)$, will be a function of the offer price, p, the distribution effort of the investment banker, e, and a parameter vector, θ, that represents the factors affecting the demand for the issue. The distribution efforts of the investment banker are the extent to which the investment banker is able to persuade investors to buy the issue, assumed to be influenced by the investment banker's superior information about the issue compared to the investors.

The investment banker and the issuer are assumed to have knowledge about the demand vector, θ, which is conditional on the parameter δ, giving the conditional density function $h(\theta|\delta)$. The investment banker is assumed to be more informed about δ than the issuer. The relationship between the proceeds and p, e and δ is given by the density function $g(x|p, e, \delta)$ induced on x by the probability distribution $h(\theta|\delta)$. The proceeds are assumed to be strictly increasing in e, the more effort the investment banker puts into the distribution the larger the proceeds to the issuer, and decreasing in δ, if the issuer has less favorable information the lower his proceeds will be.

The issuer's objective is to specify a compensation function, S, that incentivize the investment banker to exert greater sales efforts than he normally would and to using his knowledge, δ, to set an offer price that is optimal for the issuer. The investment banker will set an offering price as a function of the value he reports to the issuer, $p = p(\hat{\delta})$, where the reported knowledge of the demand, $\hat{\delta}$, may not be equal to the true knowledge parameter, δ. The issuer may accept this price, or choose to withdraw the issue. $\pi(\hat{\delta})$ will denote the probability of going forward with the issue, and $1 - \pi(\hat{\delta})$ will denote the probability that the issue will be withdrawn. The compensation function $\bar{S}(p, \hat{\delta}, x)$ will be a function of the offering price, $\hat{\delta}$, and the proceeds, x. The contract the issuer offer will be constructed on a basis of the functions $\left(p(\hat{\delta}), \bar{S}(p(\hat{\delta}), \hat{\delta}, x), \pi(\hat{\delta})\right)$.

Given $(p, \bar{S}, \pi)$, the investment banker will choose to accept or reject the offer based on his true knowledge of δ. If the investment banker proceeds with the process, the offer price will be set as $p(\hat{\delta})$ and the compensation can be expressed as the function:

$$S(\hat{\delta}, x) = \bar{S}(p(\hat{\delta}), \hat{\delta}, x) \tag{2.1}$$

If the issue is withdrawn, the investment banker will not receive any compensation. This risk must be accounted for when calculating the investment banker's income. The investment banker's income R^* can be written as:

$$R^*\big(\hat{\delta}, e, x\big) = \big[S\big(\hat{\delta}, x\big) - C(e)\big]\pi\big(\hat{\delta}\big) \tag{2.2}$$

where $C(e)$ is the cost of the distribution efforts.

The investment banker is assumed to be risk neutral and seek to maximize his expected net income:

$$R^*\big(\hat{\delta}, e, x\big) = \int R^*\big(\hat{\delta}, e, x\big) g\big(x|p\big(\hat{\delta}\big), e\ \delta\big)\ dx \tag{2.3}$$

The investment banker's distribution efforts are a function of his true information, δ, and his report, $\hat{\delta}$. His optimal distribution effort is given by:

$$e\big(\hat{\delta}, \delta\big) = argmax_e\big(R^*\big(\hat{\delta}, \delta, e\big) = \pi\big(\hat{\delta}\big) \int \big[S\big(\hat{\delta}, x\big) - C(e)\big] g\big(x|p\big(\hat{\delta}\big), e, \delta\big)\ dx\big) \tag{2.4}$$

where argmax is the argument that maximize the investment banker's income, R^*.

If the securities are issued, the issuer's net proceeds, $N(\delta)$, are given by:

$$N(\delta) = \int (x - S(\delta, x) g(x|p(\delta), e, \delta)\ dx\ ) \tag{2.5}$$

The issuer does not know the true δ when entering into the contract with the investment banker. Therefore, he must try to maximize his expected net proceeds by:

$$N = \int \pi(\delta)[N(\delta) - \bar{N}(\delta)] f(\delta)\ d\delta \tag{2.6}$$

The issuer's problem can be stated as:

$$max_{p(\delta), S(\delta,x), \pi(\delta)} \int \left(\int (x - S) g(x|p(\delta), e(\delta, \delta)\delta) dx - \bar{N}(\delta) \right) \pi(\delta) f(\delta)\ d\delta \tag{2.7}$$

$$s.t. \quad R\big(\delta, \delta, e(\delta, \delta)\big) \geq R\left(\hat{\delta}, \delta, e\big(\hat{\delta}, \delta\big)\right) \quad for\ all\ \hat{\delta}\ and\ \delta \tag{2.8}$$

$$e\big(\hat{\delta}, \delta\big) = argmax_e R\big(\hat{\delta}, \delta, e\big) \quad for\ all\ \hat{\delta}\ and\ \delta \tag{2.9}$$

$$R(\delta, \delta, e(\delta, \delta) \geq 0 \quad for\ all\ \delta \tag{2.10}$$

$$1 \geq \pi(\delta) \geq 0 \quad for\ all\ \delta \tag{2.11}$$

The constraints represent the actions of the investment banker. Since the issuer does not know the true value of δ, he will not be able to find the optimal solution of this equation set. This represents the heart of the problem of the information asymmetry between the issuer and the investment banker. Because of this, the issuer must offer a compensation to the investment banker that is sufficiently large in order to incentivize him to offer his information truthfully, as represented by equation 2.8. This requires the issuer to offer a price that is below the optimal offer price, leading to underpricing of the issue.

Rock's (1986) model is concerned with the information asymmetry that arises between investors of a new issue that have superior information about the demand for the issue, and to those investors that are less information about the demand for issue. The basic premise of Rocks model is that the price of an issue is observable, whereas the demand for the issue is not directly observable. Because of this, the price of the issue might not correspond to the demand, and uninformed investors can never be sure if a high-priced issue represents favorable information, or is caused by some extraneous factor(s). If the observed price in fact does not correspond with the level of demand, informed investors who spends a lot of time analyzing securities may be able to uncover this fact and can profit from buying mispriced securities.

When a new issue is offered, the issuer and the investment bank typically set a price, p, and a quantity, Z, of the equity to be offered. This will not be allowed to change on the day of the offering. Based on this price, there might be excessive demand or excessive supply of the issue in the market, which will not be observable until after the offering date. If there is excess demand, the investment bank typically will ration the shares, if there is excess supply the offer will not be able to sell all the shares. The value of the per share, $\tilde{v}$, is uncertain.
The issuing firm and the investment banker is assumed to be better informed than any other market participant about the future of the company. However, the market as a whole will be better informed than the issuer and the investment banker. Individual market participants may be better informed in areas such as the issuer's competitors and the appropriate discount rate for cash flows in the capital market.

Rock's model makes some assumptions, which include the assumptions that the informed demand, I, is no greater than the mean value of the shares offered, $\bar{v}Z$, the uninformed investors have homogenous expectations about the distribution of $\tilde{v}$, and that all investors have the same wealth, set equal to 1, and the same utility. The model also assumes that it is the issuing firm that dictate the price of the offering and bears the risk of the issue being undersubscribed.

Informed investors will submit orders for shares as long as the realized value per share, $\tilde{v}$, exceed the offering price, P, and they will order to the full extent of their wealth:

$$I \quad if \quad p < \tilde{v},$$
$$0 \quad if \quad p > \tilde{v} \tag{2.12}$$

The uninformed investors are not to predict the realized value per share, $\tilde{v}$. Each of the N uninformed investor will wants to invest all of his wealth, T, in a new issue, with no regard to the realized wealth per share. This leads to the combined demand for the issue to be:

$$NT + I \quad if \quad p < \tilde{v},$$
$$NT \qquad if \quad p > \tilde{v} \tag{2.13}$$

If $p < \tilde{v}$, the issuer will experience excess demand and the probability that the order will be filled is b. If $p > \tilde{v}$, there will be excess supply, and the probability of the order being filled is b'.

If the issue is oversubscribed, rationing of the orders must occur. The value of the issue will be the sum of the orders filled:

$$bNT + bI = pZ, \quad if \;\; b < 1 \tag{2.14}$$

Rearranging and taking the expectation:

$$b = \min\left(\frac{pZ}{NT+I}, 1\right), \quad if \quad p < \tilde{v} \tag{2.15}$$

$$b' = \min\left(\frac{pZ}{NT}, 1\right), \quad if \quad p > \tilde{v} \tag{2.16}$$

It follows that $b < b'$, and that the probability for an uninformed investor of receiving an underpriced issue is less than or equal to the probability of receiving an overpriced issue. This will introduce a bias against the uninformed investors, where they will only receive some the shares if the informed investors find the offer attractive and all of the shares if the informed investors find the offer to be unattractive. This has popularly been termed “The Winner’s Curse”.

Another popular term in the IPO literature is phenomenon of “Hot Issue Markets” (Ibbotson & Jaffe, 1975). They showed that the extent of the underpricing seen in IPOs varies over time. The number of companies going public also tend to vary over. Some periods are characterized by IPOs being overpriced when they are issued to the market. However, periods where IPOs are underpriced are much more common. In periods significant underpricing is observed, that coincides with a high number of companies going public are known as “Hot Issue Markets”. Ibbotson & Jaffe observed that such market conditions tend to be predictable, and that the numbers of new issues in a given months is predictable. They argued that this happens because underwriters tend to price IPOs by multiply the issuing firm’s current earnings by an industrywide P/E ratio. This P/E ratio tends to be based on observations that are several months old. Thus, new issues tend to follow past market performance.

The underpricing observed in IPOs tend to be evident fairly quickly, usually in the first day of trading, after which the share price tends to stabilize (Eckbo, 2007). Because of this, most studies investigating underpricing tend to focus on data from the first day of trading, and the IPO returns are usually estimated as the difference in the offering price and the closing price of the first day of trading.

Several studies have been conducted trying to establish a relationship between the volume of IPOs and the overall business cycle. Typically, this has been done by comparing the securities issuing market (primary market) and the securities trading market (secondary market). It has been shown that these markets influence each other (Ibbotson, Sindelar & Ritter, 1994; Ritter, 2003). Some studies have shown that it is primarily the secondary market that impacts the primary market, and that it is the performance of the stock market that will influence the IPO activity (Ritter, 2003; Kim & Weisbach, 2008). The hypothesis that it is the performance of the stock market that influences IPO performance has been termed the “IPO Cycle Hypothesis”.

The opposing view, that it is the performance of the primary market that influences the secondary market have also gained some support. This has been termed "The "Expansion Curse Hypothesis", which claims that an expansion in the primary market will lead to a decline in stock prices (Baker & Wugler, 2000; Ofek & Richardson, 2000; Braun & Larrain, 2009).

Jin, Gou, Zhou and Li (2016) showed that there is an interaction between the primary and secondary market in the China. They found that in generall, it is the secondary market that had a one-way impact on the primary market. However, when they partitioned the secondary market into bull and bear markets, they fund that during bull markets it was the secondary market that has a one-way impact on the primary market, while during bear markets the primary market had a one-way impact on the secondary market. During bear markets, this impact was negative, meaning that an increase in the IPO financing amount would significantly have a negative effect on the performance of the secondary market. However, it is worth noting that the Chinese stock market is not a fully developed free market, showing some characteristics of being an emerging market and that it has some degree of central planning.

Lowery, Officer and Schwert (2010) looked at the volatility of initial IPO returns. They found that there is a significant volatility in initial returns following a new stock issue, and that the amount of volatility tends to fluctuate over time. They found that the initial returns seen during hot IPO markets are characterized by a higher variability then the variability seen in calmer markets. When mean returns are high, volatility tends to also be high. From this, they argue that when the IPO cycle is in a hot market, it will be more difficult for the underwriters to price the issues correctly. They follow Rock's (1986) argument of information asymmetry, that the issuing company and the underwriter know more about the company's prospects than other market participants, but that the aggregate market is better informed about the demand for the company's shares. The aggregate demand for the IPOs will vary with time, and will be dependent on the type of firm that is going public. They also found that the characteristics of the issuing firm will greatly affect the pricing error seen in underwriting. Volatility tend to increase when the issuing firm is difficult to value. The characteristics of difficult to value firms are firms that are young, small and in the technology sector.

2.3 Portfolio optimization

The future state of the economy is impossible to predict without error and investing in securities is inherently a risky endeavor. For a rational investor, the risks involved in an investment is as important as the expected return. The risk faced by an investor is composed of the systematic risk, which consist of the market wide risk, and the unsystematic risk which is associated with each unique security (Bode, Kane & Marcus, 2018). Through diversification investors try to reduce the unsystematic risk as much as possible. However, it is impossible to reduce the systematic risk through diversification.

In financial theory it is assumed that investors require to get paid to take on risk. More formally, the higher the volatility of a security, the higher the expected return for the security must be for a rational investor to be willing to invest in the security. This is known as the risk premium. Expected returns are used due to the fact that the future is uncertain, and returns cannot be observed directly. The expected return is based on the historical performance of the security. When examining several securities, the historical performance of the securities is not likely to be perfectly correlated. This gives the investor the opportunity to reduce the unsystematic risk by building a portfolio of several securities.

An investors willingness to commit to risky portfolios is dependents on his level of risk aversion. Most investors are assumed to be risk averse, meaning they will only invest in risk-free assets, or assets that have a positive risk premium. In financial theory it is assumed that an investor will assign certain levels of welfare, or utility, to different portfolios based on the portfolios' expected return and risk. Utility is a central theme in economics, and the connection between expected returns and utility works as a bridge between the fields of finance and economics. An investor's utility is given by:

$$U = E(r) - \frac{1}{2}A\sigma^2 \qquad (2.17)$$

where U is the utility, E(r) is the expected return, A is the investors degree of risk aversion, and σ^2 is the variance of the returns. Equation 2.17 shows that the expected return have a positive influence on utility, whereas the volatility negatively impact the utility.

The idea of diversification has existed for a long time, but Harry Markowitz's (1952) mean-variance optimization represents a landmark in modern portfolio theory and in the construction of efficient portfolios. Using the expected returns, standard deviations and correlations, he showed that several assets can be combined in a diversified portfolio in order to maximize the expected returns for one level of risk. Alternatively, the risk can be minimized for a level of expected returns.

A portfolio consisting of n assets will have an expected return and variance given by:

$$E(r_p) = \sum_{i=1}^{n} w_i E(r_i) \tag{2.18}$$

$$\sigma_p^2 = \sum_{i=1}^{n} \sum_{j=1}^{n} w_i w_j Cov(r_i r_j) \tag{2.19}$$

where the weights for each asset, w_i, can be calculated from the covariance matrix.
From a set of risky assets, a set of different efficient portfolios can be constructed. Together these make up the efficient frontier of the risky assets. The efficient frontier represents all the risky portfolios that, for any given risk level, will give the highest return. From this set of portfolios, there will exist a well-diversified portfolio, the risky optimal portfolio, that will out-perform the all the other portfolios constructed by the same set of assets. To identify the optimal portfolio, and the weights assigned to each asset in this portfolio, investors typically use the Sharpe ratio. The Sharpe is the ratio between the risk premium and the standard deviation of any given portfolio. The optimal risky portfolio is the one that maximizes the Sharpe ratio:

$$Max\ S_p = \frac{E(r_p) - r_f}{\sigma_p}\ , \qquad s.t. \sum wi = 1 \tag{2.20}$$

When the optimal risky portfolio has been identified, the optimal complete portfolio can be constructed by combining the optimal risky portfolio with a risk-free asset, typically government bonds. The allocation between the optimal risky portfolio and the risk-free asset in the optimal complete portfolio is dependent on the investor's degree of risk aversion, A. The optimal amount invested in risky portfolio is given by:

$$y = \frac{E(r_p) - r_f}{A\sigma_p^2} \tag{2.21}$$

and the optimal amount invested in the risk-free asset is $x = 1 - y$.

What a well-diversified portfolio means, and the requirement for achieving optimal diversification, has been a topic of great interest in finance. Notable contributions to the debate have been Evans and Archer (1968), and Wagner and Lau (1971), who claimed that the effect of diversification would diminish when a portfolio contains about 10 different stocks. On the other side Statman (1987) claimed that diversification should be increased by adding more stocks, as long as the marginal benefits exceed the marginal costs. Statman argued that a well-diversified portfolio is one that eliminates the systematic risk. In order to achieve this, a portfolio must contain at least 30-40 stocks.

Newer contributions to financial theory have highlighted several drawbacks of the mean-variance optimization procedure (Chow, Jacquier, Kritzman & Lowry, 1999). One potential drawback is that the estimated risk usually is estimated with equal weights given to each observation in the respective time period. Financial markets typically go through periods where returns and volatility behave unregularly. Some periods are characterized by high volatility, while other periods are characterized by relatively low volatility. When the weights assigned to each time period are equal, investors run the risk of assigning too much weight to low volatility periods, compared to high volatility periods. This can introduce a bias in the estimation of the risk parameters, leading to construction of a portfolio that will not perform optimally in periods of high volatility. Chow et al. (1999) propose that different risk parameters should be estimated for different time periods, both for periods of financial stress and for periods experiencing little financial stress. Then different portfolios should be constructed for each time period in order to improve performance in times of both high and low volatility.

Ang and Bekaert (2002, 2004) noted that as financial markets go through cycles of low and high volatility, equity returns tend to be more strongly correlated during times of high volatility than in the less turbulent times. This strong correlation tends to occur internationally, and the authors question the benefit of international diversification. Because equity returns tend to go through periods of low and high volatility, they tried fitting a regime switching model, based on the work of Hamilton (1989), to equity returns. They showed that applying different mean-variance optimizations to the different regimes can improve a portfolio's performance, compared to a portfolio based on one simple mean-variance optimization. The optimal asset allocation during a low volatility regime is likely to be different to the optimal asset allocation during a high volatility regime, and a portfolio manager can add value by utilizing a regime switching model.

Kritzman & Li (2010) noted that in turbulent periods the return-to-risk ratio is substantially lower than in non-turbulent periods, no matter what the source of the turbulence is. When comparing turbulent periods, such as the global financial crisis of 2008, which began in the housing market, to the technology bubble collapse in 2000, and the Russian default of 1998 which contributed to the collapse of Long Term Capital Management, they found that no matter where the turbulence originated, it quickly spread throughout the financial sector, and would persist for several weeks, or longer. Since it is impossible to predict where the next financial crisis will originate, a direct result of the finding that the origin of a crisis does not matter is that previous turbulent periods can be used to stress-test portfolios more accurately than conventional methods. If traditional risk measurements, such as value at risk, is estimated using turbulent periods, rather than at the end of the investment horizon, exposures to loss can be more reliably be estimated. Mean-variance optimization can be improved by leveraging the information contained the returns and covariances observed during earlier times of financial turbulence. By blending the information from earlier turbulent periods with the in-sample expected returns and covariances, the calculation of expected returns and variance can be improved, and investors will be able to build more turbulence-resistant portfolios that also performs well under normal market conditions.

3. Method and data

3.1 Method

3.1.1 The regime switching model

Many economic and financial time series go through periods where they show different types of behavior in terms of mean and volatility. Typically, a time series will be in a particular state characterized by its mean and volatility, until it experiences a structural break, after which it will be in a state characterized by a different mean and volatility. This type of behavior makes it difficult to model the times series using a linear model. Hamilton's (1989) seminal work represents the first time where the behavior of economic variables could be influenced by economic recessions and expansions. Utilizing an unobserved Markov chain, he showed that the growth in real output could shift between one of two autoregressions, depending on whether the economy is in a recession or in an expansion.

In a regime switching model, the dependent variable is undergoing a regime switch when the parameters of the model jump between two or more constant values. These values are conditional on a state variable given by the current regime. The state variable produces estimates of the conditional probability of being in a particular state at a particular point in time. The time varying estimates of the conditional probability gives us estimates for the state transition probabilities for each state. This can be represented in a matrix form, commonly referred to as the transition matrix.

