Constructing a Volatility Risk Premium Using Gaussian Process for Regression

B.Sc. Thesis in Finance at Stockholm School of Economics

Fabian Mally* Love Marcus†

May 14, 2017

Abstract

In this thesis we investigate the volatility risk premium (VRP) on OMXS30 and S&P 500 and the predictive capabilities of Gaussian Process for regression (GP) on the volatility of those indices. The results are evaluated by comparison with corresponding predictions of a few methods from the GARCH family as well as a naive approach. Several volatility risk premia are constructed using the different forecasting methods, and their explanatory power for stock market returns is analyzed using linear regressions.

We found that the one day ahead volatility forecasts made with the GP were not as similar to the realized volatility as those made with the naive approach, and not as good at predicting the direction of change as the comparative GARCH methods. There seems to exist a volatility risk premium on the Swedish market, however not as large as the VRP on the US market, potentially indicating greater risk aversion among investors on the US market. For both markets, the volatility risk premium is found to predict future stock market excess returns, with the VRP constructed from the regular GARCH$(1, 1)$ giving the highest adjusted R^2.

Supervisor: Irina Zviadadze, Assistant Professor, Department of Finance, Stockholm School of Economics

Keywords: Volatility, Gaussian Process, GARCH, Volatility Risk Premium, Volatility Prediction, Realized Volatility, Implied Volatility

*23390@student.hhs.se

†23111@student.hhs.se

Contents

1 Introduction

Uncertainty is an integral part of financial market dynamics. For some market actors it is an unwanted source of risk which should optimally be minimized without affecting asset returns, while for others it is central source of producing asset returns. Nonetheless, it is an important consideration for investors in stock markets, as equities tend to be volatile and have given rise to numerical measures such as volatility. One indicator of volatility is the so-called ”fear-index” of implied volatility (IV), derived from prices of derivatives on financial markets. Implied volatility is, however, a biased measure, as it is calculated in a theoretical, risk-neutral, world, and tends therefore not to be the same as the actual volatility, commonly called realized volatility (RV). This discrepancy has given rise to the concept of a volatility risk premium (VRP), defined as the difference between the implied and the realized volatility. The VRP could be regarded as an indicator of the average investor's risk aversion, as it quantifies to some extent, how much investors are willing to pay to hedge against the ”volatility of volatility”. While the implied and realized volatility measures themselves have not been found to predict stock market returns, Bollerslev et. al. (2009) found that the volatility risk premium provides explanatory power for returns on the S&P 500 [BTZ09].

How to calculate the VRP is on the other hand not straightforward, as the two volatility measures should optimally reflect the same period. Previous research has focused on the source of the VRP and its characteristics and put less effort into the calculation of the variable itself, while in fact a substantial part of the work lies in forecasting a realized volatility for the following period. There is a vast amount of different volatility forecasting methods, and as of lately, new methods from the field of machine learning in computer science. In this thesis we investigate the predictive performance of such a method, namely Gaussian process for regression (GP), and compare it to standard methods from the GARCH[1] family of models as well as a naive approach on the OMXS30 and S&P 500 indexes. These volatility forecasts are then used for constructing VRPs, that are examined in terms of their explanatory power of stock market returns.

[1] General autoregressive conditional heteroskedasticity.

1.1 Thesis Motivation and Aim

Given the central role of volatility and an increasing interest in the VRP, an examination of the premium on other markets than the US is of interest. The extent to which investors are willing to pay a premium for insurance against the uncertainty of volatility, might be of interest for researchers in many fields, for instance behavioral finance. The fact that the premium has been found to predict stock market returns is of interest for academia as well as professionals, as any input that could improve a stock market forecasting is of interest for e.g. traders and investors. Finally, evaluating a new and promising machine learning method against different volatility forecasting methods is of interest, since more precise forecasts are constantly sought.

The aim of this thesis is therefore to examine the VRP and its explanatory power for future stock market returns on the Swedish and the American stock markets, and to conduct a comparison of volatility forecasting between the machine learning method of Gaussian processes for regression and other benchmark methods. The choice of the Swedish market could produce some new interesting results, while the American market is examined so that the results can be compared to previous research in the field.

1.2 Research Questions

In line with the thesis aim, two research questions are be formulated:

- What are the characteristics of the volatility risk premium on OMXS30 and does it predict future index returns?
- Is Gaussian process for regression suitable for predicting stock market volatility and the volatility risk premium (VRP)?

1.3 Previous Research

One of the first quantitative definitions of the volatility risk premium was proposed by Carr and Wu [CW09], where they defined a variance risk premium as the difference between a variance swap rate, retrieved from option prices on the market, and the realized variance. They found that for a majority of the assets they investigated, the historical variance swap rate was larger than the ex post (after the event) realized variance, i.e. they focused on an ex post VRP.

Bollerslev et. al. [BTZ09] found that the variance risk premium explains a non-trivial part of future excess returns on the S&P 500; that a high (low) premium predicts higher (lower) future stock returns. Similar results have been found for bonds [M+12] and the FX-market [DRS16]. Bollerslev et. al. received this result for ex post variance risk premia, ex post volatility risk premia and an expected variance risk premium, which essentially is an ex ante risk premium. They stress the importance of an ex ante risk premium, i.e. a VRP constructed using forecasted realized volatility (RV), as it reduces the temporal effects on the results, since the ex post risk premium is constructed using the RV of the past period, and not a forecasted one. Their ex ante variance risk premium is constructed using the forecasts of the realized volatility from the HAR-RV[2] method. This thesis will extend the analysis to the Swedish stock market and specifically extend the ex ante version of the risk premium, comparing predictability using forecasts from different forecasting methods.

Machine Learning (ML) is the field of developing software and algorithms that can learn from data and be able to perform tasks without being explicitly programmed, for example by regressing on functional forms that have not been specified in advance. Most ML-algorithms are not new, but they have gained new interest as data has become more available and the capability to perform demanding mathematical calculations on large data-sets has increased. In this thesis we apply Gaussian process for regression (GP), originally developed for geological applications, on stock market volatility. GP and similar methods have, however, been applied to financial data before.

Wilson and Ghahramani [WG10] developed a Gaussian Copula process Volatility model that performed on par with GARCH$(1,1)$ for predictions on the exchange rate volatility between the Deutschmark and the Great Britain Pound from 1984 to 1992. Titsias et. al. [TL11] applies a heteroscedastic GP on the same data-set with slightly better results, especially on long-term out-of-sample predictions. In [WHG14] Wu et. al. develops an on-line[3] heteroscedastic GP algorithm that on average outperforms GARCH, EGARCH and GJR-GARCH on one day volatility forecasts on twenty different currency exchange rates. In the draft [CD10] Chapados and Dorion apply GP on SPY[4] data over four years from 2002 with comparatively good results for

[2] Heterogenous autoregressive model of realized volatility. HAR-RV is a autoregressive fractionally integrated moving average usually evaluated using daily, weekly and monthly averaged historical RV.

[3] Continuously adding data new points when learning instead of evaluating batches.

[4] An exchange traded fund designed to track the S&P 500.

long term predictions, they do, however, include an index of the one month ahead implied volatility as a parameter.

2 Preliminaries

2.1 Implied and Realized Volatility

For the purpose of this thesis it is important to distinguish between the implied volatility, IV, and the realized volatility, RV — both are measures that try to capture the variation of a given asset. Implied volatility is the standard deviation of the returns expected in the future implied from financial derivatives on the market, which means it is the expected volatility under the risk-neutral measure. A common way to retrieve the implied volatility is to work it out backwards from the Black and Scholes model for different put and call options with a corresponding time to maturity, as the prices of the options and the other inputs into the model are readily available. There are other ways to retrieve the implied volatility as well, for instance using a model-free approach as in [BN00; CW09; JT05]. These methods are based on a weighted summation of all available put and call options.

Realized volatility is an estimation of the true standard deviation of the returns, and can be calculated ex post at time t for the period $[t-\tau, t]$ using

$$RV_{t,\tau} = \sqrt{\frac{1}{\tau}\sum_{i=0}^{\tau} r_{t-i}^2}. \tag{1}$$

where r_i is the deviation of the return from the expected return for day i, defined as $R_i - \mu = ln(P_i/P_{i-1}) - \mu$. μ is the expected value of R_i for all considered i and P_i is the closing price of the asset on day i. If $\mu = 0$ for the modeled asset, then $r_i = R_i$, which simplifies the calculations such that $RV_{t,\tau}$ is equal to the square root of the sum of τ past squared returns divided by the number of observations τ.

In a task of evaluating forecasting methods, one would compare the forecasted values with the realized volatility. The higher the frequency of the returns, the more accurate the estimation of the ex post volatility will be, but this also depends on the desired frequency of the volatility. Alternatively, one could use the daily realized range, defined as the log difference between the daily high and low of the asset, instead of daily returns. It has been shown to be a better proxy for the actual volatility, than using daily returns, with an accuracy equivalent to using intra-day returns between three and six hours [AB98; BD03].

2.2 Volatility Risk Premium

2.2.1 Definition

The volatility risk premium at time t for a future period of τ time periods is defined as the difference between the expected value of the volatility under the risk-neutral measure $\mathbb{Q}$ and the "physical" measure $\mathbb{P}$, i.e. what is usually considered to be the regular expected value [BTZ09],

$$VRP_{t,\tau} = \mathbf{E}_t^{\mathbb{Q}}[\sigma_{t,\tau}] - \mathbf{E}_t[\sigma_{t,\tau}]. \tag{2}$$

In this thesis, only a period of one month in the future will be considered for the VRP, as the data for implied volatility is limited to options of 30 days to maturity, and therefore the variable τ will be suppressed in later equations.

As clarified in section 2.1, we have $\mathbf{E}_t^{\mathbb{Q}}[\sigma_t] = IV_t$, where IV_t is the implied volatility for the subsequent month. The expected value of the volatility under the "physical" measure is more involved, but it can be approximated by $\mathbf{E}_t[RV_\tau]$ as in [BTZ09; DRS16], which can be predicted using many different forecasting methods. In other words, the issue of constructing a volatility risk premium is reduced to one of forecasting the realized volatility as accurately as possible.

2.2.2 Economic interpretation

The existence of a variance, and volatility, risk premium is based upon empirical findings, but has been theoretically argued for using the concept of variance or volatility swap contracts [CW09]. A long position in such a contract constitutes of paying a fixed variance swap rate, while receiving the actual, floating, variance rate. The swap rate is calculated as the implied variance from derivatives on the market using the risk-neutral measure and the actual variance rate can be approximated ex ante using different variance forecasting methods. The variance risk premium can then be defined as the difference between the swap rate and the actual rate, which is the equivalent of the profit or loss of taking a short position in the variance swap contract. The volatility risk premium can be calculated in a similar way.

Bollerslev et. al. [BTZ09] argues for the VRP's explanatory power of stock market returns using a stylized self-contained equilibrium model of stock market returns and volatility. The equity risk premium, an indicator of how much the stock market return must outperform a risk-free asset to justify its price, that follows from their model consists of a variance factor, which represents the classic risk-return relationship, and a factor for the volatility-of-volatility, which they argue is the true premium of volatility risk. The authors then continue by showing that the volatility-of-volatility can be measured as the difference between the risk-neutral and the objective, physical, expectation of the volatility of the next period, i.e. the volatility risk premium. By doing so, they show that the VRP serves as an explanatory variable of stock market excess returns, as it constitutes the factor for volatility risk in the equity premium of their model.

If the existence of a volatility risk premium was merely due to market inefficiencies, it should have been a temporary phenomenon according to the efficient market hypothesis, but as it has been known for over 10 years, its existence should have another explanation [JT05]. The most prominent economic interpretation and explanation of the VRP is that it serves as a proxy for the average investor's risk aversion [BH14; BGZ11]. This is because it quantifies how much volatility the average investor is willing to pay for, in excess of the predicted future volatility. A VRP of 5 for instance, means that the spread between implied and realized volatility amount to 5 percentage points, and indicates that investors are willing to pay a 25% premium if the annualized, realized, volatility is 20%. If the VRP quantifies investors' risk aversion to some extent, any potential explanatory power of stock market excess returns might be expected, as these two concepts are closely interconnected through the classic risk-return relationship of the equity risk premium.

2.3 Volatility Forecasting Methods

2.3.1 Random walk

The most simple method for estimating future realized volatility is by assuming that the volatility follows a zero drift Wiener process, or a random walk, such that $\mathbf{E}_t[\sigma_{t+1}] = \sigma_t$. In this case, the estimated future volatility is equal to the value of the past RV, calculated as in equation (1) [PG03]. This is a model-free method of estimating the future volatility, and has been used extensively by previous researchers when calculating an ex ante VRP [BGZ11; DRS16; PS14].

2.3.2 GARCH models

The return volatility is instantaneous and has to be approximated from previous returns. One way of doing this is to calculate the average of the residuals from the last fixed number of days. This method will, exactly as the naive, assume the same volatility over the next period as the last, it is, however, likely that the previous residuals have different explanatory power depending on temporal distance. Engle proposed a method for weighing previous residuals, the ARCH(q) model[5] [Eng82]

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^{q} \alpha_i \epsilon_{t-i}^2 \tag{3}$$

where α_0 is the mean, ϵ_t are the residuals at time t and α_i are the weights of the corresponding residual. The weights are determined by the best fit on the historic data.

A common generalization of this model is the GARCH(p,q) model, proposed by Bollerslev [Bol86]. Instead of basing the estimate on only the q nearest residuals, the GARCH(p,q) model recursively includes all previous residuals through the dependence on the variance of p previous periods

$$\sigma_t^2 = \omega + \sum_{i=1}^{q} \alpha_i \epsilon_{t-i}^2 + \sum_{i=1}^{p} \beta_i \sigma_{t-i}^2 \tag{4}$$

where ω is the mean and α_i, and β_i are fitted weights. The most popular model in this family of methods is the GARCH(1,1) method, perhaps due to its performance and simplicity, as Hansen and Lunde for instance found that it is not beaten by more sophisticated models in forecasting exchange rate volatility [HL05].

One disadvantage of the most basic forecasting models, including the original GARCH, is that it does not consider the asymmetric properties of volatility, i.e. that volatility tends to be higher in downswings than in upswings. This property is accounted for in the GJR-GARCH [GJR93] that besides a few technical differences from the GARCH$(1,1)$ model also includes a dummy for the sign of the last residual. The final model in the GARCH family that is introduced in this brief review is the Realized GARCH [HHS12]. Realized GARCH includes realized volatility calculated with higher frequency data than previous GARCH derivatives utilize.

[5] Autoregressive conditional heteroskedasticity model.

2.3.3 Gaussian process for regression

Gaussian process for regression (GP) is a fully probabilistic general purpose regression method based on Bayesian inference, where a prior probability distribution is updated with new information[6] when available. Before any information is taken into account by the GP, the prior encodes the current belief of the distribution. Here, a concise introduction of GP is provided, see for example [Bis06; RW05] for a more detailed reviews.

Given a set of N non-random real valued observations t_n of some underlying values y_n, and corresponding input variables $\boldsymbol{x}_n \in \boldsymbol{X}$, subject to a mean zero additive Gaussian noise with constant variance $\epsilon_n \sim \mathcal{N}(0, \sigma^2)$ such that

$$t_n = y_n + \epsilon_n, \tag{5}$$

the probability of all observed data can be expressed

$$p(\boldsymbol{t}|\boldsymbol{y}) = \mathcal{N}(\boldsymbol{t}|\boldsymbol{0}, \sigma \boldsymbol{I}_N) \tag{6}$$

with $\boldsymbol{t} = (t_1, ..., t_N)^T$ and $\boldsymbol{y} = (y_1, ..., y_N)^T$. σ is a hyperparameter that affect the shape of the approximation without requiring it to have any specific functional form, see section 2.3.4. Our goal is to assign the most probable value to previously not seen input vectors. This can be achieved by defining a marginal probability distribution $p(\boldsymbol{t})$

$$p(\boldsymbol{t}) = \int p(\boldsymbol{t}|\boldsymbol{y}) p(\boldsymbol{y}) d\boldsymbol{y} \tag{7}$$

where the condition on $\boldsymbol{X}$ has been omitted for clarity of notation. This integral can be interpreted as removing the dependence on the specific form of the underlying function by integrating over all possible functions weighted by their prior probability encoded in $p(\boldsymbol{y})$ and the the evidence $p(\boldsymbol{t}|\boldsymbol{y})$. The evidence was defined in equation (6) and the prior can be expressed as

$$p(\boldsymbol{y}) = \mathcal{N}(\boldsymbol{y}|\boldsymbol{0}, \boldsymbol{K}). \tag{8}$$

with the covariance matrix $\boldsymbol{K}$ determined by a kernel function $\boldsymbol{K}_{m,n} = k(\boldsymbol{x}_m, \boldsymbol{x}_n)$, see section 2.3.4. Mean zero is not a necessity but it simplifies notation without sacrificing generality. Given equation (8), the marginal probability can be evaluated to

$$p(\boldsymbol{t}) = \mathcal{N}(\boldsymbol{t}|\boldsymbol{0}, \boldsymbol{C}) \tag{9}$$

[6] Usually called evidence.

with the elements of $\boldsymbol{C}$ given by

$$\boldsymbol{C}_{m,n} = k(\boldsymbol{x_m}, \boldsymbol{x_n}) + \sigma\delta_{m,n} \tag{10}$$

where $\delta_{m,n}$ is the Kronecker delta[7].