In a simple model of two regimes, a time series y_t switches between the regimes according a variable, z_t which can take one of two values. If $z_t = 1$, the process is in regime 1, and if $z_t = 2$, the process is in regime 2. The switch between the states follows a first order Markov process:

$$prob[z_t = 1|z_t = 1] = \pi_{11}$$
$$prob[z_t = 2|z_t = 1] = 1 - \pi_{11}$$
$$prob[z_t = 2|z_t = 2] = \pi_{22}$$
$$prob[z_t = 1|z_t = 2] = 1 - \pi_{22}$$

where p_{11} and p_{22} denote the probability of being in the same regime as the previous time period. This can be represented as the transition matrix:

$$\prod = \begin{bmatrix} \pi_{11} & 1-\pi_{22} \\ 1-\pi_{11} & \pi_{22} \end{bmatrix}$$

Hamilton's regime switching model has shown to be a popular tool used in analyzing univariate timeseries. In its most simple case, a two-regime model based on an AR(1) process can be written as:

$$Y_t = \begin{cases} \mu_1 + \phi_1 Y_{t-1} + \sigma_1 \varepsilon_{1t}, & \varepsilon_{1t} \sim iid(0,1), \quad in\ state\ 1 \\ \mu_2 + \phi_2 Y_{t-1} + \sigma_2 \varepsilon_{2t}, & \varepsilon_{2t} \sim iid(0,1), \quad in\ state\ 2 \end{cases} \tag{3.1}$$

or simply:

$$Y_t = \mu_{s_t} + \phi_{s_t} Y_{t-1} + \sigma_{s_t} \varepsilon_{1t},\ \ \varepsilon_{1t} \sim iid(0,1) \tag{3.2}$$

where the regimes switch when s_t changes between 0 and 1. The parameters in the model is typically represented by a vector, θ:

$$\theta = \left(\mu_1, \mu_{2,}, \phi_1, \phi_2, \sigma_1, \sigma_2, \pi_{11}, \pi_{22}\right) \tag{3.3}$$

The Markov chain is commonly represented by a state indicator vector ξ_t whose ith element equals 1 if $s_t = i$ and 0 otherwise. In a two-regime chain the vector is

$$\xi_t = \begin{pmatrix} \xi_t^1 \\ \xi_t^2 \end{pmatrix} = \begin{cases} \begin{pmatrix} 1 \\ 0 \end{pmatrix}, if\ state\ 1\ rules\ at\ time\ t. \\ \begin{pmatrix} 0 \\ 1 \end{pmatrix}, if\ state\ 2\ rules\ at\ time\ t. \end{cases} \tag{3.4}$$

The states are not directly observable, and we cannot be sure which state rules at time t. However, we can assign conditional probabilities to ξ_t by utilizing all information up to time $t-1$ and the transition matrix. The conditional expectation on $\xi_{t|t-1}$ is the product of the transition matrix and the state indicator at time $t-1$:

$$\xi_{t|t-1} = \prod \xi_{t-1} \tag{3.5}$$

The model can be estimated using maximum likelihood or the Gibbs sampler. This study will utilize the MSwM package in R, which use maximum likelihood to estimate the model. Firstly, normality is assumed, starting with normal density function:

$$\varphi(x;\mu,\sigma^2)=\frac{1}{\sqrt{2\pi\sigma}}\exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right]$$

To set the initial conditions, a linear regression of the model is estimated to get the coefficients and standard deviations. In addition, $\hat{\pi}_{11}$ and $\hat{\pi}_{11}$ is set to 0.5. Then, starting at $t=1$, iterate through:

1. $f\left(Y_t \middle| X_t;\hat{\theta}\right)=\hat{\xi}^1_{t|t-1}\varphi\left(Y_t;\hat{\alpha}_1+\hat{\phi}_1X_t,\hat{\sigma}_1\right)+\hat{\xi}^2_{t|t-1}\varphi(Y_t;\hat{\alpha}_2+\hat{\phi}_2X_t,\hat{\sigma}_2)$.
2. $\xi_{t|t}=\begin{pmatrix}\hat{\xi}^1_{t|t}\\ \hat{\xi}^2_{t|t}\end{pmatrix}=\begin{pmatrix}\frac{\hat{\xi}^1_{t|t-1}\varphi\left(Y_t;\hat{\alpha}_1+\hat{\phi}_1X_t,\hat{\sigma}_1\right)}{f\left(Y_t \middle| X_t;\hat{\theta}\right)}\\ \frac{\hat{\xi}^2_{t|t-1}\varphi(Y_t;\hat{\alpha}_2+\hat{\phi}_2X_t,\hat{\sigma}_2)}{f\left(Y_t \middle| X_t;\hat{\theta}\right)}\end{pmatrix}$
3. $\hat{\xi}_{t+1|t}=\prod\hat{\xi}_{t|t}$
4. Set $t=t+1$ and repeat the process until $t=T$.

This procedure produces two outputs:

- a set of conditional densities $\left\{f\left(Y_t \middle| X_t;\hat{\theta}\right)\right\}_{t=1}^{T}$
- a set of conditional state probabilities $\left\{\hat{\xi}_{t|t}\right\}_{t=1}^{T}$

The parameters θ of the model can now be estimated by maximizing the log likelihood function:

$$\ln L(\theta)=\sum_{t=1}^{T}\ln f_t(Y_t|X_t;\theta) \qquad (3.6)$$

Following Hamilton's work, it has been demonstrated that the regime switching model readily can be expanded to multivariate models, see for example Krolzig (1997) and Ang(2002).

This study will estimate an AR(1) model for the S&P500 log returns:

$$r_t = \mu + \phi r_{t-1} + \varepsilon_t \tag{3.7}$$

where $E[r_t] = \mu$ and $\varepsilon_t \sim iid(0, \sigma^2)$.

Then a regime switching model will be fitted to the AR(1) process. Two additional regime switching modes will be estimated, where in the first model, the VIX and the TED spread will be added to the model as regressors. In the second additional model, the first day returns of IPOs are added to the model as a regressor. The regime switching models will have 2 regimes.

The performance of the different models will be compared by pseudo out-of-sample forecasting. The models will be estimated using data for the time period January 1990 to December 2019. The months of the first quarter of 2020 will be used as the testing period, and the models will be compared on the root mean squared forecast error produced by these observations:

$$RMSFE = \sqrt{\sum_{i=1}^{n} \frac{(\hat{r}_i - r_i)^2}{n}} \tag{3.8}$$

The performance of the models will also be compared by back testing the optimal risky portfolios created by each model. The optimal risky portfolios were created using mean-variance optimization for each regime. Optimal complete portfolios were not created, as the asset allocation between the risky portfolio and the risk-free asset will be dependent on the risk aversion of the individual investor. The optimal risky portfolio will be the same risky portfolio regardless of any investor’s risk aversion, as the risk aversion only will dictate the allocation choices between the optimal risky portfolio and the risk-free asset. Thus, calculating optimal complete portfolios for different levels of risk aversion would not be relevant in comparing the optimal risky portfolios create by each model.

The portfolios created will be constrained by not allowing for short selling.

$$\sum w_i = 1 \tag{3.9}$$

where w_i is the weight assigned to each asset.

Following the procedure seen in Ang & Bekaert (2004), the trading costs associated with the rebalancing needed for each regime shift will be ignored. In the real world, this is likely to greatly impact the performance of the models when back testing. However, ignoring the trading costs will simplify the analysis.

3.1.2. The GARCH model

In addition to comparing the performance between the regime switching models, the models will also be compared to a GARCH(1,1) model using pseudo out-of-sample forecasting. A distinctive feature of many financial assets is that the volatility of the asset's returns varies over time. Several discrete models have been proposed to model this behavior, such as the autoregressive conditional heteroscedasticity (ARCH), exponentially weighted moving average (EWMA) and the generalized autoregressive conditional heteroscedasticity (GARCH) models (Hull, 2018). These models differ in the weighing scheme regarding previously observed returns and volatility.

The volatility of an asset, σ_n, over m observations, is commonly set equal to the standard deviation of an asset's return, u_i, given by:

$$\sigma_n^2 = \frac{1}{m}\sum_{i=1}^{m} u_{n-i}^2 \qquad (3.10)$$

where u_i is the log returns given by:

$$u_i = \ln\left(\frac{S_i}{S_{i-1}}\right) \qquad (3.11)$$

Equation 3.10 appliess the same weights to every observation observed. This is generally not very useful in financial risk modelling, where an asset's volatility typically is more dependent on the most recent observations. One solution for applying more weight to the most recent observations is the ARCH model, given by:

$$\sigma_n^2 = \gamma V_L + \sum_{i=1}^{m} \alpha_i u_{n-i}^2 \tag{3.12}$$

where α_i is the weight given to the observation i periods ago, V_L is the long run variance and γ is the weight assigned to V_L.

The ARCH model is useful when behavior of an assets' volatility varies randomly. However, the volatility of financial assets tends to move between periods of low volatility and high volatility, often referred to as volatility clustering. The GARCH model was developed to better model this phenomenon. In the GARCH model, the volatility is given by:

$$\sigma_n^2 = \gamma V_L + \alpha u_{n-1}^2 + \beta \sigma_{n-1}^2 \tag{3.13}$$

where β is the weight given to previously observed volatility. Typically, when estimating a GARCH model, it is specified as:

$$\sigma_n^2 = \omega + \alpha u_{n-1}^2 + \beta \sigma_{n-1}^2 \tag{3.14}$$

where $\omega = \gamma V_L$.

A typical GARCH(p,q) model is estimated using the most recent p returns observed and the most recent q variances estimated. GARCH(1,1) is the most widely used model. The GARCH model is solved using maximum likelihood estimation.

$$L(\theta) = \prod_{t=1}^{m} \frac{1}{\sqrt{2\pi v_t}} e^{-\frac{u_t^2}{2v_t}} \tag{3.15}$$

where $v_t = \omega + \alpha u_{t-1}^2 + \beta \sigma_{t-1}^2$.

This study will utilize the R package rmgarch in estimating the GARCH model. The package will also provide forecasts for future variance and returns. Appendix B contains the empirical work done in this thesis using R.

3.2 Data

All data were collected for the time period the 1st of January 1990 to the 31st of March 2020. The reason for this time window is that data for the shortest time series, the VIX, starts on the 1st of January 1990. Data from the 1st of January 1990 to the 31st of December 2019 were used as the training data, and the data from the first quarter of 2020 were used for testing. The collection and processing of the data using R shown in appendix A.

The IPO data were collected from the Tomson Reuters Eikon platform, using the Equity Offerings app. The data was limited to initial public offerings in the US. All entries that had missing values for the returns on the first day of trading were removed. It is worth noting that the data from Eikon did have a lot of entries with missing values. No further investigation was done as to why the data was missing, they were simply removed. Entries that had an offering price of $5 or less were removed, on the presumption that these IPOs will not receive much attention, and thus be under analyzed by the market. This is normal procedure in the IPO literature, see for example Lowry (2001) and Ritter & Welch (2002). In order to have a dataset without a lot of missing values, the data were converted from daily returns to a monthly average return by summing up all the first day returns in a given month and divide it by the total number of IPOs in the dataset for that month. Converting the time series into monthly averages also solves the problem of dealing with days where there are more than one observed IPO.

The S&P 500 is a stock market index constructed by the largest 500 companies listed on American stock exchanges. It is assumed to be the index that best represents the US stock market. The S&P 500 data was downloaded from Yahoo Finance, using R's quantmod package.

The CBOE volatility index (VIX) is a measure of expected volatility in the markets over the next 30 days (Banerjee, Doran & Peterson, 2007). A higher VIX represents an expectation of higher volatility. The VIX data was downloaded from Yahoo Finance, using R's quantmod package.

The TED spread is the difference between the short-term US government debt and the interest rate on interbank loans. It is commonly regarded as a measure of credit risk that large international banks lend money to each other. The TED spread is calculated as the difference between the 3-Month LIBOR based on US dollars and the 3-Month Treasury Bill. The data for the TED spread was collected from the Federal Reserve Bank of St. Louis (FRED, 2020), as monthly data.

The risk-free rate of return is based on the 3-month US Treasury bill. The data was downloaded from FRED, using R's quantmod package. As the reported 3-month T-bill is in percent per year, it was converted to monthly returns by using the following formula:

$$r_{f,monthly} = \left(1 + \frac{Y_t}{100}\right)^{1/12} - 1$$

For portfolio optimization, the average of the entire time period, January 1990 to December 2019, was used.

The Vanguard Total Bond Market Index Fund Investor Shares is a fund that invests in U.S. Treasuries and mortgage backed securities of short-, medium-, and long maturities (Vanguard, 2020). The aim of the fund is to track the broad, market weighted bond index, the Bloomberg Barclays index. The data was downloaded from Yahoo Finance, using R's quantmod package.

By using the S&P 500 index and the Vanguard Total Bond Market Index Fund as the risky assets in portfolio optimization, the diversification already included in these assets are leveraged, and the full diversification required to eliminate the unsystematic risk is assumed to be achieved.

The evolution of the different data during the training period, January 1990 to December 2019 is shown in figure 3.1.

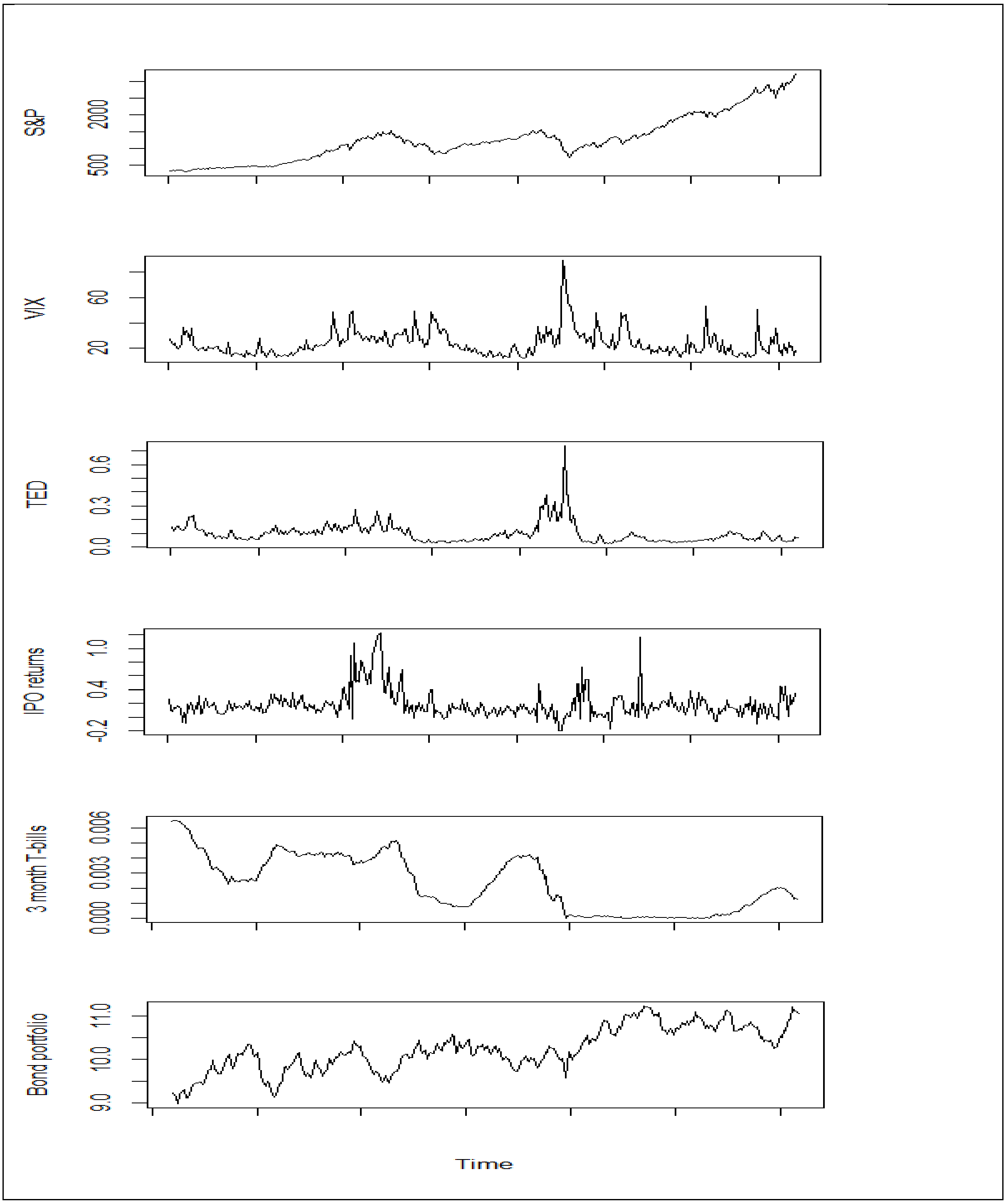

Figure 3.1. The evolution of the different data in the period January 1990 to December 2020

4. Empirical analysis

4.1 Results

4.1.1 Regime switching models

The full outputs from all the regime switching models provided by R is given in appendix C.

The S&P 500 no switching model

The results from the linear AR(1) model where the log returns of the S&P 500 are regressed on its first lags are shown in table 4.1. The estimated monthly log returns are 0.006185 with a standard error of 0.002199. The lagged S&P 500 log returns have a positive, not statistically significant, correlation with S&P 500 log returns. The Akaike information criterion, which is given in table 4.2, shows that the optimal number of lags for the model is 1 lag.

Linear S&P500 model

Estimation parameters

	Estimate	Std.error	t-value	Pr(>|t|)
μ	0.006185	0.002199	2.812	0.00519
φ	0.029268	0.052664	0.556	0.57872
R^2	0.04122			

Table 4.1 Estimated parameters of the S&P 500 AR(1) model

p	AIC
1	-1,269.724
2	-1,266.724
3	-1,263.724
4	-1,260.724
5	-1,257.724
6	-1,254.724
7	-1,251.724
8	-1,248.724
9	-1,245.724
10	-1,242.724

Table 4.2. Akaike information criteria, p = 1 results in the lowest AIC.

The S&P 500 regime switching model

The results of fitting a regime switching model to a linear model where the log returns of the S&P 500 are regressed on its first lags is given in table 4.3, and the time periods for when the model is in each regime is shown in figure 4.1.

S&P500 model

Estimation parameters

	Estimate	Std.error	t-value	Pr(>|t|)
μ_0	0.0129	0.0021	6.1429	8.103e-10
μ_1	0.0023	0.0039	0.5897	0.5554
φ_0	-0.1932	0.0828	-2.3333	0.0196
φ_1	0.0596	0.0741	0.8043	0.4212
σ_0	0.0218			
σ_1	0.0523			
R_0^2	0.0435			
R_1^2	0.0035			

Transition matrix

	Regime 1	Regime 2
Regime 1	0.963	0.030
Regime 2	0.037	0.970

Table 4.3. Results of the S&P 500 regime switching model. φ represents the coefficient of the lagged S&P 500 returns. Subscript 0 represents regime 1, and subscript 1 represents regime 2.

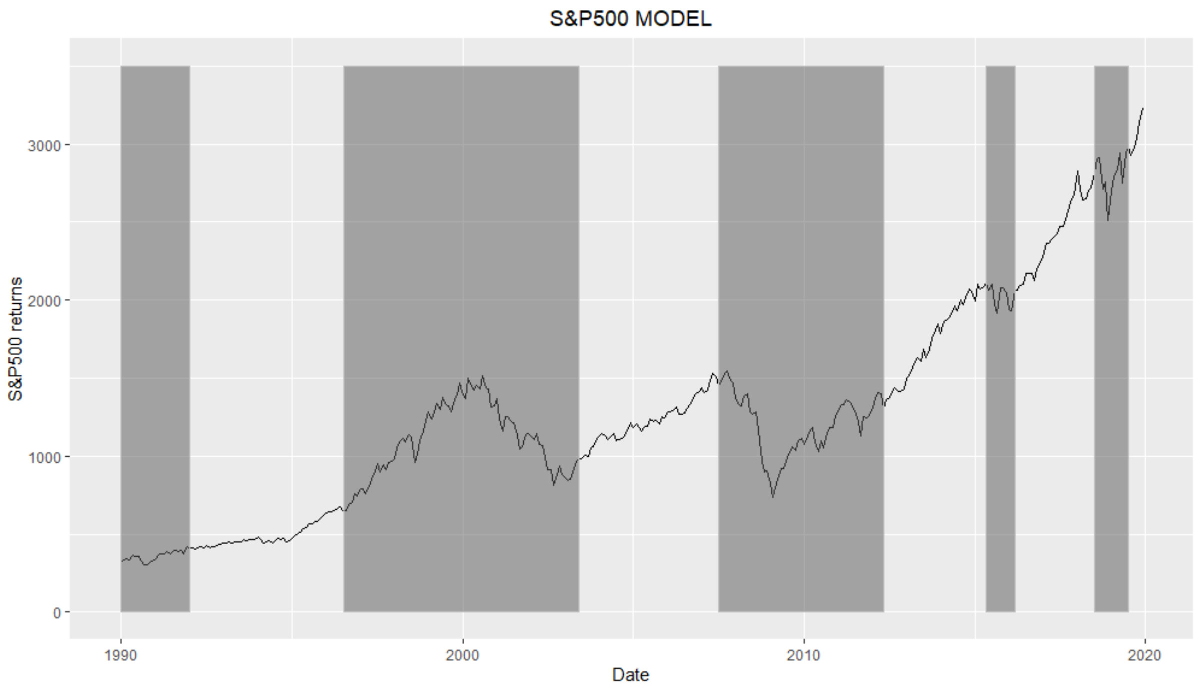

Figure 4.1. The S&P 500 index from 1990 to 2020. The shaded areas represent when the S&P500 model is in regime 1.

In regime 1 the estimated monthly log return of the S&P is 0.0129 with a standard error of 0.0021. The estimated volatility in regime 1 is 0.0218. In regime 2 the estimated monthly log return is 0.0023 with a standard error of 0.0039. The estimated volatility in regime 2 is 0.0523. The S&P 500 model successfully identified two regimes, where regime 1 is the regime with higher returns and lower volatility and regime 2 has lower returns and higher volatility. This is consistent with the "bad" / "normal" regime dichotomy reported in Ang & Bekaert (2004).

In regime 1 the lagged S&P 500 log returns have a negative, statistically significant, correlation with the S&P 500 log returns. In regime 2 the lagged S&P 500 log returns have a positive, statistically nonsignificant, correlation with the S&P 500 log returns.

According to the transition matrix, both regimes are very stable. The probability of staying in regime 1 is 0.97, and the probability of staying in regime 2 is 0.96.

The R-squared is very low in both regimes, 0.0435 in regime 1 and 0.0035 in regime 2. The model has more explanatory power in the high return, low volatility regime 1, than in regime 2. But overall the model explains very little of the variability of the S&P 500 log returns. This is in accordance with the weak form of the efficient market hypothesis, which states that the future cannot be predicted by analyzing the past (Fama, 1970).

The VIX-TED-S&P500 regime switching model

The results of fitting a regime switching model to a linear model where the log returns of the S&P 500 are regressed on its first lags and the first lags of the VIX and the TED spread is given in table 4.4. The time periods for when the model is in each regime is shown in figure 4.2.

VIX-TED-S&P500 model

Estimation parameters

	Estimate	Std.error	t-value	Pr(>|t|)
μ_0	-0.0146	0.0055	-2.6545	0.0079
μ_1	0.0099	0.0122	0.8115	0.4171
φ_{10}	-0.1767	0.0558	-3.1667	0.0015
φ_{11}	-0.0779	0.1148	-0.6786	0.4974
φ_{20}	0.0016	0.0003	5.3333	9.64e-08
φ_{21}	-0.0007	0.0005	-1.4000	0.1615
φ_{30}	0.0003	0.0005	0.6000	0.5485
φ_{31}	-0.0006	0.0006	-1.0000	0.3173
σ_0	0.0212			
σ_1	0.0464			
R_0^2	0.3620			
R_1^2	0.0547			

Transition matrix

	Regime 1	Regime 2
Regime 1	0.699	0.211
Regime 2	0.301	0.789

Table 4.4. Results of the VIX-TED-S&P regime switching model. φ_{1j} represents the coefficient of the lagged S&P 500 returns, φ_{2j} represents the coefficient of the lagged VIX and φ_{3j} represents the coefficient of the lagged TED spread. Subscript $i0$ represents regime 1, and subscript $i1$ represents regime 2.

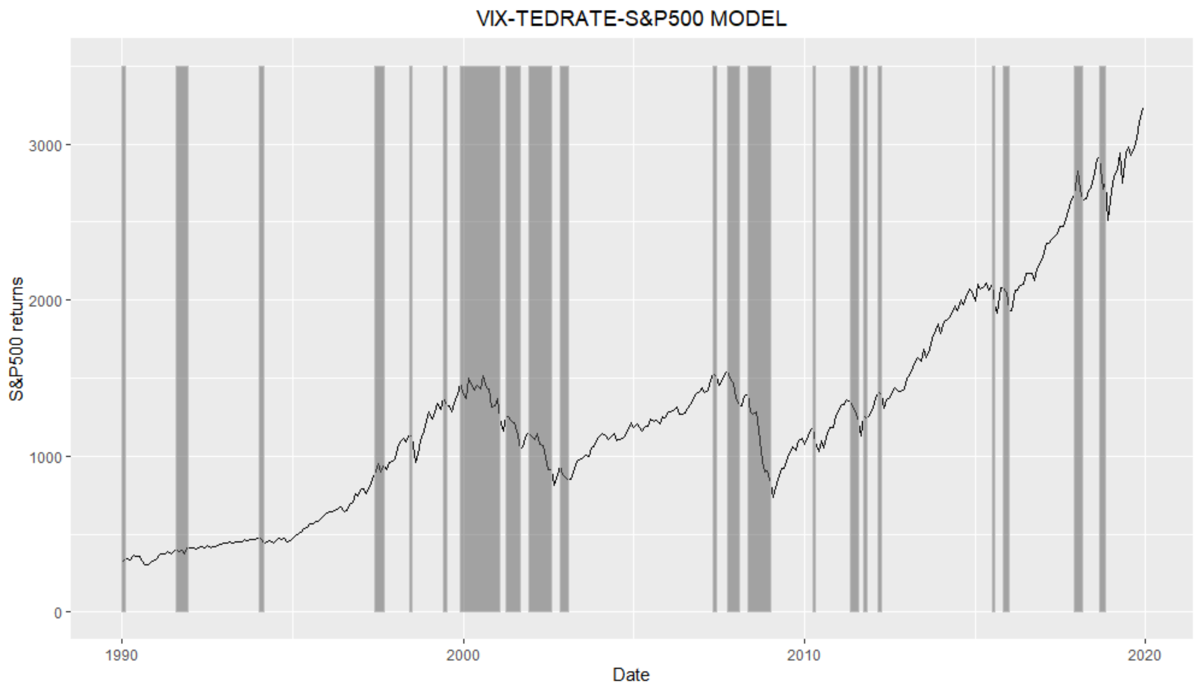

Figure 4.2. The S&P 500 index from 1990 to 2020. The shaded areas represent when the VIX-TED-SP model is in regime 1.

In regime 1 the estimated monthly log return of the S&P 500 is -0.0146 with a standard error of 0.0055. The estimated volatility in regime 1 is 0.0212. In regime 2 the estimated monthly log return is 0.0099 with a standard error of 0.0122. The estimated volatility in regime 2 is 0.0464. The model identify regime 1 as a negative return, low volatility regime, and regime 2 as a positive return, high volatility regime.

In regime 1, the lagged S&P 500 log returns have a statistically significant negative correlation to the S&P 500 log returns, whereas the lagged VIX and TED spread have a positive correlation. The VIX is statistically significant, the TED spread is not statistically significant. In regime 2, the lagged S&P 500 log returns, VIX and TED spread all have a negative correlation to the S&P log returns. None of these coefficients are statistically significant.

The R-squared is relatively high in regime 1, and the model is able to explain quite a lot of the variability of the S&P 500 log returns in regime 1. In regime 2, the R-squared is low, and the model is only able to explain a little bit of the variability of the S&P 500 log returns. According to the transition matrix, both regimes are quite unstable, with regime 1 being the least stable. The probability of staying in regime 1 is 0.699, and the probability of staying in regime 2 is 0.789.

The IPO-VIX-TED-S&P500 regime switching model

The results of fitting a regime switching model to a linear model where the log returns of the S&P 500 are regressed on its first lags and the first lags of the VIX, the TED spread and first day IPO returns is given in table 4.5. The time periods for when the model is in each regime is shown in figure 4.3.

IPO-VIX-TED-S&P500 model

Estimation parameters

	Estimate	Std.error	t-value	Pr(>|t|)
μ_0	0.0067	0.0142	0.4718	0.6371
μ_1	-0.0153	0.0054	-2.8333	0.0046
φ_{10}	-0.0755	0.1246	-0.6059	0.5446
φ_{11}	-0.1783	0.0531	-3.3578	0.0008
φ_{20}	-0.0008	0.0005	-1.6000	0.1096
φ_{21}	0.0017	0.0003	5.6667	1.46e-08
φ_{30}	-0.0005	0.0005	-1.0000	0.3173
φ_{31}	0.0001	0.0005	0.2000	0.8415
φ_{40}	0.0142	0.0309	0.4595	0.6459
φ_{41}	0.0145	0.0111	1.3063	0.1915
σ_0	0.0450			
σ_1	0.0203			
R_0^2	0.0629			
R_1^2	0.4025			

Transition matrix

	Regime 1	Regime 2
Regime 1	0.671	0.249
Regime 2	0.329	0.751

Table 4.5. Results of the IPO-VIX-TED-S&P regime switching model. φ_{1j} represents the coefficient of the lagged S&P 500 returns, φ_{2j} represents the coefficient of the lagged VIX, φ_{3j} represents the coefficient of the lagged TED spread and φ_{4j} represents the coefficient of the lagged IPO first day returns. Subscript $i0$ represents regime 1, and subscript $i1$ represents regime 2.

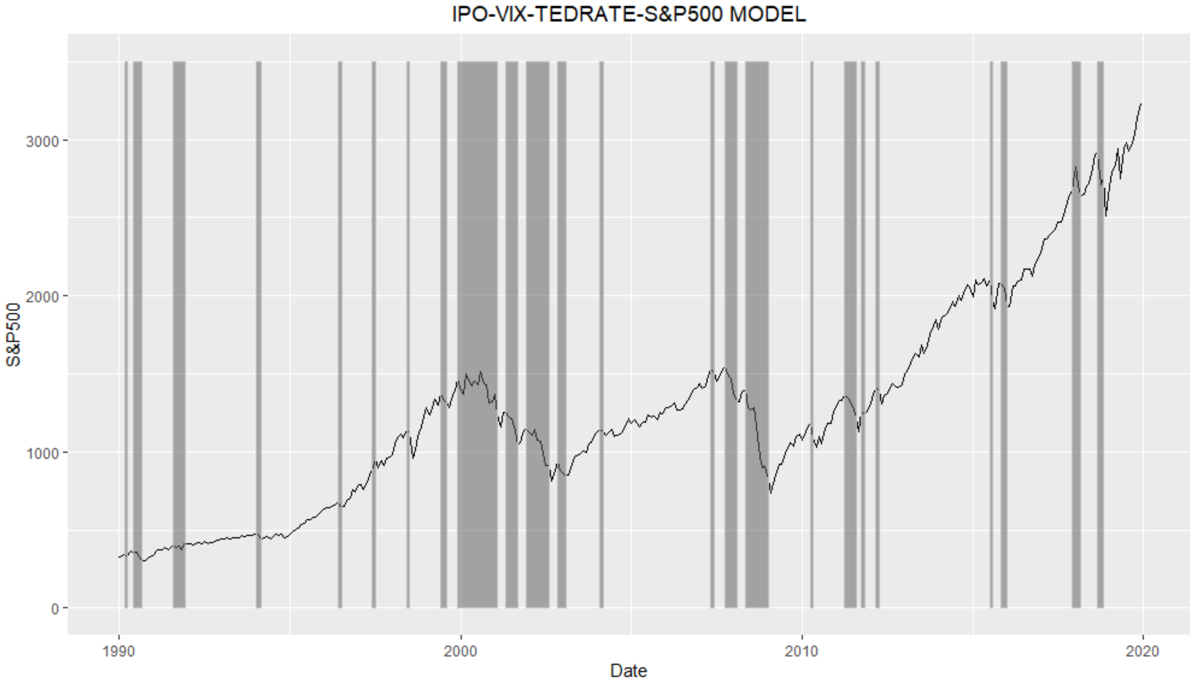

Figure 4.3. The S&P 500 index from 1990 to 2020. The shaded areas represent when the IPO-VIX-TED-SP model is in regime 2.

In regime 1, the estimated monthly log return of the S&P is 0.0067 with a standard error of 0.0142. The estimated volatility in regime 1 is 0.0450. In regime 2 the estimated monthly log return is -0.0153 with a standard error of 0.0054. The estimated volatility in regime 2 is 0.0203. The model identify regime 1 as a positive return, high volatility regime, and regime 2 as a negative return, low volatility regime.

In regime 1 the lagged S&P log returns, VIX and TED spread have a negative statistically non-significant correlation to the S&P log returns. The lagged IPO first day returns have a not statistically significant positive correlation to the S&P 500 log returns. In regime 2 the lagged S&P log returns have a statistically significant negative correlation to the S&P 500 log returns. The lagged VIX has a statistically significant positive correlation to the S&P 500 log returns. The TED spread and first day IPO returns have statistically non-significant positive correlation to the S&P 500 log returns.

The R-squared in regime 1 is low, 0.0629, so the model explains very little of the variability of the S&P 500 log returns in regime 1. In regime 2, the R-squared is high, 0.4025, so the model is able to explain a lot of the variability in regime 2.
According to the transition matrix, both regimes are quite unstable, with regime 1 being the most unstable. The probability of staying in regime 1 is 0.671, and the probability of staying in regime 2 is 0.751.

Pseudo out sample forecasting

The forecasted log returns of the S&P 500 for the first quarter of 2020 produced by the regime switching models are given in table 4.6. The only model that forecast negative returns is the VIX-TED-S&P500 model, the other models forecast positive returns for the period. The root mean squared forecast errors (RMSFE) for the regime switching models are given in table 4.7. The model with the lowest RMSFE, and that performed the best during the forecast period is the VIX-TED-S&P500 model. This is as expected since this is the only model that forecasted a decline in the S&P 500, and as is now known, the financial markets worldwide experienced large declines during the first quarter of 2020 due to the coronavirus lockdown.

	Jan 2020	Feb 2020	Mar 2020
S&P500 model log returns	0.0074	0.0132	0.0298
VIX-TED-S&P500 model log returns	-0.0096	-0.0083	-0.0215
IPO-VIX-TED-S&P500 model log returns	0.0158	0.0276	0.0879
Actual log returns	-0.0016	-0.0879	-0.1175

Table 4.6. Forecasted log returns and actual log returns for the first quarter of 2020.

	S&P500	VIX-TED-S&P500	IPO-VIX-TED-S&P500
RMSFE	0.1033	0.0722	0.1364

Table 4.7. Root mean squared errors of the models during the first quarter of 2020

4.1.2 Portfolio optimization

The no-switching model

The efficient frontier of the no switching portfolios is given in figure 4.4. The optimal risky portfolio is marked with the triangle, and the tangent represents the optimal capital asset line. The Sharpe ratio of the optimal portfolio, and its asset weights is given in figure 4.5. The optimal weights for the no-switching model is 53.63% invested in the S&P 500 index and 46.37% in the bond portfolio, resulting in a Sharpe of 0.1863.

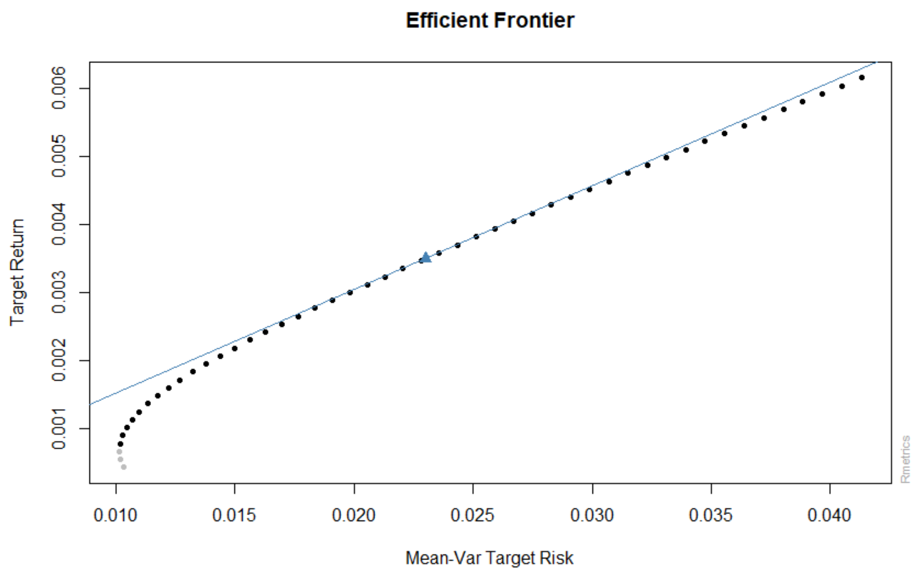

Figure 4.4. The efficient frontier of the no switching model.

Expected log returns	0.0420
StDev	0.0797
Sharpe	0.1863

Weights

S&P 500	Vanguard
0.5363	0.4637

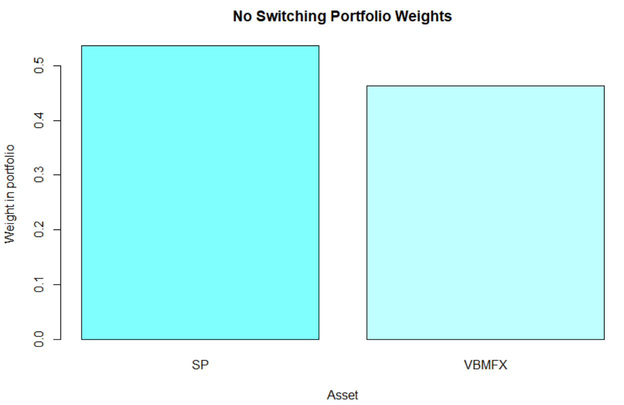

Figure 4.5. Sharpe and weights of the optimal risky portfolio.

The S&P 500 model regime switching model

Regime 1

The efficient frontier of the portfolios in regime 1 of the S&P 500 regime switching model is given in figure 4.6. The optimal risky portfolio is marked with the triangle, and the tangent represents the optimal capital asset line. The Sharpe ratio of the optimal portfolio, and its asset weights is given in figure 4.7. The optimal weights for regime 1 of the S&P 500 model is to be 100% invested in the S&P 500 index, resulting in a Sharpe of 1.1662.

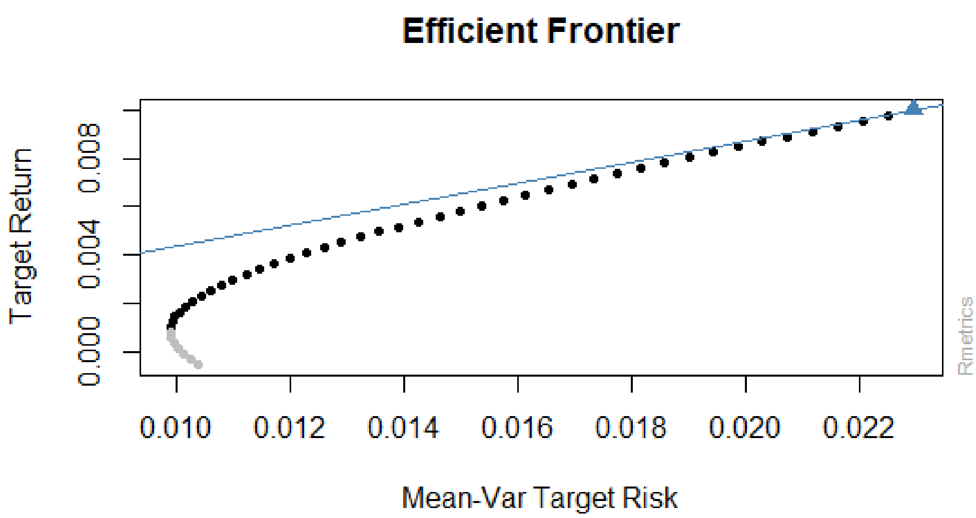

Figure 4.6 The efficient frontier of regime 1 of the S&P500 regime switching model.

Expected log returns		0.1199
StDev		0.0795
Sharpe		1.1662

Weights

	S&P 500	Vanguard
	1.0000	0.0000

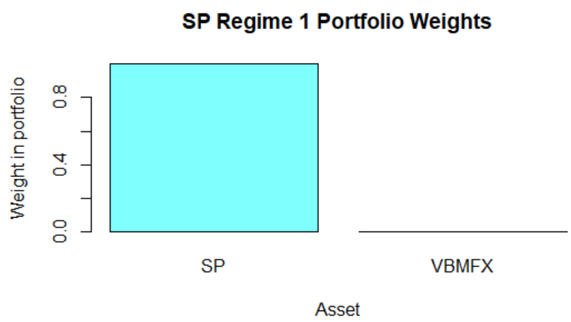

Figure 4.7. Sharpe and weights of the optimal risky portfolio.

Regime 2

The efficient frontier of the portfolios in regime 2 of the S&P 500 regime switching model is given in figure 4.8. The optimal risky portfolio is marked with the triangle, and the tangent represents the optimal capital asset line. The Sharpe ratio of the optimal portfolio, and its asset weights is given in figure 4.9. The optimal weights for regime 2 of the S&P 500 model is to be 3.89% invested in the S&P 500 index and 96.11% invested in the bond portfolio, resulting in a Sharpe of 0.5360.