The estimated value for a new input vector $\boldsymbol{x}_{N+1}$ is implied by the by the joint probability distribution over $t_1, ..., t_{N+1}$

$$p(t_{N+1}) = \mathcal{N}\left(\begin{bmatrix} \boldsymbol{t} \\ t_{N+1} \end{bmatrix} \middle| \begin{bmatrix} \boldsymbol{0} \\ 0 \end{bmatrix}, \begin{bmatrix} \boldsymbol{C}_N & \boldsymbol{k}(\boldsymbol{X}, \boldsymbol{x}_{N+1}) \\ \boldsymbol{k}(\boldsymbol{x}_{N+1}, \boldsymbol{X}) & k(\boldsymbol{x}_{N+1}, \boldsymbol{x}_{N+1}) \end{bmatrix}\right) \tag{11}$$

where $k(\boldsymbol{X}, \boldsymbol{x}_{N+1})$ is a row vector of the kernel function applied at all previous data points and the new input vector and $k(\boldsymbol{x}_{N+1}, \boldsymbol{X}) = k(\boldsymbol{X}, \boldsymbol{x}_{N+1})^T$ is the transpose. The conditional distribution $p(t_{N+1})$ is a Gaussian distribution with mean and covariance

$$m(\boldsymbol{x}_{N+1}) = \boldsymbol{k}(\boldsymbol{x}_{N+1}, \boldsymbol{X})\boldsymbol{C}_N^{-1}\boldsymbol{t} \tag{12}$$

$$\sigma(\boldsymbol{x}_{N+1}) = k(\boldsymbol{x}_{N+1}, \boldsymbol{x}_{N+1}) - \boldsymbol{k}(\boldsymbol{x}_{N+1}, \boldsymbol{X})\boldsymbol{C}_N^{-1}\boldsymbol{k}(\boldsymbol{X}, \boldsymbol{x}_{N+1}). \tag{13}$$

2.3.4 Kernel functions

The only restriction on the kernel function based on equation (9) is that the covariance matrix given by (10) has to be positive definite as this is a requirement of the Gaussian distribution. Due to the second term in equation (10), adding to the main diagonal of the kernel matrix, it is only required to be positive semi definite since the eigenvalues of $\boldsymbol{C}$ will still be strictly larger than zero.

The properties of the kernel function will affect the behaviour of the resulting regression without necessarily requiring any specific functional form, if, however, the kernel function is defined in terms of a finite set of basis functions, the functional form of the regression will be limited to linear combinations of that set. The kernel function that we have used in this thesis is the exponential kernel[8]

$$k(\boldsymbol{x}, \boldsymbol{x}') = \beta^2 \exp\left(\sqrt{\sum_{i=1}^{d} \frac{(x_i - x_i')^2}{\gamma_i^2}}\right) \tag{14}$$

[7] The Kronecker delta takes the value one when the two indexes are the same and zero otherwise.

[8] Also called Ornstein-Uhlenbeck or Matern 1/2.

since it is universal [MXZ06] and closely related to stochastic volatility [CD10]. β is a scale parameter that determines the average distance of the function from its mean, that is it determines the precision of the predictions. The other hyper parameter γ_i is called the length parameter. The length parameter determines the smoothness of the function as can be seen in Figure 1, displaying randomly generated functions from GP regressions using exponential kernels with different length parameters and zero mean.

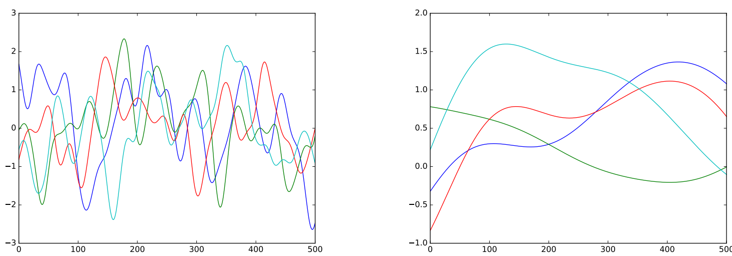

Figure 1: Randomly chosen functions from GP regressions with exponential kernels with different length parameters.

The length parameter is sometimes moved outside of the summation. When it is inside, as in equation (14), it is called automatic relevance detection (ARD) since the parameter can be assigned different values for each dimension, leading to an automatic determination of which input dimensions are most relevant.

The hyper parameters are learned from the data by maximizing the likelihood of the data given the parameter values, $p(\boldsymbol{t}|\boldsymbol{\theta})$ where $\boldsymbol{\theta}$ denotes the set of all the hyper parameters in the model. This is generally an analytically intractable problem that has to be solved using numerical methods. The log likelihood function of a Gaussian is

$$\ln p(\boldsymbol{t}|\boldsymbol{\theta}) = -\frac{1}{2}\ln|\boldsymbol{C}| - \frac{1}{2}\boldsymbol{t}^T\boldsymbol{C}^{-1}\boldsymbol{t} - \frac{N}{2}\ln(2\pi) \tag{15}$$

with the gradient

$$\frac{\partial}{\partial\theta_i}\ln p(\boldsymbol{t}|\boldsymbol{\theta}) = -\frac{1}{2}\mathrm{Tr}\left(\boldsymbol{C}^{-1}\frac{\partial\boldsymbol{C}}{\partial\theta_i}\right) + \frac{1}{2}\boldsymbol{t}^T\boldsymbol{C}^{-1}\frac{\partial\boldsymbol{C}}{\partial\theta_i}\boldsymbol{C}^{-1}\boldsymbol{t}. \tag{16}$$

Equations (15) and (16) can be used for a gradient decent minimization of the negative log likelihood. As equation (15) generally has multiple minima, the optimization algorithm should be restarted with different starting values.

2.4 Prediction performance evaluation

When comparing the accuracy of a predicted series to the actual data the most basic approach is to use some kind of norm to compare the relative distance at each data point. One group of vector norms are

$$L_w^n = \left[\sum_i w_i |r_i - d_i|^n \right]^{\frac{1}{n}} \tag{17}$$

where r_i are the values from the left vector, in this case the forecasted RV, and d_i are the values from the right vector, in this case the RV. w_i are weights corresponding to each data point.

In this thesis we will consider the the L^1 and L^2 norms which both calculate the distance between two discrete data series, the difference is in how they weigh the errors. The L^1 norm sums the absolute distance between the series, and therefore assigns the same importance to all errors, while the L^2 norm assign relatively more importance to larger errors. By including a weight $w_i = 1/d_i$, relative errors can be calculated. The standard error measurements R^2 and MAPE are relative adjustments of these L-norms.

3 Methodology

3.1 Data

When constructing a volatility risk premium using forecasts of realized volatility for the Swedish and the American stock markets, data for each respective market is required. The reason for analyzing both markets is to identify any potential differences, but also to be able to compare the results to previous research, which mainly has focused on the American market. To properly mirror the stock market of each country, the OMXS30 index was chosen for the Swedish market while the S&P 500 index was chosen for the American market. Another consideration in picking stock indices was that there had to be sufficient option trading on the index, so that an implied volatility could be retrieved for it.