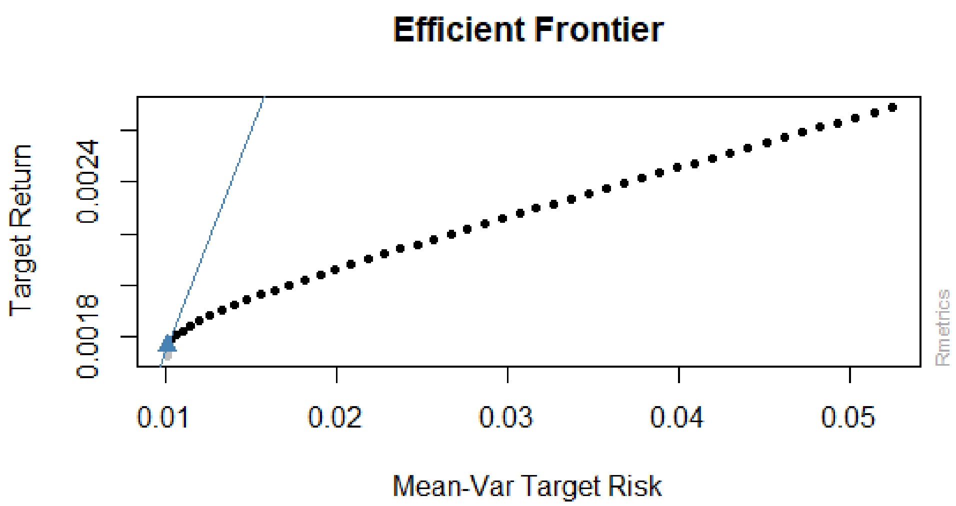

Figure 4.8. The efficient frontier of regime 2 of the S&P500 regime switching model.

Expected log returns	0.0212
StDev	0.0353
Sharpe	-0.1696

Weights

S&P 500	Vanguard
0.0389	0.9611

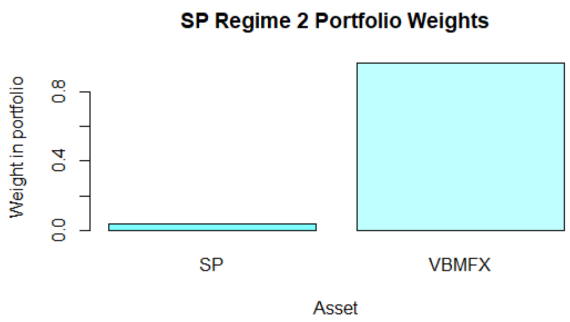

Figure 4.9. Sharpe and weights of the optimal risky portfolio.

The VIX-TED-S&P 500 model

Regime 1

The efficient frontier of the portfolios in regime 1 of the VIX-TED-S&P 500 regime switching model is given in figure 4.10. The optimal risky portfolio is marked with the triangle, and the tangent represents the optimal capital asset line. The Sharpe ratio of the optimal portfolio, and its asset weights is given in figure 4.11. The optimal weights for regime 1 of the VIX-TED-S&P500 model is to be 68.73% invested in the S&P 500 index and 31.27% invested in the bond portfolio, resulting in a Sharpe of 0.6440.

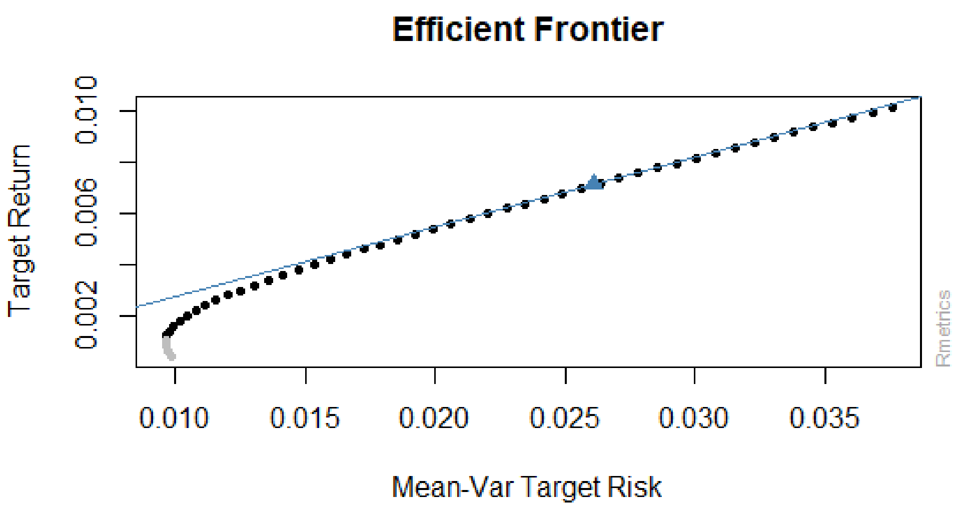

Figure 4.10. The efficient frontier of regime 1 of the VIX-TED-S&P500 regime switching model.

Expected log returns		0.0854
StDev		0.0905
Sharpe		0.6440
Weights		
	S&P 500	Vanguard
	0.6873	0.3127

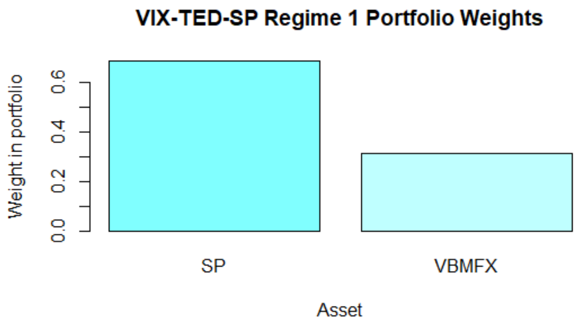

Figure 4.11. Sharpe and weights of the optimal risky portfolio.

Regime 2

The efficient frontier of the portfolios in regime 2 of the VIX-TED-S&P 500 regime switching model is given in figure 4.12. Because most of the points have negative returns, the figure is not able to mark the optimal risky portfolio and optimal capital asset line. The Sharpe ratio of the optimal portfolio, and its asset weights is given in figure 4.13. The optimal weights for regime 2 of the VIX-TED-S&P500 model is to 100% invested in the bond portfolio, resulting in a negative Sharpe of -0.4912, which means that this portfolio performs worse than the risk-free asset.

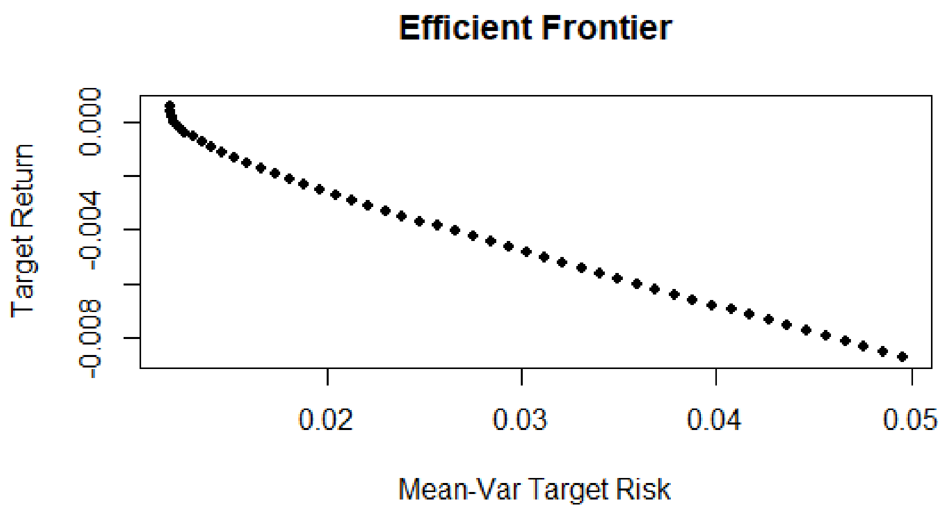

Figure 4.12. The efficient frontier of regime 2 of the VIX-TED-S&P500 regime switching model.

Expected log returns	0.0069
StDev	0.0412
Sharpe	-0.4912

Weights

S&P 500	Vanguard
0.0000	1.0000

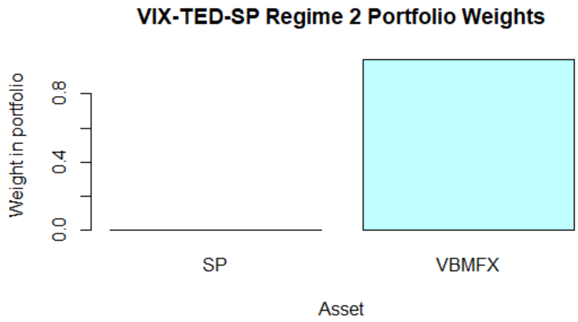

Figure 4.13. Sharpe and weights of the optimal risky portfolio.

The IPO-VIX-TED-S&P 500 model

Regime 1

The efficient frontier of the portfolios in regime 1 of the IPO-VIX-TED-S&P 500 regime switching model is given in figure 4.14. Because most of the points have negative returns, the figure is not able to mark the optimal risky portfolio and optimal capital asset line. The Sharpe ratio of the optimal portfolio, and its asset weights is given in figure 4.15. The optimal weights for regime 1 of the IPO-VIX-TED-S&P500 model is to be 100% invested in the bond portfolio, resulting in a negative Sharpe of -0.4030, and the portfolio is performing worse than the risk-free asset.

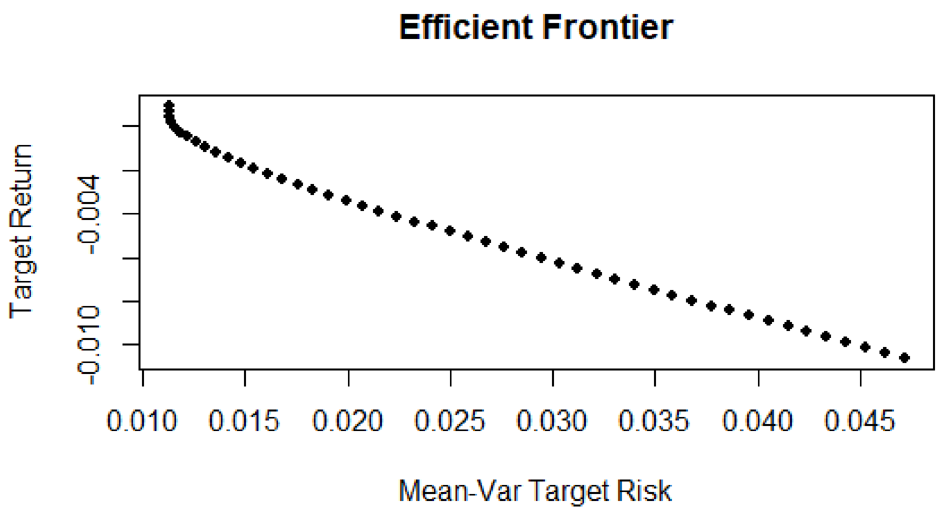

Figure 4.14. The efficient frontier of regime 1 of the IPO-VIX-TED-S&P500 regime switching model.

Expected log returns	0.0113
StDev	0.0393
Sharpe	-0.4030

Weights

S&P 500	Vanguard
0.0000	1.000

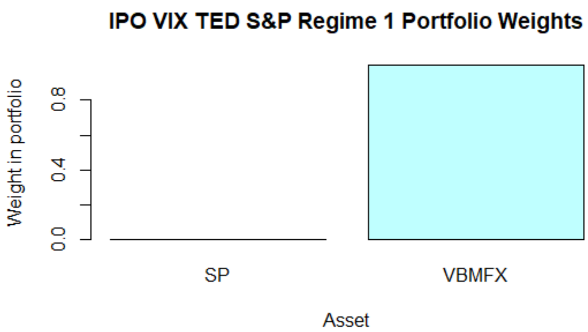

Figure 4.15. Sharpe and weights of the optimal risky portfolio.

Regime 2

The efficient frontier of the portfolios in regime 2 of the IPO-VIX-TED-S&P 500 regime switching model is given in figure 4.16. The optimal risky portfolio is marked with the triangle, and the tangent represents the optimal capital asset line. The Sharpe ratio of the optimal portfolio, and its asset weights is given in figure 4.17. The optimal weights for regime

2 of the IPO-VIX-TED-S&P500 model is to be 84.55% invested in the S&P 500 index and 15.45% invested in the bond portfolio, resulting in a Sharpe of 0.7070.

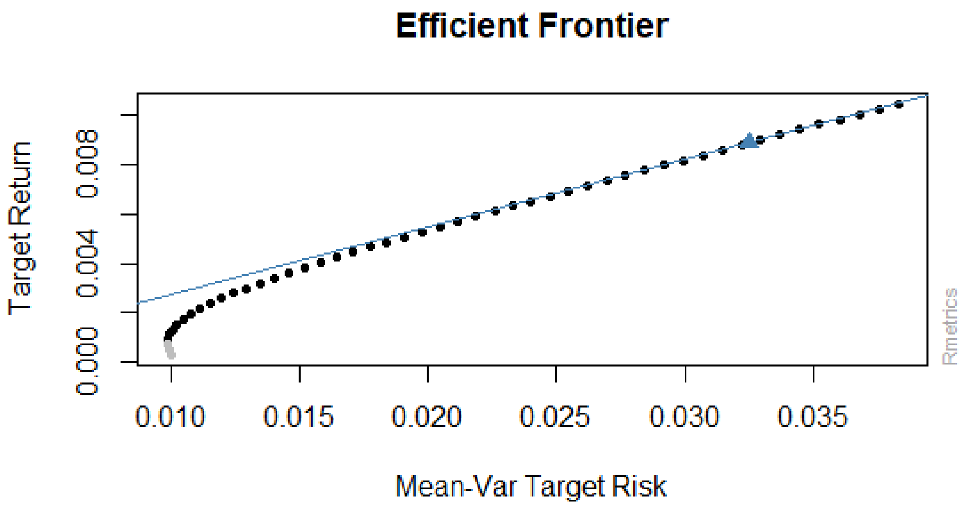

Figure 4.16. The efficient frontier of regime 2 of the IPO-VIX-TED-S&P500 regime switching model.

Expected log returns		0.1068
StDev		0.1127
Sharpe		0.7070
Weights		
	S&P 500	Vanguard
	0.8455	0.1545

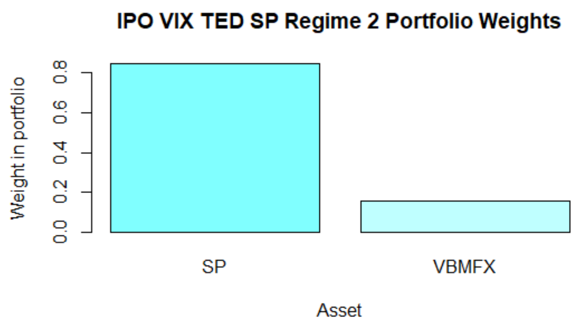

Figure 4.17. Sharpe and weights of the optimal risky portfolio.

Holding period returns

The holding period returns (HPR) is given in table 4.8. In terms of the holding period returns, the best performing model is the IPO-VIX-TED-S&P 500 regime switching model with a HPR of 313.53%, which is marginally better that the S&P500 regime switching model with a HPR of 313.23%.

	No switching model	S&P500 model	VIX-TED-S&P500 model	IPO-VIX-TED-S&P500 model
Log HPR	1.2002	1.4188	1.23127	1.4196
HPR	2.3206	3.1323	2.4253	3.1353

Table 4.8. Holding period returns for the models.

The log HPR was converted to simple HPR by:

$$HPR = \exp(log\,HPR) - 1 \tag{4.1}$$

4.1.3 GARCH(1,1) model

The estimated parameters of GARCH(1,1) model is given in table 4.9. All the estimated coefficients are statistically significant at a 95% significance level, giving the model:

$$\sigma_{t-1}^2 = 0.000077 + 0.188700u_{t-1}^2 + 0.777092\sigma_{t-1}^2 \tag{4.2}$$

	Estimate	Std.Error	t-value	Pr(>|t|)
mu	0.007734	0.001612	4.7977	0.000002
ar1	-0.073841	0.060114	-1.2283	0.219319
omega	0.000077	0.000038	2.0373	0.0416202
alpha1	0.188700	0.050186	3.7600	0.000170
beta1	0.777091	0.050932	15.2574	0.000000

Table 4.9. GARCH parameters

In figure 4.18 the squared residuals and the estimated conditional variance estimated by the GARCH(1,1) model in during the training period is shown. The plot has a typical GARCH-patterns, where there are periods with high variance clustered together and periods of relatively low variance. The volatility spikes occurs where they are expected to occur, in the early 1990's with the oil price shock, late 1990's with the Asian financial crisis and the Russian financial crisis, in the early 2000's with the dotcom bubble and in the late 2000's with the financial crisis. During the two years preceding 2020, volatility starts to rise again, after a period characterized by relatively low volatility.

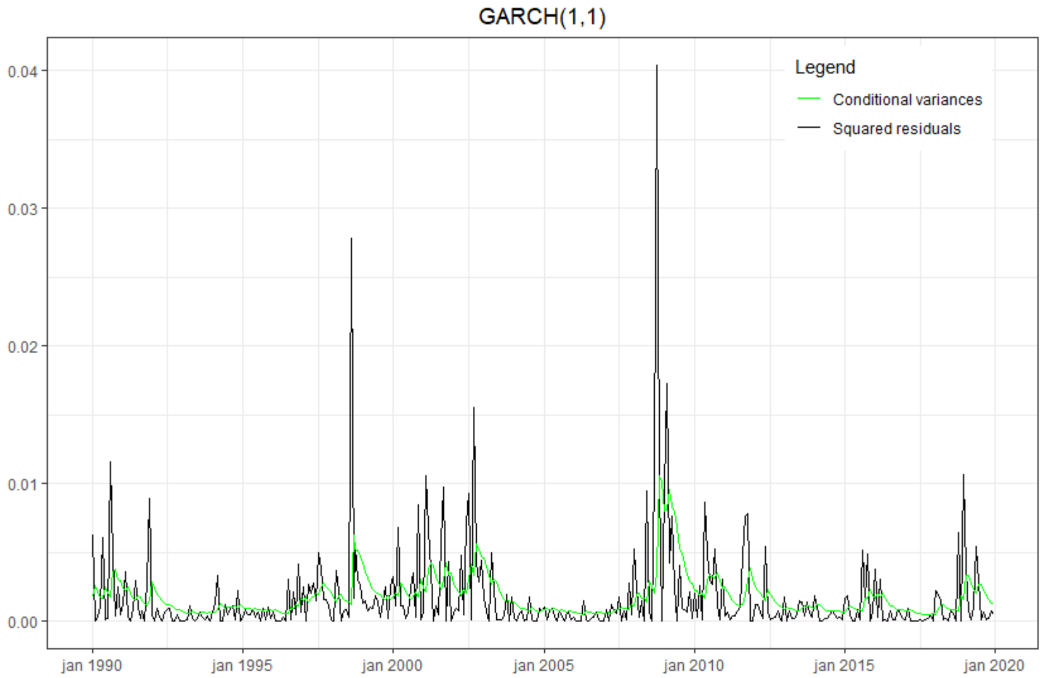

Figure 4.18. GARCH model of the training period, January 1990 to December 2019.

The forecasted variance for the first quarter of 2020, together with the variance seen during 2019, is shown in figure 4.19. As seen, the forecasted variance is not expected to change much from the conditional variance observed in December 2019. Figure 4.20 shows the squared residuals and the estimated conditional variance for the complete time period of the data, from January 1990 to March 2020. The plot shows that the conditional variance is increasing a lot during the first quarter of 2020, due the coronavirus recession. The fall in the global stock markets started on the 20th of February. From the 24th to the 28th of February many stock markets around the world reported the worst one-week performance since the financial crisis. On March 8th an oil price war broke out between Russia and OPEC, causing further falls in the global financial markets. The following week the S&P 500 and Dow Jones Industrial Average hit "circuit breakers" on several occasions, meaning that the trading will be halted to stop panic selling. These events led to a variance that is a lot higher than the variance forecasted by the model.

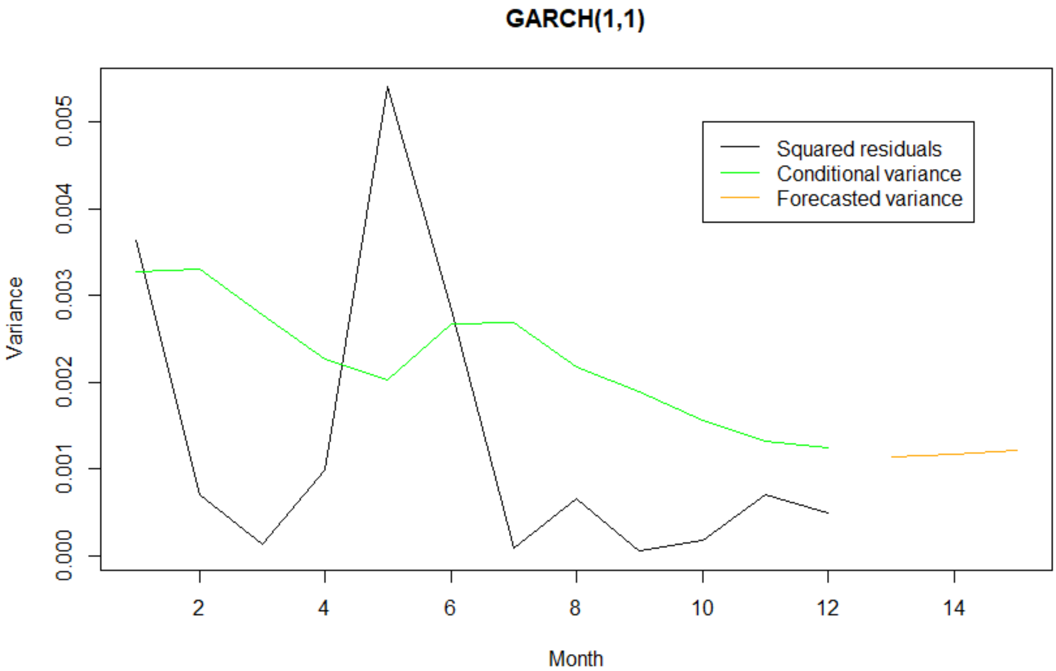

Figure 4.19. The variance estimated from the training period GARCH model, zoomed in on the year 2019, together with the forecasted variance for the first quarter of 2020. Observation 13, 14 and 15 corresponds to January, February and March 2020, respectively.

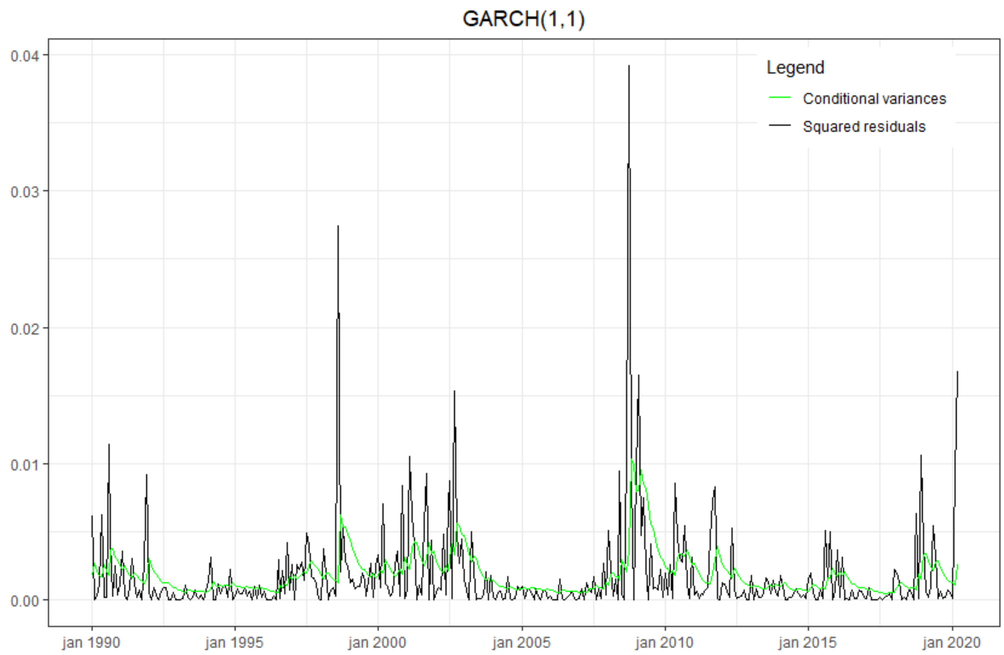

Figure 4.2. GARCH model with the first quarter of 2020 included

The log returns forecasted by the GARCH(1,1) model for the first quarter of 2020 is given in table 4.10, together with the observed log returns and the calculated root mean square forecast error.

	Jan 2020	Feb 2020	March
Estimated log returns	0.0062	0.0078	0.0077
Observed log returns	-0.0016	-0.0879	-0.1175
Forecast error	1e-04	0.00916	0.01569
RMSFE			0.0911

Table 4.10. Forecast error of returns based on the GARCH(1,1) model

4.2 Discussion

As shown in table 4.5 for the IPO-VIX-TED-S&P500 regime switching model, the lagged first day IPO returns have a positive correlation to the S&P 500 log returns in both regimes, meaning that if the first day returns of the IPOs are negative, they would have a negative impact on the performance of the S&P 500. However, the coefficients are not statistically significant. In regime 2, which is the regime with negative mean log returns of the S&P 500, the coefficient has a p-value of 0.1915. In this regime it has lower likelihood of being created by noise, but it is still outside of the traditional levels of significance at 90% significance, or above.

When looking at the pseudo out-of-sample forecasts in table 4.7, the model that best forecasted the log returns of the first quarter of 2020 is the VIX-TED-S&P500RS model, and the worst performing model is the IPO-VIX-TED-S&P500RS model. Based on the RMSFEs, adding the traditional risk metrics as regressors to the S&P500RS model will improve the forecasts for the testing period used in this thesis. However, adding the IPO-returns as a regressor to the VIX-TED-S&P500RS model does not add value to the forecast. In fact, the forecast got worse when the IPOs were added.

The R-squared is generally low for the regime switching models, and the models were generally not able to explain a lot of the variation observed in the S&P 500 log returns. This supports the weak form of the efficient market hypothesis, which states that the future cannot be predicted by studying what happened in the past. However, in the negative returns regime of the VIX-TED-S&P500RS model and the IPO-VIX-TED-S&P500RS model the R-square was 0.3620 and 0.4025, respectively. And it seems that both of these models are able to explain quite a lot of the variation that is happening when the markets are falling. The model including the IPO returns does have a higher R-squared, but the act of adding more regressors to a model generally increase value of the R-squared even when the added regressor does not add any explanatory power (Stock & Watson, 2015). It is therefore not possible to say that the model including IPO returns has better explanatory power based on the value of the R-squared alone.

When comparing the pseudo out-of-sample forecasts off the regime switching models with the GARCH model, table 4.10, the regime switching model with the lowest RMSFE, the VIX-TED-S&P500RS model, outperforms the GARCH model. However, the GARCH model outperforms the other regime switching models.

The models in this study did not use a rolling window and were not updated during the testing period. In the normal world, the models would be re-estimated between each point in the testing period, which could have affected the forecasting performance, as the increase in volatility experienced during the testing period would be included in the models.

When comparing the holding period returns for the different models, table 4.8 shows that the S&P500RS model and the IPO-VIX-TED-S&P500RS model by far outperform the other two models. All the regime switching models outperformed the model that did not allow for switching. This shows that, for the time period that this study is concerned, applying a regime switching model to the S&P 500 log returns does add value, consistent with the previous research regarding regime switching models, see for example Ang & Bekaert (2002; 2004).

The IPO-VIX-TED-S&P500RS model outperformed the S&P500RS on the first decimal, 313.53% vs 313.23% returns respectively. This difference is negligible, and the trading costs are not taken onto account. When comparing figure 4.1 and figure 4.3, it is apparent that the IPO-VIX-TED-S&P500 RS model would lead to much higher trading costs, as it would require rebalancing of the portfolio 48 times, opposed to 9 rebalancing required by the S&P500RS model. The difference in stability of the models is also demonstrated in the transition matrixes, table 4.5-4.5, where the transition matrix of the S&P500 RS model shows that the stability of the regimes is very high, whereas for the other two regime switching models, the regimes are quite unstable, leading to many rebalancing required for these two models.

When comparing the Sharpe ratios, figures 4.5-4.17, the S&P500RS model have the superior Sharpe ratio in both regimes, compared to the other regime switching models, and its Sharpe ration is superior to the no-switching model in both regimes. The higher Sharpe ratio of the S&P500RS model compared to the IPO-VIX-TED-S&P500RS model, would allow an investor to apply leverage to the S&P500RS model without taking on more risk than the risk experienced by an investor using the IPO-VIX-TED-S&P500RS model. This leverage would increase the returns of the S&P500RS model, and lead to higher returns than the 0,3% higher returns experienced by the IPO-VIX-TED-S&P500RS model.

5. Conclusion

As noted in the data section, the IPO data had a lot of entries with missing values. The reason for this was not explored. Were they postponed, or did the offering not go through at all. This could be a topic for further research.

The results of the study showed that IPO returns had a non-statistically significant positive correlation with the S&P 500 log returns, table 4.5. In the negative returns regime, the IPO returns showed the highest level of significance, with a p-value of 0.1915.

When comparing the performance of the models using pseudo out-of-sample forecasts for the S&P 500 log returns during the first quarter of 2020, the model that included the IPO returns performed the worst. The best performing model in forecasting was the regime switching model that included the VIX and the TED spread as regressors. The VIX-TED-S&P500 regime switching model had a lower RMSFE value than all the models, including the GARCH model. The GARCH model however, outperformed the other regime switching models in the pseudo out-of-forecast.

When the regime switching models were back tested using portfolio optimization, the best performing model was the regime switching model that included the IPO returns, with a holding period return (HPR) of 313.53%. This model which marginally outperformed the S&P 500 regime switching model, with a HPR of 313.23%. These two models outperformed the regime switching model that included the VIX and TED spread as regressor and the model that did not allow for switching by a large margin, which had HPR of 242.53% and 323.06%

respectively. All the regime switching models outperformed the model that did not allow for regime switching, showing that in general regime switching models will outperform a linear model in portfolio optimization.

The difference between the HPR of the regime switching model that included IPO returns and the S&P 500 regime switching model was 0.3%. The back testing was done without considering transaction costs associated with the rebalancing required at each regime shift. The S&P 500 regime switching model required rebalancing 9 times during the period, whereas the regime switching model that included the IPO returns required rebalancing 48 times in the same time period. It is likely that the transaction costs associated with the rebalancing would eat up those 0.3% excess returns. In addition to transaction costs, the S&P 500 regime switching model had a higher Sharpe ratio in both regimes, meaning that the risk adjusted returns for this model were higher than for the IPO regime switching model. This means that almost the same HPR was achieved for the period, but at a much lower risk for the S&P 500 regime switching model.

The results in this study does not indicate that adding IPO returns to the modelling the S&P 500 returns add much value. The coefficients for the IPO returns were not statistically significant. The regime switching model did not have the best performance in the pseudo out-of-sample forecast. The optimal risky portfolio created by the IPO model was able to generate the highest holding period returns marginally beating the S&P 500 regime switching model, however this was with a much lower Sharpe ratio, so the risk adjusted returns for this portfolio was lower than for the S&P 500 ratio.

This study used a different time horizon than most other studies on the interaction between the performance of IPOs and the performance of the stock market. Other studies tend to use longer time horizons, but this study was limited by the time horizon of the VIX index. This study was simplified by using monthly data, as opposed to using daily data commonly seen in research regarding IPOs. The IPO dataset collected from the Tomson Reuters Eikon database did have a lot of missing data. Further research could be done by investigating these planned offerings, and see if there is useful information contained in identifying if initial public offerings get postponed or cancelled.

References

Ang, A & Bekaert, G. (2002). International asset allocation with regime shifts. *The Review of Financial Studies*. 15(4), 1137-1187.

Ang, A & Bekaert, G. (2004). How regimes affect asset allocation. *Financial Analyst Journal*. 60(2), 86-99.

Banerjee, P. S., Doran, J. S. & Peterson, D. R. (2007). Implied volatility and future returns. *Journal of Banking and Finance, 31,* 3183-3199.

Baker, M. & Wurgler, J. (2000). The equity shares in new issues and aggregate stock returns. *The Journal of Finance*. 55(5), 2219-2257.

Baron, D. P. (1982). A model of the demand for investment banking advising and distributing services for new issues under asymmetric information: Delegation and the information problem. *Journal of Finance*. 35, 1115-1138.

Bloomberg. (2019, October 12th). Uber, Peloton: Unprofitable Companies Offer IPOs At Record Rate. Collected from. https://www.bloomberg.com/graphics/2019-unprofitable-ipo-record-uber-wework-peloton/

Bollerslev, T. (1986). Generalized autoregressive conditional heteroscedasticity. *Journal of Econometrics*, 31, 307-327.

Braun, M & Larrain, B. (2009). Do IPOs affect the prices of other stocks? Evidence from emerging markets. *Review of Financial Studies*. 22(4), 1505-1544.

Chauvet, M., Senyuz, Z. & Yoldas, E. (2015). What does financial volatility tell us about macroeconomic fluctuations? *Journal of Economic Dynamics & Control*. 52. 340-360.

Chicago Board Options Exchange. (2019, October 9th.). VIX-index. Collected from http://www.cboe.com/vix

Chow, G., Jacquier, E., Kritzman, M. & Lowry, K. (1999). Optimal portfolios in good times and bad. *Financial Analyst Journal*. 55(3), 65-73.

Christiansen, C., Schmeling, M. & Schrimpf, A. (2012). A comprehensive look at financial prediction by economic variables. *Journal of Applied Econometrics*. 27(6), 956-977.

Diebold, F. X. & Rudebusch, G. D. (1996). Measuring business cycles: A modern perspective. *The Review of Economics and Statistics, 78*(1), 76-77.

Eckbo, B. E. (2007). *Handbook of Corporate Finance*. Amsterdam: Elsevier B.V.

Evans, J. L. & Archer, S. H. (1968). Diversification and the reduction of dispersion: an empirical analysis. *The Journal of Finance*. 23(5), 761-767.

Fama, E. F. (1970). Efficient Capital Markets: A Review of Theory and Empirical Work. *The Journal of Finance*. 25(2). 383-417.

FRED. (2020, April 1st) TED rate. Collected from: https://fred.stlouisfed.org/series/TEDRATE

Hamilton, J. D. (1983). Oil and the macroeconomy since Wold War II. *Journal of Political Economy*. 91, 220-248.

Hamilton, J. D. (1989). A new approach to the economic analysis of non-stationary time series and business cycles. *Econometrics*. 57(2). 357-384.

Harding, D. & Pagan, A. (2002). Dissecting the cycle: a methodological investigation. *Journal of Monetary Economics, 49,* 365-381.

Hull, J. C. (2018). *Risk Management and Financial Institutions*. New Jersey: Wiley & Sons.

Ibbotson, R. G. (1975). Price performance of common stock new issues. *Journal of Financial Economics*. 2, 235-272.

Ibbotson, R. G. & Jaffe, J. F. (1975). "Hot Issue" markets. *The Journal of Finance*. 30(4), 1027-1042.

Ibbotson, R. G., Sindelar, J. L. & Ritter, J. R. (1994). The market's problem with the pricing of initial public offerings. *Journal of Applied Corporate Finance*. 7(1), 66-74.

Jin, Y., Gou, J., Zhou, X. & Li, S. P. (2016). IPO market cycles and expansion curse: Evidence from Chinese IPO market. *Investment Analyst Journal*. 45(1), 46-62.

Jones, C. M. & Kaul, G. (1996). Oil and the stock market. *Journal of Finance*. 51, 463-491.

Kim, W. & Weisbach, M. S. (2008). Motivations for public equity offers: An international perspective. *Journal of Financial Economics*. 87(2), 281-307.

Kritzman, M & Li, Y. (2010). Skulls, financial turbulence, and risk management. *Financial Analyst Journal*. 66(5), 30-41.

Krolzig, H. M. (1997*). Markov Switching Vector Autoregressions: Modelling, Statistical Inference and Application to Business Cycle Analysis*. Berlin: Springer.

Logue, D. (1973). On the pricing of unseasoned equity issues: 1965-1969. *Journal of Financial ad Quantitive Analysis*. 8(1), 91-103.

Lowry, M. (2001). Why does IPO volume fluctuate so much? *Journal of Financial Economics*. 67, 3-40.

Macrotrends. (2019, October 8th). S&P 500 10 Year Daily Chart. Collected from: https://www.macrotrends.net/2488/sp500-10-year-daily-chart

Markowitz, H. (1952). Portfolio Selection. *The Journal of Finance*. 7(1), 77-91.

Ofek, E. & Richardson, M. (2004). Illiquidity spillovers: Theory and evidence from European telecom bond issuance. Job-Market Paper.

Mishkin, F. S. (2019). *The Economics of Money, Banking, and Financial Markets.* Harlow: Pearson Education Limited.

Paye, B. S. (2012). 'Deja vol': Predictive regressions for aggregate stock market volatility using macroeconomic variables. *Journal of Financial Economics*. 106, 527-546.

Ritter, J. R. & Welch, I. (2002). A review of IPO activity, pricing and allocations. *The Journal of Finance*. 57(4), 1795-1828.

Rock, K. (1986). Why new issues are underpriced. *Journal of Financial Economics.* 15, 187-212.

Schwert, G. W. (1989a). Why does stock market volatility change over time? *Journal of Finance.* 44, 1115-1153.

Schwert, G. W. (1989b). Business cycles, financial crises and stock volatility. *Carnegie-Rochester Conference Series on Public Policy*. 31, 83-125.

Statman, M (1987). How many stocks make a diversified portfolio? *The Journal of Financial and Quantitative Analysis*. 22(3), 353-363.

Stock, J. H. & Watson, M. W. (2015). *Introduction to Econometrics*. Harlow: Pearson Education Limited.

Vanguard(2020, April 1st). The Vanguard Total Bond Market Index Fund Investor Shares. Collected from:
https://institutional.vanguard.com/web/cfv/product-details/fund/0084.

Wagner, W. H. & Lau, S. C. (1971). The effect of diversification on risk. *Financial Analyst Journal*. 27(6), 48-53.

Wang, Y., Wei, Y., Wu, C. & Yin, L (2018). Oil and the short-term predictability of stock return volatility. *Journal of Empirical Finance*. 47, 90-104.

Welch, I. (1989). Seasoned offerings, imitation costs, and the underpricing of initial public offerings. *Journal of Finance*. 44, 421-449.