The implied volatility for the S&P 500 was retrieved using the CBOE VIX index, which is calculated using a model-free approach based on 30-day put and call options on the S&P 500. Implied volatility data for OMXS30 was retrieved using a similar index, the SIX Volatility index, which on the other hand is calculated using the Black-Scholes formula [Inf14]. Bollerslev [BTZ09] found that model-free estimated implied volatility outperformed those based on any specific model in constructing variance risk premia, but there has yet to arrive a model-free index for the Swedish stock market and calculating one is beyond the scope of this thesis.

Regarding data for the VRP regressions, consumer price index data was collected from the website of the OECD [OEC17], while the 3-Month treasury bill data and currency data for the US, along with default spread data, was collected from the website of St. Louis Federal Reserve [FRE17a; FRE17b]. Data for the Swedish 3-Month treasury bill[9] and currency strength was collected from the website of the Swedish Riksbank [Rik17a; Rik17b].

The data was collected from 5 May 2004 to 22 March 2017 as that was as early as the SIX Volatility index stretched. The stock indices were collected in daily and monthly price data, to calculate realized volatility and to calculate monthly returns respectively, while the implied volatility indices were daily data on 30-days ahead implied volatility. All data except for the regression variables was collected from Thomson Reuters Datastream.

[9] Statsskuldväxel.

3.2 Calculating ex post realized volatility

Calculating the underlying realized volatility of each stock market is essential in evaluating the forecasting power of the various models and an important input variable for some of the models. The daily realized range defined as $\ln(P_{high}/P_{low})$, where P_{high} and P_{low} are the daily high and low prices respectively, has been shown to be a less biased estimate of the realized volatility than the daily returns if multiplied with a factor $(4 * \ln(2))^{-1}$ [MV07]. A rolling window estimation was done $\tau = 21$ days as the length of the rolling window, chosen since it was the average number of trading days in a given 30-day period in the data set over the estimation period. This way the monthly ex post realized volatilities were calculated. However, to make them comparable to the implied volatilities that are reported in percentage form and in annualized numbers, the calculated RV was multiplied by $\sqrt{100^2}$ and by $\sqrt{252}$, the square root of the average number of trading days each year. Ultimately, the formula used for calculating the realized volatility was

$$RV_{t,\tau} = \sqrt{\frac{100^2}{4 * \ln(2)} \frac{252}{\tau} \sum_{i=0}^{\tau} \ln(P_{t-i,high}/P_{t-i,low})^2}. \tag{18}$$

3.3 Forecasting volatility

Volatility was forecasted using Gaussian processes for regression (GP), GARCH methods and a random-walk, naive, approach. The best case scenario would be to accurately forecast 21 days ahead, as that is the equivalent window as the implied volatility is supposed to reflect, but preliminary results indicated that most of the different methods did not perform well for forecasting such long period ahead. As the goal with using a forecasted volatility is to decrease the temporal difference between the risk-neutral and the physical measure of expected future volatility, any forecast horizon could prove useful in constructing ex ante VRPs. Therefore, 1 day ahead, rolling window forecasts were carried out for the different methods.

3.3.1 GARCH methods

The GARCH methods were implemented in the R environment [R D08] using the rugarch package [Gha15]. Learning and forecasts were made using functions provided in the rugarch package. All of the GARCH methods were specified with a model for the mean following an ARMA[10](0,0) and a variance model of order (1,1) for GARCH and GJR-GARCH, whereas the realized GARCH was defined by its default parameters, (2,1). The GARCH methods had a learning period of approximately 3300 training days, which was the maximum length that the available data allowed for. The choice of including GARCH(1,1) as a volatility forecasting method is because it is one of the standard methods in the finance field, whereas the choice of including the GJR-GARCH was to try to capture the asymmetric properties of volatility. The choice of including the realized GARCH model was to try to include the past realized volatility as calculated using equation 18, as the GARCH methods otherwise calculate the past volatility themselves.

3.3.2 GP for regression

The GP for regression were implemented in the Python programming language [Ros95] using the GPy package [GPy12].

One of the more significant differences between GP and GARCH regressions in this context is that the GP can be trained with additional explanatory variables. In this project, however, we have only included the weekday in addition to the historical realized volatility and return residuals. Macroeconomic variables are difficult to include since they are generally calculated and released infrequently, and often delayed compared to daily data. Macro variables therefore have higher potential as input variables for longer forecast horizons.

Weekdays were included in the form of an array with five elements representing each day. The current day was indicated with a one and the other days with zeros. In order to provide a balance between the access to data and the number of input parameters 16 days of return residuals were supplied to the GP and two days of previous volatility. The GP was also supplied future volatility as training data over the training window, ending the same number of days before the prediction is made as the forecast horizon, as it would otherwise be trained with future data compared to the day the forecast is made. The GP was trained with 300 trading days of previous data, as this seemed to give a good balance between performance and computational complexity.

[10] Auto Regressive Moving Average. The (0,0) version is essentially white noise.

3.4 VRP regression

To estimate a potential relationship between the constructed volatility risk premia and the excess returns for the different stock markets, linear regressions were run. As some data for the Swedish market was missing, such as dividend yield and historical P/E-numbers, these variables were dropped in the regressions for the American VRP as well. Instead, a set of macroeconomic variables were included in the regression models; inflation, currency strength and the credit default spread, as these variables were readily available for both countries over the entire considered period.

The change in consumer prices was included as a measure of inflation, which theoretically should affect the stock market [DeF+91]. The percentage change of the respective currencies' strength against a basket of foreign currencies, denoted TCW for Total Competitiveness Weights, was included since it theoretically might influence the S&P 500 and OMXS30, because they are composed of the largest companies in each country, with considerable foreign operations. Finally, the default spread, defined as the spread between the corporate bond yield of Moody's BAA and AAA credit ratings, was included as it could work as a leading indicator of the business cycle [GL00].

Due to the macro variables only being available on a monthly basis, the regressions were run using monthly data. This meant that the daily, rolling window, data of monthly volatility estimates had to be aggregated using monthly averages, and the daily excess log returns of the stock markets were aggregated using summation.

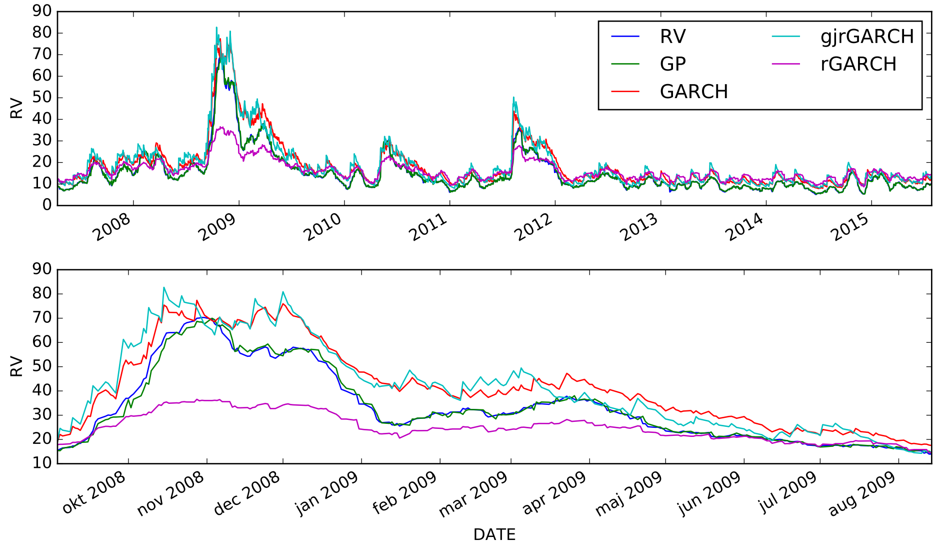

Figure 2: One day ahead forecast of S&P 500 volatility. The top panel displays the whole period for which predictions were generated and the bottom an interval including the 2008 financial crisis.

4 Results

4.1 Volatility Forecasts

One day forecast by the GP and GARCH methods are plotted together with the actual RV in Figures 2-3 for the S&P 500 and the OMXS30 respectively. The plots in the top panels span the entire term for which we had complete data, from 13 April 2007 to 28 Juli 2015 for a total of 2 078 trading days. From the figures it is apparent that the GARCH methods generally give higher estimates than the RV calculated with equation (18), and that GARCH and GJR-GARCH react sharply at volatility jumps. The realized GARCH does not have this tendency, instead it underestimates the volatility after jumps, most notably in 2008.

The GP follows the RV closely and for the sake of visual clearness only the GP without weekday dummies is included in the figures. As can be seen from the lower panels, the GP predictions are often delayed one day after the RV when there are volatility jumps. The addition of the weekday dummies have little effect on the forecasted volatility as can be seen in Tables 1-2.

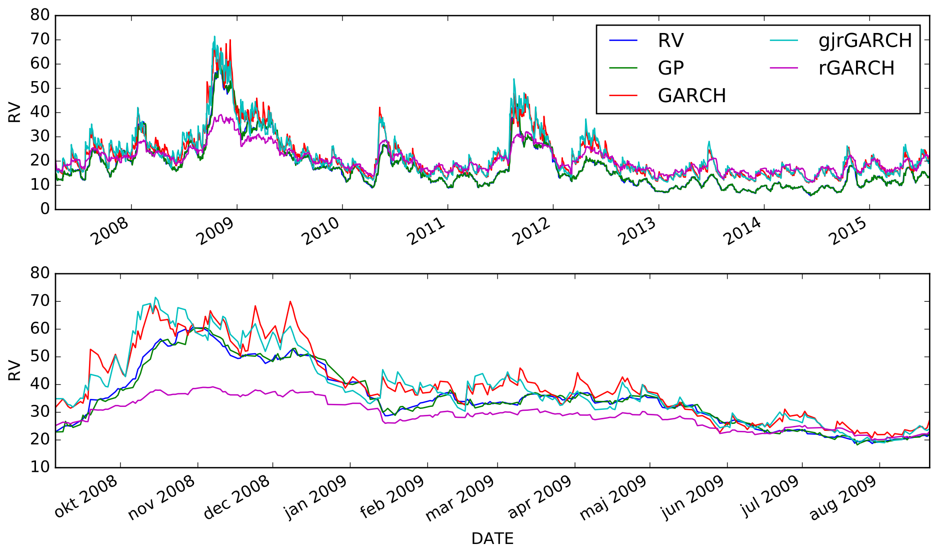

Figure 3: One day ahead forecast of OMXS30 volatility. The top panel displays the whole period for which predictions were generated and the bottom an interval including the 2008 financial crisis.

The GP performed slightly better on OMXS30 with dummies than without, but the opposite is true on S&P 500.

It can also be seen from Tables 1-2 that the GP is outperformed by the naive predictions for all measurements included. Due to the vertical dislocation of the GARCH predictions relative to the RV the L-norms are not applicable for determining the prediction accuracy for these methods.

The relative smoothness of the GP and the naive predictions compared to the different GARCH method predictions makes the Kendall correlation higher between these series and the RV. Since the Kendall correlation is calculated on the complete data-set we have also included the proportion of predictions that have the right sign of change compared to the last volatility. In this category all the GARCH methods outperform the naive method and the GP, with the realized GARCH as the highest scoring method.

4.2 GP predictive accuracy

The uncertainty of GP easily available since the outputs consist of the mean and standard deviation of a Gaussian distribution. This cannot, however,

Table 1: S&P500 prediction fit

	Naive	GP	GP weekday	GARCH	GJR-GARCH	real-GARCH
L^1	**756**	946	955	7 846	8 063	8 026
L^2	**904**	1 338	1 396	42 999	54 318	54 318
L^1_w	**52**	62	62	582	591	625
L^2_w	**45**	60	60	2 466	2973	2 910
Corr	**0.949**	0.942	0.942	0.854	0.791	0.869
Sign	0.608	0.576	0.570	0.670	0.588	**0.742**

Table 2: OMXS30 prediction fit

	Naive	GP	GP weekday	GARCH	GJR-GARCH	real-GARCH
L^1	**817**	994	986	9 964	10 509	10 150
L^2	**893**	1 233	1 219	64 890	75 578	67 148
L^1_w	**47**	56	56	720	756	764
L^2_w	**41**	54	53	4 400	5 080	4 719
Corr	**0.958**	0.951	0.951	0.763	0.723	0.797
Sign	0.598	0.578	0.575	0.655	0.633	**0.712**

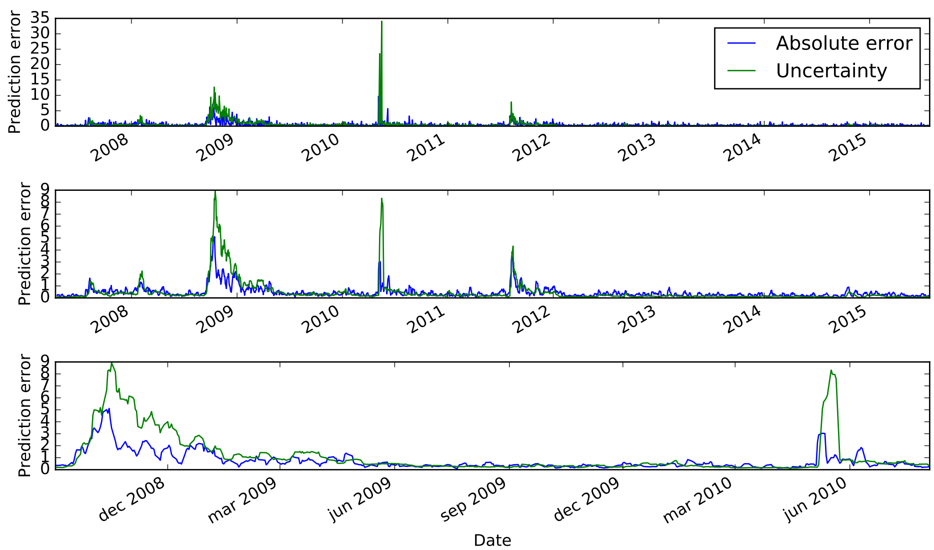

Figure 4: Absolute errors ($|\mathrm{RV} - \mathrm{GP}_{pred}|$) and standard deviation of the GP forecasts on S&P 500. The first panel displays the daily values while the second and third panels display five days trailing averages.

be used to improve the predictive accuracy, but can potentially determine how much confidence is put into the forecast at different times. Figure 4 displays the absolute value of the difference between the actual RV and the GP predictions. The top panel displays the daily values while the middle and bottom panels display five days trailing averages in order to make the plots more accessible. The first two panels display the entire period and the third is focused on a period covering the financial crisis in 2008 and the 2010 Flash Crash.

From the figure it is apparent that the standard deviation of the predictions did not rise at the same time as the predictive errors in June 2010. In 2008, however, the standard deviation increased simultaneously with the forecast error. The standard errors of the output could therefore potentially be used as a measure of the uncertainty of the predictions, but the predictions can be erroneous without it showing in the standard errors before the event.

Another property that is apparent from the Figure is that the volatility jumps in 2008, 2010 and 2011, have delayed effects on the outputted standard deviation on later days. This effect is due to a few days with different properties get disproportionately big effects on the learning.

4.3 Summary statistics of VRP regressions

As can be seen in Tables 3-4, the different forecasting methods gave varying estimates of the volatility risk premium, with the Gaussian processes and the naive approach both being the largest in the US and in the Swedish regressions, followed by the VRP calculated from the realized GARCH forecasts. Only the GP VRP that provided the highest explanatory power was included in the regressions. While the performance of the two, in terms of adjusted R^2, was very similar, the VRP constructed from the GP without weekday dummies, was slightly better.

The average risk premium across all forecasting methods is 5.0 for the American market and 2.0 for the Swedish market, a difference of three percentage points. In other words, there is a larger spread between the realized and implied volatility in the American market, perhaps indicating a larger degree of risk aversion in this market than in the Swedish.

The volatility risk premia from the GARCH and the GJR-GARCH have similar characteristics, and are on both markets approximately 4 percentage points lower than the GP-forecasted VRP and around 2 percentage points lower than the realized GARCH VRP. As the realized GARCH and the Gaussian processes models both take into account the past realized volatility in the forecasts, whereas the other GARCH models are based on past returns and volatility calculated only by the models themselves, this might serve as an explanation since it seems like the estimates are on average lower, if the past realized volatility is included in the model. For the Swedish market, this difference, perhaps upward bias, in the GARCH-forecasts means that the VRPs are negative on average, which is not in line with previous research and puts to question these GARCH-forecasts.