Appendix A. Data preparation.

```r
library(quantmod)
library(tidyverse)
library(tseries)
library(forecast)
library(zoo)

setwd("C:/Users/Kristoffer/Documents/Master/R scripts")

# Create new environment
sp500 <- new.env()

# Download data for S&P500
getSymbols("^GSPC", env = sp500, src = "yahoo", from = as.Date("1989-12-29"), to = as.Date("2020-03-31"))

# Create tibble from the S&P500 data
GSPC <- sp500$GSPC
monthly <- to.monthly(GSPC)
tib <- monthly %>% fortify.zoo %>% as_tibble()

# Inspect tibble
class(tib)
head(tib)
tail(tib)
summary(tib)

# Will work with the S&P closing data from 01/01/1960 to 31/03/2020

# Plot the time series and test for stationarity
ggplot(tib, aes(x = Index, y = GSPC.Close)) +
  geom_line()
adf.test(tib$GSPC.Close)
# Both the plot and the ADF test show that the time series is not stationary


# Take the returns and log returns of the time series
log_close = log(tib$GSPC.Close)
log_returns = log_close - lag(log_close, k = 1)
log <- na.omit(log_returns)
returns <- (tib$GSPC.Close - lag(tib$GSPC.Close, k = 1))/lag(tib$GSPC.Close, k=1)
returns

# New tibble with returns and log returns added
tib_returns <- add_column(tib, Returns = returns) %>% drop_na()
str(tib_returns)
tib_returns2 <- add_column(tib_returns, Log.Returns = log) %>% drop_na()

# Returns are stationary in the adf test
ggplot(tib_returns, aes(x = Index, y = Returns)) +
  geom_line()
adf.test(tib_returns$Returns)


# Log returns are stationary
ggplot(tib_returns2, aes(x = Index, y = Log.Returns)) +
```

```
  geom_line()
adf.test(tib_returns2$Log.Returns)

# Download data for VIX
getSymbols("^VIX", env = sp500, src = "yahoo", from = as.Date("1990-01-01"), to = as.Date("2020-03-31"))
vix <- sp500$VIX
head(vix)
vix_monthly <- to.monthly(vix)
str(vix_monthly$vix.High)


# Plot and check for stationarity in VIX. Stationary at a 90% significance level
ggplot(vix_monthly, aes(x = Index, y = vix.High)) +
  geom_line()
adf.test(vix_monthly$vix.High)

# Adding VIX to tibble
head(tib_returns2)
str(tib_returns2)
vix_tb <- vix_monthly %>% fortify.zoo() %>%  as.tibble()
joined_vix <- left_join(tib_returns2, vix_tb)
head(joined_vix)


# Load data for the TEDspread
ted.rate <- read.csv("TEDRATE(1).csv")
head(ted.rate)
str(ted.rate)
ted.rate$TEDRATE <- ted.rate$TEDRATE/100

# Plot and check for stationarity in the TED spread. Stationary at 95% significance level
ts.plot(ted.rate$TEDRATE)
adf.test(ted.rate$TEDRATE)

# Adding TED spread to tibble
ted_tb <- ted.rate %>% as.tibble()
head(ted_tb)
ted_month <- as.yearmon(ted_tb$DATE)
ted_tb$DATE <- ted_month
tedr <- rename(ted_tb, Index = DATE)
head(tedr)

joined_vix_ted <- left_join(joined_vix, tedr)
head(joined_vix_ted)

# Load data for IPOs
ipos <- read.csv("Eikondata.csv", sep=';', stringsAsFactors = FALSE) %>%  as.tibble()
head(ipos)
ipos$Percent.Change.Offer.Price.to.First.Closing.Price <- gsub(",", ".", ipos$Percent.Change.Offer.Price.to.First.Closing.Price)
ipos$Offer.Price <- gsub(",", ".", ipos$Offer.Price)
head(ipos)

# Remove NAs from returns
ipos$Returns <- sapply(ipos$Percent.Change.Offer.Price.to.First.Closing.Price, as.numeric)/100
```

```
head(ipos)
ipos_no_na <- drop_na(ipos)
head(ipos_no_na)

# Remove IPOs where offer price is less than $5
ipos_no_na$Offer.Price <- sapply(ipos_no_na$Offer.Price, as.numeric)
ipos_no_low_offer <- ipos_no_na[which(ipos_no_na$Offer.Price >= 5),]
head(ipos_no_low_offer)

# Select only date and first day returns
ipo_returns <- ipos_no_low_offer[c("Issue.Date", "Returns")]
head(ipo_returns)
tail(ipo_returns)
ipo_returns$Issue.Date <- as.Date(ipo_returns$Issue.Date, "%d. %m. %Y")
ipo_returns$Issue.Date <- as.yearmon(ipo_returns$Issue.Date)

# Take monthly average and prepare data for adding it to joined tibble
ipo_average <- aggregate(Returns ~ Issue.Date, ipo_returns, mean)
n <- dim(ipo_average)[1]
asdf <- ipo_average %>% as.tibble()
ipo_ready <- rename(asdf, Index = Issue.Date)
ipo_ready <- rename(ipo_ready, FD.Returns = Returns)
ipo_ready

# Joined all data
joined_vix_ted_ipos <- left_join(joined_vix_ted, ipo_ready)
head(joined_vix_ted_ipos)

# Make tibble with only the data that is needed
data_set <- joined_vix_ted_ipos[c("Index", "GSPC.Close", "Returns", "Log.Returns", "vix.High", "TEDRATE", "FD.Returns")]
head(data_set)
data_set <- rename(data_set, SP.Close = GSPC.Close)
data_set <- rename(data_set, Date = Index)
data_set <- rename(data_set, SP.Log.Returns = Log.Returns)
data_set <- rename(data_set, First.Day.Returns = FD.Returns)
data_set <- rename(data_set, SP.Returns = Returns)

# Replace NAs in first day returns with previous value
ipo_dataset <- data_set
ipo_dataset$First.Day.Returns <- na.locf(ipo_dataset$First.Day.Returns)
ipo_dataset$First.Day.Returns

# Check if ipo data are stationary. Stationary at a 95% sign level
adf.test(ipo_dataset$First.Day.Returns)

# Get data for Vanguard Total Bond Market Index Fund and add returns and log returns to dataset
getSymbols("VBMFX", from = as.Date("1989-12-29"), to = as.Date("2020-03-31"))
VBMFX_monthly <- to.monthly(VBMFX)
head(VBMFX_monthly)
VBMFX_tib <- VBMFX_monthly %>% fortify.zoo %>% as_tibble()
VBMFX_returns <- (VBMFX_tib$VBMFX.Close - lag(VBMFX_tib$VBMFX.Close, k = 1))/lag(VBMFX_tib$VBMFX.Close, k=1)
VBMFX_log_close <- log(VBMFX_tib$VBMFX.Close)
```

```
VBMFX_log_returns <- VBMFX_log_close - lag(VBMFX_log_close, k = 1)
VBMFX.Returns <- na.omit(VBMFX_returns)
VBMFX.Log.Returns <- na.omit(VBMFX_log_returns)
VBMFX_returns <- add_column(ipo_dataset, VBMFX.Returns)  %>% drop_na()
vbmfx_data <- add_column(VBMFX_returns, VBMFX.Log.Returns) %>% drop_na()
# ADD 3-month T-bills as the risk free rate
getSymbols("DGS3MO", src = "FRED", from = as.Date("1989-12-29"), to = as.Date("2020-03-31"))
head(DGS3MO)
t.bills <- to.monthly(DGS3MO)
head(t.bills)
ts.plot(t.bills$DGS3MO.Close)
risk_free_percent <- t.bills$DGS3MO.Close
risk_free <- risk_free_percent/100
str(risk_free)
rf <- (1+risk_free[97:459,])^(1/12) -1
tail(rf)
colnames(rf) <- "rf"
head(rf)
ts.plot(rf)
final_data <- add_column(vbmfx_data, rf)
head(final_data)
# Save the dataset to csv file
write_csv(final_data, "ipo_dataset.csv")
# Plot all the time seres data
ts.plot(final_data$SP.Close[1:360], ylab="S&P")
ts.plot(final_data$vix.High[1:360], ylab="VIX")
ts.plot(final_data$TEDRATE[1:360], ylab="TED")
ts.plot(final_data$First.Day.Returns[1:360], ylab="IPO returns")
ts.plot(final_data$rf[1:360], ylab="3 month T-bills")
ts.plot(VBMFX_monthly$VBMFX.Close[2:361], ylab="Bond portfolio")
```

Appendix B.1. Regime switching models.

```
library(tidyverse)
library(MSwM)
library(tseries)
library(stargazer)
library(xts)
library(zoo)
set.seed(1234)
# Set wd
setwd("C:/Users/Kristoffer/Documents/Master/R scripts")
# Load data
ipo_data <- read.csv("ipo_dataset.csv", stringsAsFactors = FALSE) %>% as_tibble()
# The dataset
head(ipo_data)
```

```r
tail(ipo_data)

# Split into training dataset (1990-2019) and testing dataset (2020)
training_data <- ipo_data[1:360,]
tail(training_data)
test_data <- ipo_data[361:363,]
test_data

SP.Returns <- training_data$SP.Returns
SP.Log.Returns <- training_data$SP.Log.Returns
VIX <- training_data$vix.High
TEDRATE <- training_data$TEDRATE
First.Day.Returns <- training_data$First.Day.Returns

# Test for stationarity, S&P price is not stationary, log returns are stationary at a 95% significance level
# All other time series are stationary at a 95% significance level
adf.test(training_data$SP.Close)
adf.test(training_data$SP.Log.Returns)
adf.test(training_data$vix.High)
adf.test(training_data$TEDRATE)
adf.test(training_data$First.Day.Returns)



# Model an AR(1) model of S&P500 log returns
sp.model <- lm(SP.Log.Returns ~ lag(SP.Log.Returns, k = 1))
summary(sp.model)

# Check Akaike information criteria for different lag lengths, 1 lag is optimal lag length
iterations <- 10
variables <- 2
output <- matrix(ncol = variables, nrow = iterations)
aicvec <- c()
for(i in 1:iterations){
  n <-AIC(sp.model, k = i)
  output[i, ] <- c(i, n)
}
output
stargazer(output, type = "html")

# Autoregressive Markov Switching model fitted to S&P500 log returns
sp.model.mswm <- msmFit(sp.model, k = 2, sw = c(TRUE, TRUE, TRUE))

# VIX-TED-S&P500 regime switching model
vix.ted.sp.model <- lm(training_data$SP.Log.Returns ~ lag(SP.Log.Returns, k= 1) +
                         lag(VIX, k = 1) + lag(TEDRATE, k = 1))
vix.ted.sp.model.mswm <- msmFit(vix.ted.sp.model, k = 2, sw = c(TRUE, TRUE, TRUE, TRUE, TRUE))

# IPO-VIX-TED-S&P500 regime switching model
ipo.vix.ted.sp.model <- lm(training_data$SP.Log.Returns ~ lag(SP.Log.Returns, k= 1) +
                             lag(VIX, k = 1) + lag(TEDRATE, k = 1) +
                             lag(First.Day.Returns, k = 1))
```

```
ipo.vix.ted.sp.model.mswm <- msmFit(ipo.vix.ted.sp.model, k = 2, sw = c(TRUE, TRUE, TRUE, TRUE, TRUE, TRUE))

# Summary for each of the Markov switching models
summary(sp.model.mswm)
summary(vix.ted.sp.model.mswm)
summary(ipo.vix.ted.sp.model.mswm)

#=================================================================

# Plotting
par(mar=c(1,1,1,1))
tib <- as_tibble(training_data)


# S&P 500 model
# Regime 1: 25:78, 162:210, 269:304, 315:342, 355:359
# Regime 2: 1:24, 79:161, 211:268, 305:314, 343:354, 360:
sp.xone <- c(as.yearmon(training_data$Date[1]),
as.yearmon(training_data$Date[79]),
as.yearmon(training_data$Date[211]),
as.yearmon(training_data$Date[305]),
as.yearmon(training_data$Date[343]),
as.yearmon(training_data$Date[360]))
sp.xtwo <- c(as.yearmon(training_data$Date[25]),
as.yearmon(training_data$Date[162]),
as.yearmon(training_data$Date[269]),
as.yearmon(training_data$Date[315]),
as.yearmon(training_data$Date[355]),
as.yearmon(training_data$Date[360]))
sp.shade <- data.frame(x1 = sp.xone, x2 = sp.xtwo, y1=c(-0.2,-0.2), y2=c(0.2,0.2))
sp.shader <- sp.shade[1:5,]

ggplot() +
  ggtitle("S&P500 MODEL") +
  theme(plot.title = element_text(hjust = 0.5)) +
  labs(x = "Date", y = "S&P500") +
  geom_line(data = tib, aes(x = as.yearmon(Date), y = SP.Close)) +
  geom_rect(data=sp.shader, aes(xmin=x1, xmax=x2, ymin=0, ymax=3500), color="grey", alpha=0.5)

# VIX TED SP model
# Regime 1: 2:8, 20:23, 49:50, 90:92, 102, 114:115, 120:133, 136:140, 144:151, 155:157, 209, 214:217, 221:228, 244, 257:259, 262, 267, 307, 311:312, 336:338, 345:346
# Regime 2: 1, 9:19, 24:48, 51:89, 93:101, 103:113, 116:119, 134:135, 141:143, 152:154, 158:208, 210:213, 218:220, 229:243, 245:256, 260:261, 263:266, 268:306, 308:310, 313:335, 339:344, 347:
vtsp.xone <- c(as.yearmon(training_data$Date[1]),
as.yearmon(training_data$Date[9]),
as.yearmon(training_data$Date[24]),
as.yearmon(training_data$Date[51]),
as.yearmon(training_data$Date[93]),
as.yearmon(training_data$Date[103]),
as.yearmon(training_data$Date[116]),
                as.yearmon(training_data$Date[134]),
as.yearmon(training_data$Date[141]),
as.yearmon(training_data$Date[152]),
as.yearmon(training_data$Date[158]),
```

```
    as.yearmon(training_data$Date[210]),
    as.yearmon(training_data$Date[218]),
                    as.yearmon(training_data$Date[229]),
    as.yearmon(training_data$Date[245]),
    as.yearmon(training_data$Date[260]),
    as.yearmon(training_data$Date[263]),
    as.yearmon(training_data$Date[268]),
    as.yearmon(training_data$Date[308]),
                    as.yearmon(training_data$Date[313]),
    as.yearmon(training_data$Date[339]),
    as.yearmon(training_data$Date[347]),
    as.yearmon(training_data$Date[360]),
    as.yearmon(training_data$Date[360]))

vtsp.xtwo <- c(as.yearmon(training_data$Date[2]),
    as.yearmon(training_data$Date[20]),
    as.yearmon(training_data$Date[49]),
    as.yearmon(training_data$Date[90]),
    as.yearmon(training_data$Date[102]),
    as.yearmon(training_data$Date[114]),
                    as.yearmon(training_data$Date[120]),
    as.yearmon(training_data$Date[136]),
    as.yearmon(training_data$Date[144]),
    as.yearmon(training_data$Date[155]),
    as.yearmon(training_data$Date[209]),
    as.yearmon(training_data$Date[214]),
                    as.yearmon(training_data$Date[221]),
    as.yearmon(training_data$Date[244]),
    as.yearmon(training_data$Date[257]),
    as.yearmon(training_data$Date[262]),
    as.yearmon(training_data$Date[267]),
    as.yearmon(training_data$Date[307]),
                    as.yearmon(training_data$Date[311]),
    as.yearmon(training_data$Date[336]),
    as.yearmon(training_data$Date[345]),
    as.yearmon(training_data$Date[359]),
    as.yearmon(training_data$Date[360]),
    as.yearmon(training_data$Date[360]))

vtsp.shade <- data.frame(x1 = vtsp.xone, x2 = vtsp.xtwo, y1=c(-
    0.2,-0.2), y2=c(0.2,0.2))
vtsp.shader <- vtsp.shade[1:21,]

ggplot() +
  labs(x = "Date", y = "S&P500") +
  theme(plot.title = element_text(hjust = 0.5)) +
  ggtitle("VIX-TEDRATE-S&P500 MODEL") +
  geom_line(data = tib, aes(x = as.yearmon(Date), y = SP.Close)) +
  geom_rect(data=vtsp.shader, aes(xmin=x1, xmax=x2, ymin=0,
    ymax=3500), color="grey", alpha=0.5)


# IPO VIX TED SP model
# Regime 1: 3, 6:8, 20:23, 49:50, 78, 90, 102, 114:115, 120:133,
    137:140, 144:151, 155:157, 170, 209, 214:217, 221:228, 244, 256:259,
    262, 267, 307, 311:312, 336:338, 345:346
# Regime 2: 1:2, 4:5, 9:19, 24:48, 51:77, 79:89, 91:101, 103:113,
    116:119, 134:136, 141:143, 152:154, 158:169, 171:208, 210:213,
    218:220, 229:243, 245:255, 260:261, 263:266, 268:306, 308:310,
    313:335, 339:344, 347:
```

```
ivtsp.xone <- c(as.yearmon(training_data$Date[4]),
  as.yearmon(training_data$Date[9]),
  as.yearmon(training_data$Date[24]),
  as.yearmon(training_data$Date[51]),
  as.yearmon(training_data$Date[79]),
                     as.yearmon(training_data$Date[91]),
  as.yearmon(training_data$Date[103]),
  as.yearmon(training_data$Date[116]),
  as.yearmon(training_data$Date[134]),
  as.yearmon(training_data$Date[141]),as.yearmon(training_data$Date[152
  ]),
                     as.yearmon(training_data$Date[158]),
  as.yearmon(training_data$Date[171]),
  as.yearmon(training_data$Date[210]),
  as.yearmon(training_data$Date[218]),
  as.yearmon(training_data$Date[229]),
  as.yearmon(training_data$Date[245]),
                     as.yearmon(training_data$Date[260]),
  as.yearmon(training_data$Date[263]),
  as.yearmon(training_data$Date[268]),
  as.yearmon(training_data$Date[308]),
  as.yearmon(training_data$Date[313]),
  as.yearmon(training_data$Date[339]),
                     as.yearmon(training_data$Date[347]),
  as.yearmon(training_data$Date[359]),
  as.yearmon(training_data$Date[360]))

ivtsp.xtwo <- c(as.yearmon(training_data$Date[3]),
  as.yearmon(training_data$Date[6]),
  as.yearmon(training_data$Date[20]),
  as.yearmon(training_data$Date[49]),
  as.yearmon(training_data$Date[78]),
  as.yearmon(training_data$Date[90]),
                      as.yearmon(training_data$Date[102]),
  as.yearmon(training_data$Date[114]),
  as.yearmon(training_data$Date[120]),
  as.yearmon(training_data$Date[137]),
  as.yearmon(training_data$Date[144]),
  as.yearmon(training_data$Date[155]),
                      as.yearmon(training_data$Date[170]),
  as.yearmon(training_data$Date[209]),
  as.yearmon(training_data$Date[214]),
  as.yearmon(training_data$Date[221]),
  as.yearmon(training_data$Date[244]),
  as.yearmon(training_data$Date[256]),
                      as.yearmon(training_data$Date[262]),
  as.yearmon(training_data$Date[267]),
  as.yearmon(training_data$Date[307]),
  as.yearmon(training_data$Date[311]),
  as.yearmon(training_data$Date[336]),
  as.yearmon(training_data$Date[345]),
                      as.yearmon(training_data$Date[359]),
  as.yearmon(training_data$Date[360]))

ivtsp.shade <- data.frame(x1 = ivtsp.xone, x2 = ivtsp.xtwo, y1=c(-
  0.2,-0.2), y2=c(0.2,0.2))
ivtsp.shader <- ivtsp.shade[1:24,]

ggplot() +
  ggtitle("IPO-VIX-TEDRATE-S&P500 MODEL") +
  theme(plot.title = element_text(hjust = 0.5)) +
```

```r
  labs(x = "Date", y = "S&P500") +
  geom_line(data = tib, aes(x = as.yearmon(Date), y = SP.Close)) +
  geom_rect(data=ivtsp.shader, aes(xmin=x1, xmax=x2, ymin=0, ymax=3500), color="grey", alpha=0.5)

#===================================================================

# Pseudo out of sampling forecast using transition matrix

# S&P500 model
sp.matrix <- as.matrix(sp.model.mswm@transMat)
sp.matrix

# In December 2019 the model is in regime 2
sp.initial.state = c(0,1)

sp.jan <- sp.matrix %*% sp.initial.state
sp.jan
sp.feb <- sp.matrix %*% sp.jan
sp.feb
sp.march <- sp.matrix %*% sp.feb
sp.march

# Predict that the model will be in regime 2 in all of 2020:Q1

# RMSFE
sp.forecast.log.returns.jan <- 0.01286700 - (0.19318686 * training_data$SP.Log.Returns[360])
sp.forecast.log.returns.feb <- 0.01286700 - (0.19318686 * test_data$SP.Log.Returns[1])
sp.forecast.log.returns.mar <- 0.01286700 - (0.19318686 * test_data$SP.Log.Returns[2])
sp.forecasterror.jan <- test_data$SP.Log.Returns[1] - sp.forecast.log.returns.jan
sp.forecasterror.feb <- test_data$SP.Log.Returns[2] - sp.forecast.log.returns.feb
sp.forecasterror.mar <- test_data$SP.Log.Returns[3] - sp.forecast.log.returns.mar
sp.RMSFE <- sqrt(sp.forecasterror.jan^2 + sp.forecasterror.feb^2 + sp.forecasterror.mar^2)

# Forecast VIX-TED-S&P500 model using transition matrix
vt.matrix <- as.matrix(vix.ted.sp.model.mswm@transMat)
vt.matrix

# December 2019, model is in regime 2
vt.initial.state = c(0,1)

vt.jan <- vt.matrix %*% vt.initial.state
vt.jan
vt.feb <- vt.matrix %*% vt.jan
vt.feb
vt.mar <- vt.matrix %*% vt.feb
vt.mar

# Predict that the model will be in regime 2 in all of 2020:Q1

# RMSFE
vt.forecast.log.returns.jan <- 0.009936377 - (0.07794005 * training_data$SP.Log.Returns[360]) - (0.000737021 *
```

```
   training_data$vix.High[360]) - (0.05794994 *
   training_data$TEDRATE[360])
vt.forecast.log.returns.feb <- 0.009936377 - (0.07794005 *
   test_data$SP.Log.Returns[1]) - (0.000737021 * test_data$vix.High[1])
   - (0.05794994 * test_data$TEDRATE[1])
vt.forecast.log.returns.mar <- 0.009936377 - (0.07794005 *
   test_data$SP.Log.Returns[2]) - (0.000737021 * test_data$vix.High[2])
   - (0.05794994 * test_data$TEDRATE[2])
vt.forecasterror.jan <- test_data$SP.Log.Returns[1] -
   vt.forecast.log.returns.jan
vt.forecasterror.feb <- test_data$SP.Log.Returns[2] -
   vt.forecast.log.returns.feb
vt.forecasterror.mar <- test_data$SP.Log.Returns[3] -
   vt.forecast.log.returns.mar
vt.RMSFE <- sqrt(vt.forecasterror.jan^2 + vt.forecasterror.feb^2 +
   vt.forecasterror.mar^2)

# Forecast IPO-VIX-TED-S&P500 model using transition matrix
ip.matrix <- as.matrix(ipo.vix.ted.sp.model.mswm@transMat)
ip.matrix

# December 2019, model is in regime 1
ip.initial.state = c(1,0)

ip.jan <- ip.matrix %*% ip.initial.state
ip.jan
ip.feb <- ip.matrix %*% ip.jan
ip.feb
ip.march <- ip.matrix %*% ip.feb
ip.march

# Predict that the model will be in regime 1 in all of 2020:Q1

# RMSFE
ip.forecast.log.returns.jan <- -0.015263413 - (0.1782138 *
   training_data$SP.Log.Returns[360]) + (0.0016995085*
   training_data$vix.High[360]) + (0.007586086 *
   training_data$TEDRATE[360]) + (0.01454385 *
   training_data$First.Day.Returns[360])
ip.forecast.log.returns.feb <- -0.015263413 - (0.1782138 *
   test_data$SP.Log.Returns[1]) + (0.0016995085 * test_data$vix.High[1])
   + (0.007586086 * test_data$TEDRATE[1]) + (0.01454385 *
   test_data$First.Day.Returns[1])
ip.forecast.log.returns.mar <- -0.015263413 - (0.1782138 *
   test_data$SP.Log.Returns[2]) + (0.0016995085 * test_data$vix.High[2])
   + (0.007586086 * test_data$TEDRATE[2]) + (0.01454385 *
   test_data$First.Day.Returns[2])
ip.forecasterror.jan <- test_data$SP.Log.Returns[1] -
   ip.forecast.log.returns.jan
ip.forecasterror.feb <- test_data$SP.Log.Returns[2] -
   ip.forecast.log.returns.feb
ip.forecasterror.mar <- test_data$SP.Log.Returns[3] -
   ip.forecast.log.returns.mar
ip.RMSFE <- sqrt(ip.forecasterror.jan^2 + ip.forecasterror.feb^2 +
   ip.forecasterror.mar^2)

# Output table
header <- c("", "S&P500 model", "VIX-TED-S&P500 model", "IPO-VIX-
   TED-S&P500 model")
RMSFE <- c("RMSFE", round(sp.RMSFE, 4), round(vt.RMSFE, 4),
   round(ip.RMSFE, 4))
```

```
tabl <- rbind(header, RMSFE)
stargazer(tabl, type="html")

headr <- c("", "Jan 2020", "Feb 2020", "Mar 2020")
row1 <- c("S&P500 model", round(sp.forecast.log.returns.jan, 4), round(sp.forecast.log.returns.feb, 4), round(sp.forecast.log.returns.mar, 4))
row2 <- c("VIX-TED-S&P500 model", round(vt.forecast.log.returns.jan, 4), round(vt.forecast.log.returns.feb, 4), round(vt.forecast.log.returns.mar, 4))
row3 <- c("IPO-VIX-TED-S&P500 model", round(ip.forecast.log.returns.jan, 4), round(ip.forecast.log.returns.feb, 4), round(ip.forecast.log.returns.mar, 4))
row4 <- c("Actual returns", round(test_data$SP.Log.Returns[1], 4), round(test_data$SP.Log.Returns[2], 4), round(test_data$SP.Log.Returns[3], 4))
tab <- rbind(headr, row1, row2, row3, row4)
stargazer(tab, type="html")
```