Studying the correlation matrices for both countries, one can notice that there is on average, higher correlation between stock market excess return and the risk premia for the S&P 500 than for the OMXS30. In the American market, the GARCH and the GJR-GARCH VRPs are more distinctly correlated with returns compared to the other VRPs, than in the Swedish market. Furthermore, the high correlation between the Gaussian processes calculated VRP and the one derived from a naive approach, is to be expected as the GP predicted volatility was quite similar to the previous period's volatility.

The plots of the VRPs, see Figure 5, sheds light on the volatile nature of the VRP. The VRPs for both markets look fairly similar, with a more obvious spread between the GARCH-calculated VRP and the others in the Swedish market than in the US. An interesting difference between the plots is that the VRP on OMXS30 generally is smaller than that of the S&P500, illustrating the previously mentioned difference in averages.

4.4 VRP Regression Results

As can be seen from the regression output in Tables 5-6, the results show that there is a significant relationship between the VRP and stock market excess returns. This means, in the case of the GARCH-VRP, that an increase in the VRP, i.e. the percentage point spread between IV and RV, with 1 percentage point is, on average, followed by an increase in the following month's excess returns by 0.56 percentage points for the S&P 500, and 0.59 percentage points for OMXS30. If the VRP is an indicator of investor risk aversion, the results show that an increase in risk aversion among stock market investors is correlated with higher stock market excess return.

For the US market, the significance of the coefficient estimate for the VRPs depends on which forecasting method that was used, as only the

Table 3: Summary Statistics of Variables Used in US Regressions

	Ret.	VRPgp	VRPgch	VRPgjr	VRPreal	VRPnav	Infl.	TCW	Def.Sp.
Nr. obs	100	100	100	100	100	100	100	100	100
Max	10.23	15.648	8.6	9.477	28.408	15.761	5.6	6.684	56.627
Mean	0.269	6.779	3.096	3.001	5.473	6.771	1.884	0.135	0.604
Min	-18.595	1.009	-9.688	-9.8	-0.844	1.114	-2.097	-4.103	-25
Std.Dev.	4.711	2.456	2.745	3.174	5.576	2.484	1.549	1.828	10.703
				Correlation matrix					
Ret.	1	0.129	0.341	0.42	-0.047	0.15	-0.334	-0.129	-0.216
VRPgp		1	0.614	0.527	0.424	0.996	-0.148	-0.149	-0.263
VRPgch			1	0.863	-0.247	0.623	-0.064	-0.199	-0.414
VRPgjr				1	-0.277	0.55	-0.254	-0.355	-0.559
VRPreal					1	0.408	-0.091	0.128	0.315
VRPnav						1	-0.16	-0.17	-0.296
Infl.							1	0.155	0.359
TCW								1	0.495
Def.Sp.									1

Table 4: Summary Statistics of Variables Used in Swedish Regressions

	Ret.	VRPgp	VRPgch	VRPgjr	VRPreal	VRPnav	Infl.	TCW	Def.Sp.
Nr. obs	99	99	99	99	99	99	99	99	99
Max	15.664	10.67	6.58	6.567	23.699	10.481	4.372	4.652	56.627
Mean	0.099	4.511	-0.116	-0.336	1.566	4.519	1.153	0.058	0.722
Min	-18.563	-1.952	-6.081	-8.236	-4.438	-2.253	-1.872	-5.635	-25
Std.Dev.	5.159	2.14	2.577	2.72	5.064	2.142	1.525	1.583	10.691
				Correlation matrix					
Ret.	1	0.226	0.252	0.235	0.126	0.227	-0.395	-0.026	-0.091
VRPgp		1	0.565	0.404	0.388	0.998	0.019	0.103	-0.037
VRPgch			1	0.892	0.379	0.552	0.003	-0.27	-0.273
VRPgjr				1	0.336	0.399	-0.106	-0.349	-0.337
VRPreal					1	0.377	0.335	0.246	0.333
VRPnav						1	0.01	0.1	-0.045
Infl.							1	0.105	0.366
TCW								1	0.334
Def.Sp.									1

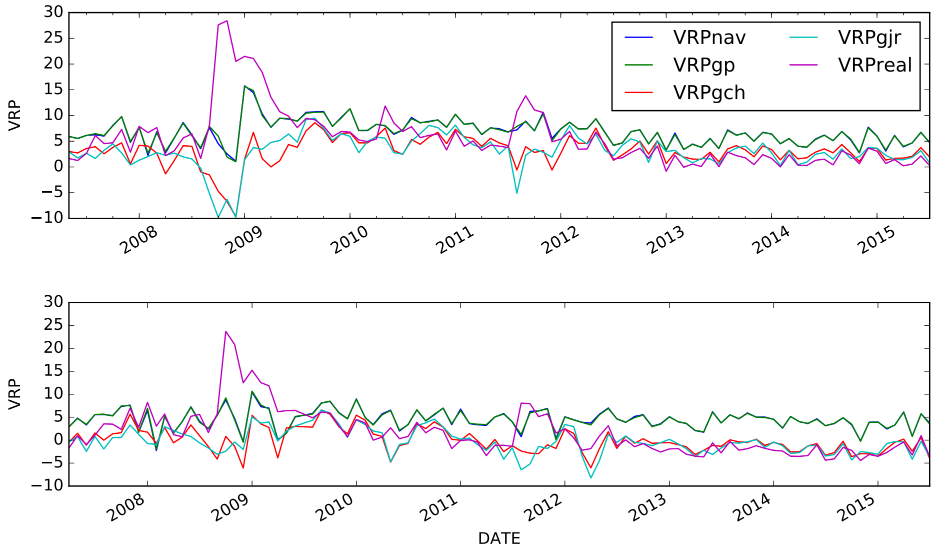

Figure 5: Volatility risk premia predicted by the different forecasting methods for S&P 500 and OMXS30 respectively.

GARCH and GJR-GARCH gave significant estimates, while the Swedish market shows a significant correlation between the VRPs and equity excess returns no matter the forecasting method. All of the estimates for the coefficients of the risk premia in the regression equations are positive, and even if some of them are insignificant, the results indicate that stock market excess return and the VRP are positively correlated.

Regarding the explanatory power of the VRP regressions for next month's stock market excess return, the GARCH forecasted VRP performs best, followed by the one from the GJR-GARCH. The adjusted R^2s for the US market GARCH and GJR-GARCH VRP regressions are 0.24 for both, while for the OMXS30 the same values are 0.19 and 0.17 respectively. The fact that the adjusted R^2s are lower for the Swedish market regressions, is a trend holds for all different VRPs except the one using a realized volatility calculated using the realized-GARCH method.

Regarding the macro-variables included in the regressions, only inflation has a significant correlation with the following month's excess return, while the currency strength and default spread are insignificant. In line with what would be expected, an increase in consumer prices, i.e. higher inflation, leads to lower excess return the following month. The addition of the macro-variables, increased the explanatory power of the models, and did not have any effect on the significance of the VRP coefficient estimates.

The regression results are in line with previous research, as Bollerslev et. al. (2009) found that the coefficient for the VRP in regressions on S&P 500 monthly excess return is 0.39 [BTZ09].

4.5 Robustness of regressions

The regressions that were carried out to identify and quantify a potential relationship between the VRP and the following month's stock market excess return were thoroughly checked for common errors that arise in linear regressions. The regression diagnostics were quite similar for the models including VRPs from GP, naive and realized GARCH, whereas they differed a bit for the GARCH and GJR-GARCH VRP regression models. Therefore, plot diagnostics for the GP and the GJR-GARCH are attached in appendix, see section 5.

First, as the regression models are constructed in a linear fashion, if the relationship between the explanatory variables and the dependent variable is non-linear this might show up in the residuals of the regression. Therefore, plots of the residuals against the fitted values of the dependent variable were examined, see Figure 6. A linear relationship would show up as the dots evenly spread out around a horizontal line, without any obvious pattern [Kab15]. As can be seen from the red trend lines, no obvious pattern is observed and the linearity assumption of the regressions are fulfilled.

Second, another assumption of the linear regression models is that the residuals have an expected value of zero, with finite variance [Han17]. An easy way to check this assumption is if the residuals lie on a linear line when plotted against e.g. an $\mathcal{N}(0,1)$ distribution. As can be seen in QQ-plots in Figure 7, this seems to be the case, with the exception of some outliers, notably observation 18 and 22 in the US data and observation 18 and 14 in the Swedish data. Removal of these observations to enhance the regression models would however create a bias, as these observations are related to the high volatility jumps of the 2008 financial crisis. Therefore they are kept in the model.