Appendix B.2. Portfolio optimization.

```
library(tseries)
library(stargazer)
library(fPortfolio)
set.seed(1234)

# Set wd
setwd("C:/Users/Kristoffer/Documents/Master/R scripts")

# Load data
ipo_data <- read.csv("ipo_dataset.csv", stringsAsFactors = FALSE)

# The dataset
head(ipo_data)
tail(ipo_data)

# Split into training dataset (1990-2019) and testing dataset (2020)
training_data <- ipo_data[1:360,]
test_data <- ipo_data[361:363,]

# The mean risk free rate for the whole period will be used in all Sharpe calculations
rf <- mean(training_data$rf)

#==================================================================

# S&P 500 model no switching
R <- cbind(training_data$SP.Log.Returns, training_data$VBMFX.Log.Returns)
head(R)
colnames(R) <- c("SP", "VBMFX")
dates <- seq.Date(as.Date("1990-01-01"), as.Date("2019-12-01"), by = "month")
```

```
date <- format(dates, "%Y-%m-%d")
rownames(R) <- date

# Efficient frontier
portfolioreturns <- as.timeSeries(R)
effFrontier <- portfolioFrontier(portfolioreturns, constraints = "LongOnly")
plot(effFrontier,c(1,3))
frontierWeights <- getWeights(effFrontier)
risk_return <- frontierPoints(effFrontier)

# Tangency portfolio / Sharpe
tangencyport <- tangencyPortfolio(portfolioreturns, spec = portfolioSpec(), constraints = "LongOnly")
tangencyport
tangency_weights <- getWeights(tangencyport)
tangency_ret <- getTargetReturn(tangencyport)
tangency_sd <- getTargetRisk(tangencyport)
barplot(tangency_weights, main="No Switching Portfolio Weights", xlab = "Asset", ylab = "Weight in portfolio", col=cm.colors(ncol(frontierWeights)+2))
ns.shar <- ((tangency_ret[1]*12) - rf*12)/(tangency_sd[1]*sqrt(12))
ns.shar
ns.weights <- round(tangency_weights, 4)

# Holding period log returns
ns.returns <- R * tangency_weights
ns.holding.peroid.log.returns <- sum(ns.returns)
ns.holding.peroid.log.returns

# Weights table
header1 <- c("Sharpe","")
shar <- c(round(ns.shar, 4), "")
header2 <- c("Weights","")
headers <- c( "S&P 500", " Vanguard")
weight <- c(ns.weights[1], ns.weights[2])
tabl <- rbind(header1, shar, header2, headers, weight)
stargazer(tabl, type = "html")


#===================================================================

# S&P500 model
sp.r1.sp.r <- c(training_data$SP.Log.Returns[25:75], training_data$SP.Log.Returns[162:210],
                training_data$SP.Log.Returns[269:304], training_data$SP.Log.Returns[315:342],
                training_data$SP.Log.Returns[355:359])

sp.r2.sp.r <- c(training_data$SP.Log.Returns[1:24], training_data$SP.Log.Returns[77:161],
                training_data$SP.Log.Returns[211:268], training_data$SP.Log.Returns[305:314],
training_data$SP.Log.Returns[343:354],training_data$SP.Log.Returns[360])


sp.r1.vb.r <- c(training_data$VBMFX.Log.Returns[25:75], training_data$VBMFX.Log.Returns[162:210],
```

```
                    training_data$VBMFX.Log.Returns[269:304], training_data$VBMFX.Log.Returns[315:342],
                    training_data$VBMFX.Log.Returns[355:359])

sp.r2.vb.r <- c(training_data$VBMFX.Log.Returns[1:24], training_data$VBMFX.Log.Returns[77:161],
                    training_data$VBMFX.Log.Returns[211:268], training_data$VBMFX.Log.Returns[305:314],
                    training_data$VBMFX.Log.Returns[343:354], training_data$SP.Returns[360])

sp.r1.date <- c(date[25:75], date[162:210],
                    date[269:304], date[315:342],
                    date[355:359])

sp.r2.date <- c(date[1:24], date[77:161],
                    date[211:268], date[305:314],
                    date[343:354], date[360])

R1 <- cbind(sp.r1.sp.r, sp.r1.vb.r)
head(R1)
colnames(R1) <- c("SP", "VBMFX")
head(sp.r1.date)
rownames(R1) <- sp.r1.date

# Efficient frontier
portfolioreturns <- as.timeSeries(R1)
effFrontier <- portfolioFrontier(portfolioreturns, constraints = "LongOnly")
plot(effFrontier,c(1,3))
frontierWeights <- getWeights(effFrontier)
risk_return <- frontierPoints(effFrontier)

# Tangency portfolio / Sharpe
tangencyport <- tangencyPortfolio(portfolioreturns, spec = portfolioSpec(), constraints = "LongOnly")
tangencyport
tangency_weights <- getWeights(tangencyport)
tangency_ret <- getTargetReturn(tangencyport)
tangency_sd <- getTargetRisk(tangencyport)
barplot(tangency_weights, main="SP Regime 1 Portfolio Weights", xlab = "Asset", ylab = "Weight in portfolio", col=cm.colors(ncol(frontierWeights)+2))
sp.r1.shar <- ((tangency_ret[1]*12) - rf*12)/(tangency_sd[1]*sqrt(12))
sp.r1.shar
sp.r1.weights <- round(tangency_weights, 4)

# Holding period log returns
sp.r1.returns <- R1 * tangency_weights
sp.r1.holding.peroid.returns <- sum(sp.r1.returns)
sp.r1.holding.peroid.returns

# Weights table
header1 <- c("Sharpe","")
shar <- c(round(sp.r1.shar, 4), "")
header2 <- c("Weights","")
headers <- c( "S&P 500", " Vanguard")
weights <-c(sp.r1.weights[1], sp.r1.weights[2])
tabl <- rbind(header1, shar, header2, headers, weights)
stargazer(tabl, type = "html")
```

```
# Regime 2
R2 <- cbind(sp.r2.sp.r, sp.r2.vb.r)
head(R2)
colnames(R2) <- c("SP", "VBMFX")
rownames(R2) <- sp.r2.date

# Efficient frontier
portfolioreturns <- as.timeSeries(R2)
effFrontier <- portfolioFrontier(portfolioreturns, constraints = "LongOnly")
plot(effFrontier,c(1,3))
frontierWeights <- getWeights(effFrontier)
risk_return <- frontierPoints(effFrontier)

# Tangency portfolio / Sharpe
tangencyport <- tangencyPortfolio(portfolioreturns, spec = portfolioSpec(), constraints = "LongOnly")
tangencyport
tangency_weights <- getWeights(tangencyport)
tangency_ret <- getTargetReturn(tangencyport)
tangency_sd <- getTargetRisk(tangencyport)
barplot(tangency_weights, main="SP Regime 2 Portfolio Weights", xlab = "Asset", ylab = "Weight in portfolio", col=cm.colors(ncol(frontierWeights)+2))
sp.r2.shar <- ((tangency_ret[1]*12) - rf*12)/(tangency_sd[1]*sqrt(12))
sp.r2.shar
sp.r2.weights <- round(tangency_weights, 4)

# Holding period log returns
sp.r2.returns <- R2 * tangency_weights
sp.r2.holding.peroid.returns <- sum(sp.r2.returns)
sp.r2.holding.peroid.returns
sp.holding.peroid.log.returns <- sp.r1.holding.peroid.returns + sp.r2.holding.peroid.returns

# Weights table
header1 <- c("Sharpe","")
shar <- c(round(sp.r2.shar, 4), "")
header2 <- c("Weights","")
headers <- c( "S&P 500", " Vanguard")
weights <-c(sp.r2.weights[1], sp.r2.weights[2])
tabl <- rbind(header1, shar, header2, headers, weights)
stargazer(tabl, type = "html")

#=================================================================

# VIX-TED-S&P500 model
vt.r1.sp.r <- c(training_data$SP.Log.Returns[1], training_data$SP.Log.Returns[9:19], training_data$SP.Log.Returns[24:48], training_data$SP.Log.Returns[51:89], training_data$SP.Log.Returns[93:101], training_data$SP.Log.Returns[103:113], training_data$SP.Log.Returns[116:119],
                 training_data$SP.Log.Returns[134:135], training_data$SP.Log.Returns[141:143], training_data$SP.Log.Returns[152:154], training_data$SP.Log.Returns[158:208],
```

```
training_data$SP.Log.Returns[210:213],
training_data$SP.Log.Returns[218:220],
training_data$SP.Log.Returns[229:243],
training_data$SP.Log.Returns[245:256],
training_data$SP.Log.Returns[260:261],
training_data$SP.Log.Returns[263:266],
training_data$SP.Log.Returns[268:306],
training_data$SP.Log.Returns[308:310],
training_data$SP.Log.Returns[313:335],
training_data$SP.Log.Returns[339:344],
training_data$SP.Log.Returns[347:360])

vt.r2.sp.r <- c(training_data$SP.Log.Returns[2:8],
training_data$SP.Log.Returns[20:23],
training_data$SP.Log.Returns[49:50],
training_data$SP.Log.Returns[90:92],
training_data$SP.Log.Returns[102],
training_data$SP.Log.Returns[114:115],
training_data$SP.Log.Returns[120:133],
training_data$SP.Log.Returns[136:140],
training_data$SP.Log.Returns[144:151],
training_data$SP.Log.Returns[155:157],
training_data$SP.Log.Returns[209],
training_data$SP.Log.Returns[214:217],
training_data$SP.Log.Returns[221:228],
training_data$SP.Log.Returns[244],
training_data$SP.Log.Returns[257:259],
training_data$SP.Log.Returns[262], training_data$SP.Log.Returns[267],
training_data$SP.Log.Returns[307],
training_data$SP.Log.Returns[311:312],
training_data$SP.Log.Returns[336:338],
training_data$SP.Log.Returns[345:346])

vt.r1.vb.r <- c(training_data$VBMFX.Log.Returns[1],
training_data$VBMFX.Log.Returns[9:19],
training_data$VBMFX.Log.Returns[24:48],
training_data$VBMFX.Log.Returns[51:89],
training_data$VBMFX.Log.Returns[93:101],
training_data$VBMFX.Log.Returns[103:113],
training_data$VBMFX.Log.Returns[116:119],
training_data$VBMFX.Log.Returns[134:135],
training_data$VBMFX.Log.Returns[141:143],
training_data$VBMFX.Log.Returns[152:154],
training_data$VBMFX.Log.Returns[158:208],
training_data$VBMFX.Log.Returns[210:213],
training_data$VBMFX.Log.Returns[218:220],
training_data$VBMFX.Log.Returns[229:243],
training_data$VBMFX.Log.Returns[245:256],
training_data$VBMFX.Log.Returns[260:261],
training_data$VBMFX.Log.Returns[263:266],
training_data$VBMFX.Log.Returns[268:306],
training_data$VBMFX.Log.Returns[308:310],
training_data$VBMFX.Log.Returns[313:335],
training_data$VBMFX.Log.Returns[339:344],
training_data$VBMFX.Log.Returns[347:360])

vt.r2.vb.r <- c(training_data$VBMFX.Log.Returns[2:8],
training_data$VBMFX.Log.Returns[20:23],
training_data$VBMFX.Log.Returns[49:50],
training_data$VBMFX.Log.Returns[90:92],
```

```
training_data$VBMFX.Log.Returns[102],
training_data$VBMFX.Log.Returns[114:115],
                 training_data$VBMFX.Log.Returns[120:133],
training_data$VBMFX.Log.Returns[136:140],
training_data$VBMFX.Log.Returns[144:151],
training_data$VBMFX.Log.Returns[155:157],
training_data$VBMFX.Log.Returns[209],
training_data$VBMFX.Log.Returns[214:217],
                 training_data$VBMFX.Log.Returns[221:228],
training_data$VBMFX.Log.Returns[244],
training_data$VBMFX.Log.Returns[257:259],
training_data$VBMFX.Log.Returns[262],
training_data$VBMFX.Log.Returns[267],
training_data$VBMFX.Log.Returns[307],
                 training_data$VBMFX.Log.Returns[311:312],
training_data$VBMFX.Log.Returns[336:338],
training_data$VBMFX.Log.Returns[345:346])

vt.r1.date <- c(date[1], date[9:19], date[24:48], date[51:89],
date[93:101], date[103:113], date[116:119],
                 date[134:135], date[141:143], date[152:154],
date[158:208], date[210:213], date[218:220],
                 date[229:243], date[245:256], date[260:261],
date[263:266], date[268:306], date[308:310],
                 date[313:335], date[339:344], date[347:360])

vt.r2.date <- c(date[2:8], date[20:23], date[49:50], date[90:92],
date[102], date[114:115],
                 date[120:133], date[136:140], date[144:151],
date[155:157], date[209], date[214:217],
                 date[221:228], date[244], date[257:259], date[262],
date[267], date[307],
                 date[311:312], date[336:338], date[345:346])

# Regime 1
R1 <- cbind(vt.r1.sp.r, vt.r1.vb.r)
head(R1)
colnames(R1) <- c("SP", "VBMFX")
head(sp.r1.date)
rownames(R1) <- vt.r1.date

# Efficient frontier
portfolioreturns <- as.timeSeries(R1)
effFrontier <- portfolioFrontier(portfolioreturns, constraints =
"LongOnly")
plot(effFrontier, c(1,3))
frontierWeights <- getWeights(effFrontier)
risk_return <- frontierPoints(effFrontier)

# Tangency portfolio / Sharpe
tangencyport <- tangencyPortfolio(portfolioreturns, spec =
portfolioSpec(), constraints = "LongOnly")
tangencyport
tangency_weights <- getWeights(tangencyport)
tangency_ret <- getTargetReturn(tangencyport)
tangency_sd <- getTargetRisk(tangencyport)
barplot(tangency_weights, main="VIX-TED-SP Regime 1 Portfolio
Weights", xlab = "Asset", ylab = "Weight in portfolio",
col=cm.colors(ncol(frontierWeights)+2))
vt.r1.shar <- ((tangency_ret[1]*12) -
rf*12)/(tangency_sd[1]*sqrt(12))
```

```
vt.r1.shar
vt.r1.weights <- round(tangency_weights, 4)

# Holding period log returns
vt.r1.returns <- R1 * tangency_weights
vt.r1.holding.peroid.returns <- sum(vt.r1.returns)
vt.r1.holding.peroid.returns

# Weights table
header1 <- c("Sharpe","")
shar <- c(round(vt.r1.shar, 4), "")
header2 <- c("Weights","")
headers <- c( "S&P 500", " Vanguard")
weights <-c(vt.r1.weights[1], vt.r1.weights[2])
tabl <- rbind(header1, shar, header2, headers, weights)
stargazer(tabl, type = "html")

# Regime 2
R2 <- cbind(vt.r2.sp.r, vt.r2.vb.r)
head(R2)
colnames(R2) <- c("SP", "VBMFX")
rownames(R2) <- vt.r2.date

# Efficient frontier
portfolioreturns <- as.timeSeries(R2)
effFrontier <- portfolioFrontier(portfolioreturns, constraints = "LongOnly")
frontierWeights <- getWeights(effFrontier)
risk_return <- frontierPoints(effFrontier)
risk_return_points <- frontierPoints(effFrontier)
plot(risk_return_points, pch = 16, cex=0.9, xlab = "Mean-Var Target Risk", ylab = "Target Return", main = "Efficient Frontier")

# Tangency portfolio / Sharpe
tangencyport <- tangencyPortfolio(portfolioreturns, spec = portfolioSpec(), constraints = "LongOnly")
tangencyport
tangency_weights <- getWeights(tangencyport)
tangency_ret <- getTargetReturn(tangencyport)
tangency_sd <- getTargetRisk(tangencyport)
barplot(tangency_weights, main="VIX-TED-SP Regime 2 Portfolio Weights", xlab = "Asset", ylab = "Weight in portfolio", col=cm.colors(ncol(frontierWeights)+2))
vt.r2.shar <- ((tangency_ret[1]*12) - rf*12)/(tangency_sd[1]*sqrt(12))
vt.r2.shar
vt.r2.weights <- round(tangency_weights, 4)

# Holding period retuns
vt.r2.returns <- R2 * tangency_weights
vt.r2.holding.peroid.returns <- sum(vt.r2.returns)
vt.r2.holding.peroid.returns
vt.holding.peroid.log.returns <- vt.r1.holding.peroid.returns + vt.r2.holding.peroid.returns
vt.holding.peroid.log.returns

# Weights table
header1 <- c("Sharpe","")
shar <- c(round(vt.r2.shar, 4), "")
header2 <- c("Weights","")
headers <- c( "S&P 500", " Vanguard")
```

```
weights <-c(vt.r2.weights[1], vt.r2.weights[2])
tabl <- rbind(header1, shar, header2, headers, weights)
stargazer(tabl, type = "html")

#====================================================================

# IPO-VIX-TED-S&P500 model
ip.r1.sp.r <- c(training_data$SP.Log.Returns[3],
  training_data$SP.Log.Returns[6:8],
  training_data$SP.Log.Returns[20:23],
                    training_data$SP.Log.Returns[49:50],
  training_data$SP.Log.Returns[78], training_data$SP.Log.Returns[90],
                    training_data$SP.Log.Returns[102],
  training_data$SP.Log.Returns[114:115],
  training_data$SP.Log.Returns[120:133],
                    training_data$SP.Log.Returns[137:140],
  training_data$SP.Log.Returns[144:151],
  training_data$SP.Log.Returns[155:157],
                    training_data$SP.Log.Returns[170],
  training_data$SP.Log.Returns[209],
  training_data$SP.Log.Returns[214:217],
                    training_data$SP.Log.Returns[221:228],
  training_data$SP.Log.Returns[244],
  training_data$SP.Log.Returns[256:259],
                    training_data$SP.Log.Returns[262],
  training_data$SP.Log.Returns[267], training_data$SP.Log.Returns[307],
                    training_data$SP.Log.Returns[311:312],
  training_data$SP.Log.Returns[336:338],
  training_data$SP.Log.Returns[345:346])

ip.r2.sp.r <- c(training_data$SP.Log.Returns[1:2],
  training_data$SP.Log.Returns[4:5],
  training_data$SP.Log.Returns[9:19],
                    training_data$SP.Log.Returns[24:48],
  training_data$SP.Log.Returns[51:77],
  training_data$SP.Log.Returns[79:89],
                    training_data$SP.Log.Returns[91:101],
  training_data$SP.Log.Returns[103:113],
  training_data$SP.Log.Returns[116:119],
                    training_data$SP.Log.Returns[134:136],
  training_data$SP.Log.Returns[141:143],
  training_data$SP.Log.Returns[152:154],
                    training_data$SP.Log.Returns[158:169],
  training_data$SP.Log.Returns[171:208],
  training_data$SP.Log.Returns[210:213],
                    training_data$SP.Log.Returns[218:220],
  training_data$SP.Log.Returns[229:243],
  training_data$SP.Log.Returns[245:255],
                    training_data$SP.Log.Returns[260:261],
  training_data$SP.Log.Returns[263:266],
  training_data$SP.Log.Returns[268:306],
                    training_data$SP.Log.Returns[308:310],
  training_data$SP.Log.Returns[313:335],
  training_data$SP.Log.Returns[339:344],
                    training_data$SP.Log.Returns[347:360])

ip.r1.vb.r <- c(training_data$VBMFX.Log.Returns[3],
  training_data$VBMFX.Log.Returns[6:8],
  training_data$VBMFX.Log.Returns[20:23],
```

```
                     training_data$VBMFX.Log.Returns[49:50], training_data$VBMFX.Log.Returns[78], training_data$VBMFX.Log.Returns[90],
                     training_data$VBMFX.Log.Returns[102], training_data$VBMFX.Log.Returns[114:115], training_data$VBMFX.Log.Returns[120:133],
                     training_data$VBMFX.Log.Returns[137:140], training_data$VBMFX.Log.Returns[144:151], training_data$VBMFX.Log.Returns[155:157],
                     training_data$VBMFX.Log.Returns[170], training_data$VBMFX.Log.Returns[209], training_data$VBMFX.Log.Returns[214:217],
                     training_data$VBMFX.Log.Returns[221:228], training_data$VBMFX.Log.Returns[244], training_data$VBMFX.Log.Returns[256:259],
                     training_data$VBMFX.Log.Returns[262], training_data$VBMFX.Log.Returns[267], training_data$VBMFX.Log.Returns[307],
                     training_data$VBMFX.Log.Returns[311:312], training_data$VBMFX.Log.Returns[336:338], training_data$VBMFX.Log.Returns[345:346])

ip.r2.vb.r <- c(training_data$VBMFX.Log.Returns[1:2], training_data$VBMFX.Log.Returns[4:5], training_data$VBMFX.Log.Returns[9:19],
                     training_data$VBMFX.Log.Returns[24:48], training_data$VBMFX.Log.Returns[51:77], training_data$VBMFX.Log.Returns[79:89],
                     training_data$VBMFX.Log.Returns[91:101], training_data$VBMFX.Log.Returns[103:113], training_data$VBMFX.Log.Returns[116:119],
                     training_data$VBMFX.Log.Returns[134:136], training_data$VBMFX.Log.Returns[141:143], training_data$VBMFX.Log.Returns[152:154],
                     training_data$VBMFX.Log.Returns[158:169], training_data$VBMFX.Log.Returns[171:208], training_data$VBMFX.Log.Returns[210:213],
                     training_data$VBMFX.Log.Returns[218:220], training_data$VBMFX.Log.Returns[229:243], training_data$VBMFX.Log.Returns[245:255],
                     training_data$VBMFX.Log.Returns[260:261], training_data$VBMFX.Log.Returns[263:266], training_data$VBMFX.Log.Returns[268:306],
                     training_data$VBMFX.Log.Returns[308:310], training_data$VBMFX.Log.Returns[313:335], training_data$VBMFX.Log.Returns[339:344],
                     training_data$VBMFX.Log.Returns[347:360])

ip.r1.date <- c(date[3], date[6:8], date[20:23],
                     date[49:50], date[78], date[90],
                     date[102], date[114:115], date[120:133],
                     date[137:140], date[144:151], date[155:157],
                     date[170], date[209], date[214:217],
                     date[221:228], date[244], date[256:259],
                     date[262], date[267], date[307],
                     date[311:312], date[336:338], date[345:346])

ip.r2.date <- c(date[1:2], date[4:5], date[9:19],
                     date[24:48], date[51:77], date[79:89],
                     date[91:101], date[103:113], date[116:119],
                     date[134:136], date[141:143], date[152:154],
```

```
                    date[158:169], date[171:208], date[210:213],
                    date[218:220], date[229:243], date[245:255],
                    date[260:261], date[263:266], date[268:306],
                    date[308:310], date[313:335], date[339:344],
                    date[347:360])

# Regime 1
R1 <- cbind(ip.r1.sp.r, ip.r1.vb.r)
head(R1)
colnames(R1) <- c("SP", "VBMFX")
head(ip.r1.date)
rownames(R1) <- ip.r1.date

# Efficient frontier
portfolioreturns <- as.timeSeries(R1)
effFrontier <- portfolioFrontier(portfolioreturns, constraints = "LongOnly")
frontierWeights <- getWeights(effFrontier)
risk_return <- frontierPoints(effFrontier)
risk_return_points <- frontierPoints(effFrontier)
plot(risk_return_points, pch = 16, cex=0.9, xlab = "Mean-Var Target Risk", ylab = "Target Return", main = "Efficient Frontier")

# Tangency portfolio / Sharpe
tangencyport <- tangencyPortfolio(portfolioreturns, spec = portfolioSpec(), constraints = "LongOnly")
tangencyport
tangency_weights <- getWeights(tangencyport)
tangency_ret <- getTargetReturn(tangencyport)
tangency_sd <- getTargetRisk(tangencyport)
barplot(tangency_weights, main="IPO VIX TED S&P Regime 1 Portfolio Weights", xlab = "Asset", ylab = "Weight in portfolio", col=cm.colors(ncol(frontierWeights)+2))
ip.r1.shar <- ((tangency_ret[1]*12) - rf*12)/(tangency_sd[1]*sqrt(12))
ip.r1.shar
ip.r1.weights <- round(tangency_weights, 4)

# Holding period log returns
ip.r1.returns <- R1 * tangency_weights
ip.r1.holding.peroid.returns <- sum(ip.r1.returns)
ip.r1.holding.peroid.returns

# Weights table
header1 <- c("Sharpe","")
shar <- c(round(ip.r1.shar, 4), "")
header2 <- c("Weights","")
headers <- c( "S&P 500", " Vanguard")
weights <-c(ip.r1.weights[1], ip.r1.weights[2])
tabl <- rbind(header1, shar, header2, headers, weights)
stargazer(tabl, type = "html")

# Regime 2
R2 <- cbind(ip.r2.sp.r, ip.r2.vb.r)
head(R2)
colnames(R2) <- c("SP", "VBMFX")
rownames(R2) <- ip.r2.date

# Efficient frontier
portfolioreturns <- as.timeSeries(R2)
```

```r
effFrontier <- portfolioFrontier(portfolioreturns, constraints = "LongOnly")
plot(effFrontier,c(1,3))
frontierWeights <- getWeights(effFrontier)
risk_return <- frontierPoints(effFrontier)

# Tangency portfolio / Sharpe
tangencyport <- tangencyPortfolio(portfolioreturns, spec = portfolioSpec(), constraints = "LongOnly")
tangencyport
tangency_weights <- getWeights(tangencyport)
tangency_ret <- getTargetReturn(tangencyport)
tangency_sd <- getTargetRisk(tangencyport)
barplot(tangency_weights, main="IPO VIX TED SP Regime 2 Portfolio Weights", xlab = "Asset", ylab = "Weight in portfolio", col=cm.colors(ncol(frontierWeights)+2))
ip.r2.shar <- ((tangency_ret[1]*12) - rf*12)/(tangency_sd[1]*sqrt(12))
ip.r2.shar
ip.r2.weights <- round(tangency_weights, 4)

# Holding period log returns
ip.r2.returns <- R2 * tangency_weights
ip.r2.holding.peroid.returns <- sum(ip.r2.returns)
ip.r2.holding.peroid.returns
ip.holding.peroid.log.returns <- ip.r1.holding.peroid.returns + ip.r2.holding.peroid.returns
ip.holding.peroid.log.returns

# Weights table
header1 <- c("Sharpe","")
shar <- c(round(ip.r2.shar, 4), "")
header2 <- c("Weights","")
headers <- c( "S&P 500", " Vanguard")
weights <-c(ip.r2.weights[1], ip.r2.weights[2])
tabl <- rbind(header1, shar, header2, headers, weights)
stargazer(tabl, type = "html")

#=================================================================

# Compare holding period log returns
ns.holding.peroid.log.returns
sp.holding.peroid.log.returns
vt.holding.peroid.log.returns
ip.holding.peroid.log.returns

# Relative holding period returns
ns.holding.period.returns <- exp(ns.holding.peroid.log.returns) - 1
sp.holding.period.returns <- exp(sp.holding.peroid.log.returns) - 1
vt.holding.period.returns <- exp(vt.holding.peroid.log.returns) - 1
ip.holding.period.returns <- exp(ip.holding.peroid.log.returns) - 1

# Output
hed <- c("No switching model", "S&P500 model", "VIX-TED-S&P500 model", "IPO-VIX-TED-S&P500 model")
row1 <- c(ns.holding.peroid.log.returns, sp.holding.peroid.log.returns, vt.holding.peroid.log.returns, ip.holding.peroid.log.returns)
row2 <- c(ns.holding.period.returns, sp.holding.period.returns, vt.holding.period.returns, ip.holding.period.returns)
tabl <- rbind(hed, row1, row2)
```

```
stargazer(tabl, type="html")
```

Appendix B.3. GARCH model.

```
library(tidyverse)
library(rmgarch)
library(zoo)
library(stargazer)

setwd("C:/Users/Kristoffer/Documents/Master/R scripts")

# Load data
data <- read.csv("ipo_dataset.csv", stringsAsFactors = FALSE)

 log.return <- data$SP.Log.Returns
 log.return


 # Use the training data
 sp.returns <- log.return[0:360]

 # Specify that we use an ARIMA(1,0) i.e. AR(1) model
 ug_spec <- ugarchspec(mean.model = list(armaOrder=c(1,0)))

 ug_fit <- ugarchfit(spec = ug_spec, data = sp.returns)
 ug_fit@fit$coef
 ug_var <- ug_fit@fit$var
 ug_res2 <- (ug_fit@fit$residuals)^2

 plot(ug_res2, type ="l", ylab = "S&P 500 volatility")
 lines(ug_var, col="green")

 ug.df <- cbind(data$Date[0:360], ug_res2, ug_var)
 colnames(ug.df) <- c("Date", "Res", "Var")
 tib <- as_tibble(ug.df)
 tib$Res <- as.numeric(tib$Res)
 tib$Var <- as.numeric(tib$Var)

 colors <- c("Squared residuals" = "black", "Conditional variances" = "green")

 ggplot(data = tib) +
   geom_line(aes(x = as.yearmon(Date), y = Res, group=1, color = "Squared residuals")) +
   geom_line(aes(x = as.yearmon(Date), y = Var, group=1, color = "Conditional variances")) +
   labs(x = "", y = "", color = "Legend") +
   theme_bw() +
   ggtitle("GARCH(1,1)") +
   theme(plot.title = element_text(hjust = 0.5)) +
   scale_color_manual(values = color) +
   theme(legend.position = c(0.85,0.9))
```

```r
# Forecasting
ug_fore <- ugarchforecast(ug_fit, n.ahead=3)
ug_fore

ug_f <- ug_fore@forecast$sigmaFor
plot(ug_f, type="l", ylab = "Forecasted variance", xlab = "Month", ylim = c(0, 0.05))

ug_var_t <- c(tail(ug_var, 12), rep(NA,3))
ug_res2_t <- c(tail(ug_res2, 12), rep(NA, 3))
ug_f <- c(rep(NA, 12), (ug_f)^2)

plot(ug_res2_t, type="l", ylab = "Variance", xlab = "Month", main = "GARCH(1,1)")
lines(ug_f, col="orange")
lines(ug_var_t, col="green")
legend(10,0.005,legend = c("Squared residuals", "Conditional variance", "Forecasted variance"), col = c("Black", "Green", "Orange"), lty=1)

# Model with complete dataset
complete.ug_fit <- ugarchfit(spec = ug_spec, data = log.return)
complete.ug_fit@fit$coef
complete.ug_var <- complete.ug_fit@fit$var
complete.ug_res2 <- (complete.ug_fit@fit$residuals)^2

complete.ug.df <- cbind(data$Date, complete.ug_res2, complete.ug_var)
colnames(complete.ug.df) <- c("Date", "Res", "Var")
tib <- as_tibble(complete.ug.df)
tib$Res <- as.numeric(tib$Res)
tib$Var <- as.numeric(tib$Var)

colors <- c("Squared residuals" = "black", "Conditional variances" = "green")
tail(tib)

ggplot(data = tib) +
  geom_line(aes(x = as.yearmon(Date), y = Res, group=1, color = "Squared residuals")) +
  geom_line(aes(x = as.yearmon(Date), y = Var, group=1, color = "Conditional variances")) +
  labs(x = "", y = "", color = "Legend") +
  theme_bw() +
  ggtitle("GARCH(1,1)") +
  theme(plot.title = element_text(hjust = 0.5)) +
  scale_color_manual(values = color) +
  theme(legend.position = c(0.85,0.9))

# Forecast error
jan.2020 <- (ug_fore@forecast$seriesFor[1] - data$SP.Log.Returns[361])^2
feb.2020 <- (ug_fore@forecast$seriesFor[2] - data$SP.Log.Returns[362])^2
mar.2020 <- (ug_fore@forecast$seriesFor[3] - data$SP.Log.Returns[363])^2
RMSFE <- sqrt((jan.2020 + feb.2020 + mar.2020)/3)
header <- c("Jan 2020", "Feb 2020", "March", "RMSFE")
estimates <- c(round(ug_fore@forecast$seriesFor[1], 4), round(ug_fore@forecast$seriesFor[2], 4), round(ug_fore@forecast$seriesFor[3], 4))
```

```
observed <- c(round(data$SP.Log.Returns[361], 4),
  round(data$SP.Log.Returns[362], 4), round(data$SP.Log.Returns[363],
  4))
row <- c(round(jan.2020, 4), round(feb.2020, 4), round(mar.2020,
  4), round(RMSFE, 4))
tabl <- rbind(header, estimates, observed, row)
stargazer(tabl, type="html")
```

Appendix C.1 Results of the S&P500 regime switching model

```
Markov Switching Model

Call: msmFit(object = sp.model, k = 2, sw = c(TRUE, TRUE, TRUE))

        AIC       BIC    logLik
  -1343.032 -1303.965 675.5159

Coefficients:

Regime 1
---------
                              Estimate Std. Error t value Pr(>|t|)
(Intercept)(S)                  0.0023     0.0039  0.5897   0.5554
lag(SP.Log.Returns, k = 1)(S)   0.0596     0.0741  0.8043   0.4212

Residual standard error: 0.05229265
Multiple R-squared: 0.003487

Standardized Residuals:
         Min           Q1          Med           Q3          Max
-0.182216704 -0.006466105  0.001145065  0.011132658  0.104290533

Regime 2
---------
                              Estimate Std. Error t value  Pr(>|t|)
(Intercept)(S)                  0.0129     0.0021  6.1429 8.103e-10 ***
lag(SP.Log.Returns, k = 1)(S)  -0.1932     0.0828 -2.3333   0.01963 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.02181118
Multiple R-squared: 0.04346

Standardized Residuals:
          Min            Q1           Med            Q3           Max
-0.0568646128 -0.0052576090  0.0005339147  0.0072434052  0.0388749132

Transition probabilities:
           Regime 1   Regime 2
Regime 1 0.97025881 0.03696034
Regime 2 0.02974119 0.96303966
```

Appendix C.2 Results of the VIX-TED-S&P500 regime switching model

```
Markov Switching Model

Call: msmFit(object = vix.ted.sp.model, k = 2, sw = c(TRUE, TRUE, TRUE,
    TRUE, TRUE))

        AIC       BIC    logLik
  -1361.246 -1283.113 688.6229

Coefficients:

Regime 1
---------
                              Estimate Std. Error t value Pr(>|t|)
(Intercept)(S)                  0.0099     0.0122  0.8115   0.4171
lag(SP.Log.Returns, k = 1)(S)  -0.0779     0.1138 -0.6845   0.4937
lag(VIX, k = 1)(S)             -0.0007     0.0005 -1.4000   0.1615
lag(TEDRATE, k = 1)(S)         -0.0006     0.0006 -1.0000   0.3173
```

```
Residual standard error: 0.04642525
Multiple R-squared: 0.05466

Standardized Residuals:
         Min           Q1          Med           Q3          Max 
-0.142950955 -0.002683327  0.009145703  0.018694798  0.110592186 

Regime 2
---------
                              Estimate Std. Error t value  Pr(>|t|)    
(Intercept)(S)                 -0.0146     0.0055 -2.6545  0.007943 ** 
lag(SP.Log.Returns, k = 1)(S)  -0.1767     0.0557 -3.1724  0.001512 ** 
lag(VIX, k = 1)(S)              0.0016     0.0003  5.3333 9.644e-08 ***
lag(TEDRATE, k = 1)(S)          0.0003     0.0005  0.6000  0.548506    
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.02122698
Multiple R-squared: 0.362

Standardized Residuals:
         Min           Q1          Med           Q3          Max 
-0.035506057 -0.013625254 -0.002066081  0.009259954  0.039065222 

Transition probabilities:
          Regime 1  Regime 2
Regime 1 0.6992455 0.2108881
Regime 2 0.3007545 0.7891119
```

Appendix C.3 Results of the IPO-VIX-TED-S&P500 regime switching model

```
Markov Switching Model

Call: msmFit(object = ipo.vix.ted.sp.model, k = 2, sw = c(TRUE, TRUE, 
    TRUE, TRUE, TRUE, TRUE))

        AIC       BIC   logLik
  -1360.423 -1262.756 690.2113

Coefficients:

Regime 1
---------
                                 Estimate Std. Error t value Pr(>|t|)
(Intercept)(S)                     0.0067     0.0138  0.4855   0.6273
lag(SP.Log.Returns, k = 1)(S)     -0.0762     0.1222 -0.6236   0.5329
lag(VIX, k = 1)(S)                -0.0008     0.0005 -1.6000   0.1096
lag(TEDRATE, k = 1)(S)            -0.0534     0.0542 -0.9852   0.3245
lag(First.Day.Returns, k = 1)(S)   0.0145     0.0299  0.4849   0.6277

Residual standard error: 0.04502278
Multiple R-squared: 0.06282

Standardized Residuals:
         Min           Q1          Med           Q3          Max 
-0.139855365 -0.002424712  0.009663693  0.019358460  0.111580936 

Regime 2
---------
                                 Estimate Std. Error t value  Pr(>|t|)    
(Intercept)(S)                    -0.0153     0.0054 -2.8333 0.0046070 ** 
lag(SP.Log.Returns, k = 1)(S)     -0.1783     0.0529 -3.3705 0.0007503 ***
lag(VIX, k = 1)(S)                 0.0017     0.0003  5.6667 1.456e-08 ***
lag(TEDRATE, k = 1)(S)             0.0084     0.0498  0.1687 0.8660326    
lag(First.Day.Returns, k = 1)(S)   0.0144     0.0109  1.3211 0.1864680    
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.02032249
Multiple R-squared: 0.4014

Standardized Residuals:
         Min           Q1          Med           Q3          Max 
-0.034094009 -0.012431139 -0.002353037  0.008120321  0.035926625 

Transition probabilities:
         Regime 1  Regime 2
Regime 1 0.672469 0.2471704
Regime 2 0.327531 0.7528296
```