Furthermore, to examine if the residuals are identical or not, an important consideration since robust standard errors were used in the regressions, scale-location plots were analyzed, see Figure 8. If the points in the plots are

Table 5: Results from monthly regressions, US

	Dependent variable:				
	Market excess return				
	GP	GARCH	gjrGARCH	rGARCH	Naive
	(1)	(2)	(3)	(4)	(5)
VRPgp	0.239				
	(0.172)				
VRPgch		0.562^{***}			
		(0.164)			
VRPgjr			0.550^{***}		
			(0.161)		
VRPreal				0.033	
				(0.081)	
VRPnav					0.270
					(0.172)
Infl.	-0.850^{***}	-1.026^{***}	-0.801^{***}	-0.840^{***}	-0.854^{***}
	(0.281)	(0.283)	(0.291)	(0.297)	(0.283)
TCW	−0.028	−0.005	0.117	−0.077	−0.022
	(0.256)	(0.257)	(0.266)	(0.263)	(0.257)
Def.Sp.	−0.046	0.008	0.019	−0.061	−0.042
	(0.047)	(0.050)	(0.053)	(0.051)	(0.048)
Constant	0.632	0.709	0.374	2.012^{**}	0.438
	(1.370)	(0.810)	(0.871)	(0.882)	(1.378)
Adj. R Sq.	0.153	0.237	0.235	0.131	0.157
Observations	100	100	100	100	100
Residual Std. Error (df = 95)	3.474	3.707	3.655	3.734	3.525
Note:			*p<0.1; **p<0.05; ***p<0.01		

Table 6: Results from monthly regressions, SWE

	Dependent variable:				
	Market excess return				
	GP	GARCH	gjrGARCH	rGARCH	Naive
	(1)	(2)	(3)	(4)	(5)
VRPgp	0.483**				
	(0.206)				
VRPgch		0.591***			
		(0.172)			
VRPgjr			0.526***		
			(0.170)		
VRPreal				0.231**	
				(0.092)	
VRPnav					0.479**
					(0.205)
Infl.	−1.046***	−1.203***	−1.087***	−1.204***	−1.045***
	(0.308)	(0.294)	(0.296)	(0.306)	(0.307)
TCW	−0.034	0.258	0.391	0.061	−0.035
	(0.295)	(0.284)	(0.292)	(0.286)	(0.294)
Def.Sp.	0.033	0.056	0.054	−0.005	0.034
	(0.046)	(0.045)	(0.046)	(0.045)	(0.046)
Constant	−0.644	1.772***	1.816***	1.478***	−0.628
	(1.070)	(0.529)	(0.537)	(0.536)	(1.070)
Adj. R Sq.	0.134	0.189	0.173	0.148	0.133
Observations	99	99	99	99	99
Residual Std. Error (df = 94)	3.752	3.784	3.886	3.332	3.825
Note:					*p<0.1; **p<0.05; ***p<0.01

randomly scattered across a horizontal line, this would indicate homoscedasticity [Kab15]. This seems not to be the case in our regressions, supporting the assumption of heteroscedasticity and robust standard errors.

Finally, as the power of a regression model can be weakened by multicollinearity, i.e. high correlation between the explanatory variables, an analysis of the interdependence between the independent variables in the regressions were carried out. As can be seen in the correlation matrices in Tables 3-4, the largest correlation between variables in the US regressions is between the VRP constructed from the GJR-GARCH and the default spread variable, amounting to -0.55. For the Swedish market, the largest correlation is between inflation and the default spread, amounting 0.366. These correlations should not give rise to any significant multicollinearity issues [Han17].

5 Summary

The GP regression did not produce one day ahead predictions as similar to the next day realized volatility as the naive approach, or as good at predicting of the direction of change as the GARCH methods. The best predictor of the direction of change to the next day was the realized GARCH, which is also the newest addition of the GARCH methods included in this thesis. The predictions by the realized GARCH did, however, often underestimate the effect of volatility jumps, making GARCH$(1, 1)$ the best performer of the GARCH methods when looking at closeness of fit to the ex post RV.

The mean of the VRPs on the American market indicates that the average investor is willing to pay a premium of 5 percentage points volatility to hedge themselves against its uncertain nature. This is the equivalent of paying a premium of more than a third of the annualized volatility on the S&P 500, which amounted to 14.85% over the considered period. For the Swedish market, investors were willing to pay a premium of 2 percentage points volatility to hedge themselves, the equivalent of paying a premium of approximately 11% of the annualized volatility on the OMXS30, which amounted to 17.6% over the considered period. In other words, the existence of a volatility risk premium on the Swedish market is verified, giving some potential insight into investor risk aversion.

Considering the volatile nature of the risk premium illustrated in Figure 5, this thesis suggests that the VRP is not constant over time. The biggest differences arise in times of large volatility jumps, where the VRP decreases, and sometimes even reach values below zero, as the event takes place, illustrating the unexpected nature of volatility shocks as the realized volatility increases considerably relative to the implied volatility. These decreases in the VRP are followed by the largest observed values during the considered period, effectively illustrating how investors irrationally missprice volatility following a jump such as the 2008 financial crisis, as the actual volatility decreased quicker than investors believed. To rationally explain these characteristics, one could consider the concept of black swan events from the field of behavioral finance [Tal07]. Such an event is extremely difficult to foresee and results in outcomes that deviates significantly from previous trends. The events that give rise volatility jumps in stock markets could be seen as black swans, and investors willingness to pay a premium for volatility, a hedge against these unpredictable events, as they tend to lead to negative stock returns.

There are similarities between the VRP of the Swedish market and the US market, as both were positive on average, however with the American VRP being larger in most cases. The results indicate that the average investor on the US market is more risk averse than the average investor on the Swedish market. The comparison of magnitudes is merely an indication and one must take into consideration that the two different countries' implied volatility data are constructed using different methods. This could potentially mean that there is a bias in one of the countries' implied volatility data, weakening the comparison of magnitudes to some extent.

The VRP predicts stock market returns in both the Swedish and the US market. Previous research has found that using a model-free approach was essential in constructing a VRP that provides explanatory power for stock market returns [BTZ09]. This thesis suggests the opposite, as significant results were received for both markets, the reliance on a model-free approach is not essential for predicting stock market returns. Regarding the fit of the regression models for the different volatility forecasting methods, the GARCH-based model gave the best adjusted R^2, which is a bit surprising considering its negative mean for the OMXS30. The GP-based model was not the best volatility forecasting method for constructing a VRP with high explanatory power for stock market excess return, further putting to question its volatility forecasting power. An interesting note is that the GP-VRP without weekday dummies slightly outperformed the model with dummies included.

The conclusions regarding the performance of the GP as a volatility forecast method and its usefulness in constructing a VRP relies heavily on the calculation of ex post realized volatility, as this is the main input data for the algorithm. If equation 18 created a biased estimate of RV, it will have translated into the forecasts of the Gaussian process method. However, when comparing the realized volatility calculated to data of realized volatility from the Oxford-Man Institute of Quantitative Finance, the two time series are similar [GS09]. The Oxford-calculated RV is on average 0.22 percentage points annualized volatility higher than the one used in this thesis, showing a small downward bias in the RV used in this thesis, that is slightly more prominent during the highly volatile period of the 2008 financial crisis.

We did not see any benefit on the GP predictions when including the weekday as an input variable, this can be an effect of the weekday not adding any explanatory value, but it can also be dependent on how the information is encoded. An alternative encoding method that we did not test is to include a variable with the number of trading, or non-trading, days before the prediction date. Another possibility is that the GP requires more specialized kernel functions for evaluating potential periodicity in the volatility. One important direction of future research is to look at more specialized kernel functions for stock market volatility predictions.

Another possible deficiency with the GP implementations used for this examination is that they have been homoscedastic. Heteroscedastic GP, does, however, require significantly more processing time for training and evaluation. One way of mitigating this is by developing a specialized online[11] GP that is suitable for stock market volatility data.

For future research, one could try to construct VRPs based on volatility forecasts with a longer horizon, to further reduce the temporal difference between the implied and the realized volatility. Another suggestion is for future research to investigate the strength of GP volatility forecast using an ex post realized constructed from intraday returns, as opposed to a daily realized range, which could potentially give a better estimate of future volatility.

[11]Continuously adding data new points when learning instead of evaluating batches.

References

[AB98] Torben G Andersen and Tim Bollerslev. "Answering the skeptics: Yes, standard volatility models do provide accurate forecasts". In: *International economic review* (1998), pp. 885–905.

[BD03] Michael W Brandt and Francis X Diebold. *A no-arbitrage approach to range-based estimation of return covariances and correlations.* Tech. rep. National Bureau of Economic Research, 2003.

[BGZ11] Tim Bollerslev, Michael Gibson, and Hao Zhou. "Dynamic estimation of volatility risk premia and investor risk aversion from option-implied and realized volatilities". In: *Journal of econometrics* 160.1 (2011), pp. 235–245.

[BH14] Geert Bekaert and Marie Hoerova. "The VIX, the variance premium and stock market volatility". In: *Journal of Econometrics* 183.2 (2014), pp. 181–192.

[Bis06] Christopher M. Bishop. *Pattern Recognition and Machine Learning (Information Science and Statistics).* Secaucus, NJ, USA: Springer-Verlag New York, Inc., 2006. ISBN: 0387310738.

[BN00] Mark Britten-Jones and Anthony Neuberger. "Option prices, implied price processes, and stochastic volatility". In: *The Journal of Finance* 55.2 (2000), pp. 839–866.

[Bol86] Tim Bollerslev. "Generalized autoregressive conditional heteroskedasticity". In: *Journal of econometrics* 31.3 (1986), pp. 307–327.

[BTZ09] Tim Bollerslev, George Tauchen, and Hao Zhou. "Expected stock returns and variance risk premia". In: *Review of Financial studies* 22.11 (2009), pp. 4463–4492.

[CD10] Nicolas Chapados and Christian Dorion. "Gaussian Processes and Non-Parametric Volatility Forecasting". In: *Citeseer (Draft)* (2010).

[CW09] Peter Carr and Liuren Wu. "Variance risk premiums". In: *Review of Financial Studies* 22.3 (2009), pp. 1311–1341.

[DeF+91] Robert H DeFina et al. "Does inflation depress the stock market". In: *Business Review, Federal Reserve Bank of Philadelphia* 3 (1991), pp. 3–12.

[DRS16] Pasquale Della Corte, Tarun Ramadorai, and Lucio Sarno. "Volatility risk premia and exchange rate predictability". In: *Journal of Financial Economics* 120.1 (2016), pp. 21–40.

[Eng82] Robert F Engle. "Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation". In: *Econometrica: Journal of the Econometric Society* (1982), pp. 987–1007.

[FRE17a] St. Louis Federal Reserve FRED. *3-Month Treasury Bill: Secondary Market Rate [TB3MS]*. https://fred.stlouisfed.org/series/TB3MS. Accessed: 2017-04-30. 2017.

[FRE17b] St. Louis Federal Reserve FRED. *Trade Weighted U.S. Dollar Index: Major Currencies [DTWEXM]*. https://fred.stlouisfed.org/series/DTWEXM. Accessed: 2017-04-30. 2017.

[Gha15] Alexios Ghalanos. *rugarch: Univariate GARCH models.* R package version 1.3-6. 2015.

[GJR93] Lawrence R Glosten, Ravi Jagannathan, and David E Runkle. "On the relation between the expected value and the volatility of the nominal excess return on stocks". In: *The journal of finance* 48.5 (1993), pp. 1779–1801.

[GL00] Mark Gertler and Cara S Lown. *The information in the high yield bond spread for the business cycle: evidence and some implications*. Tech. rep. National bureau of economic research, 2000.

[GPy12] GPy. *GPy: A Gaussian process framework in python*. http://github.com/SheffieldML/GPy. 2012.

[GS09] Neil Shephard Gerd Heber Asger Lunde and Kevin K. Sheppard. *Oxford-Man Institute's realized library, version 0.2*. http://realized.oxford-man.ox.ac.uk/data/download. Oxford-Man Institute, University of Oxford. Accessed: 2017-04-30. 2009.

[Han17] Bruce E. Hansen. *Econometrics*. University of Wisconsin. Wisconsin, USA, 2017. URL: http://www.ssc.wisc.edu/~bhansen/econometrics/Econometrics.pdf.

[HHS12] Peter Reinhard Hansen, Zhuo Huang, and Howard Howan Shek. "Realized garch: a joint model for returns and realized measures of volatility". In: *Journal of Applied Econometrics* 27.6 (2012), pp. 877–906.

[HL05] Peter R Hansen and Asger Lunde. "A forecast comparison of volatility models: does anything beat a GARCH (1, 1)?" In: *Journal of applied econometrics* 20.7 (2005), pp. 873–889.

[Inf14] SIX Financial Information. *SIX Volatility Index*. http://www.six.se/hs-fs/hub/370766/file-1935114639-pdf/SIX_Sites/SIX_Index/Produktblad/SIX_Volatility_Index/SIXVX_Volatility_Index.pdf?t=1487692999096, 2014.

[JT05] George J Jiang and Yisong S Tian. "The model-free implied volatility and its information content". In: *Review of Financial Studies* 18.4 (2005), pp. 1305–1342.

[Kab15] Robert Kabacoff. *R in action: data analysis and graphics with R*. Manning Publications Co., 2015.

[M+12] Philippe Mueller, Andrea Vedolin, Yu-min Yen, et al. *Bond variance risk premia*. Financial Markets Group, London School of Economics and Political Science, 2012.

[MV07] Martin Martens and Dick Van Dijk. "Measuring volatility with the realized range". In: *Journal of Econometrics* 138.1 (2007), pp. 181–207.

[MXZ06] Charles A Micchelli, Yuesheng Xu, and Haizhang Zhang. "Universal kernels". In: *Journal of Machine Learning Research* 7.Dec (2006), pp. 2651–2667.

[OEC17] OECD. *Inflation (CPI) (indicator)*. https://data.oecd.org/price/inflation-cpi.htm. Accessed: 2017-04-30. 2017. DOI: 10.1787/eee82e6e-en.

[PG03] Ser-Huang Poon and Clive WJ Granger. "Forecasting volatility in financial markets: A review". In: *Journal of economic literature* 41.2 (2003), pp. 478–539.

[PS14] Marcel Prokopczuk and Chardin Wese Simen. "The importance of the volatility risk premium for volatility forecasting". In: *Journal of Banking & Finance* 40 (2014), pp. 303–320.

[R D08] R Development Core Team. *R: A Language and Environment for Statistical Computing*. ISBN 3-900051-07-0. R Foundation for Statistical Computing. Vienna, Austria, 2008. URL: http://www.R-project.org.

[Rik17a] Sveriges Riksbank. *Sök räntor & valutakurser, statsskuldsväxel 3 månader*. http://www.riksbank.se/sv/Rantor-och-valutakurser/Sok-rantor-och-valutakurser. Accessed: 2017-04-30. 2017.

[Rik17b] Sveriges Riksbank. *Sök räntor & valutakurser, TCW-index*. http://www.riksbank.se/sv/Rantor-och-valutakurser/Sok-rantor-och-valutakurser. Accessed: 2017-04-30. 2017.

[Ros95] Guido Rossum. *Python Reference Manual*. Tech. rep. Amsterdam, The Netherlands, The Netherlands, 1995.

[RW05] Carl Edward Rasmussen and Christopher K. I. Williams. *Gaussian Processes for Machine Learning (Adaptive Computation and Machine Learning)*. The MIT Press, 2005. ISBN: 026218253X.

[Tal07] Nassim Nicholas Taleb. *The black swan: The impact of the highly improbable*. Vol. 2. Random house, 2007.

[TL11] Michalis K Titsias and Miguel Lázaro-Gredilla. "Variational heteroscedastic Gaussian process regression". In: *Proceedings of the 28th International Conference on Machine Learning (ICML-11)*. 2011, pp. 841–848.

[WG10] Andrew Wilson and Zoubin Ghahramani. "Copula processes". In: *Advances in Neural Information Processing Systems*. 2010, pp. 2460–2468.

[WHG14] Yue Wu, José Miguel Hernández-Lobato, and Zoubin Ghahramani. "Gaussian process volatility model". In: *Advances in Neural Information Processing Systems*. 2014, pp. 1044–1052.

Appendix

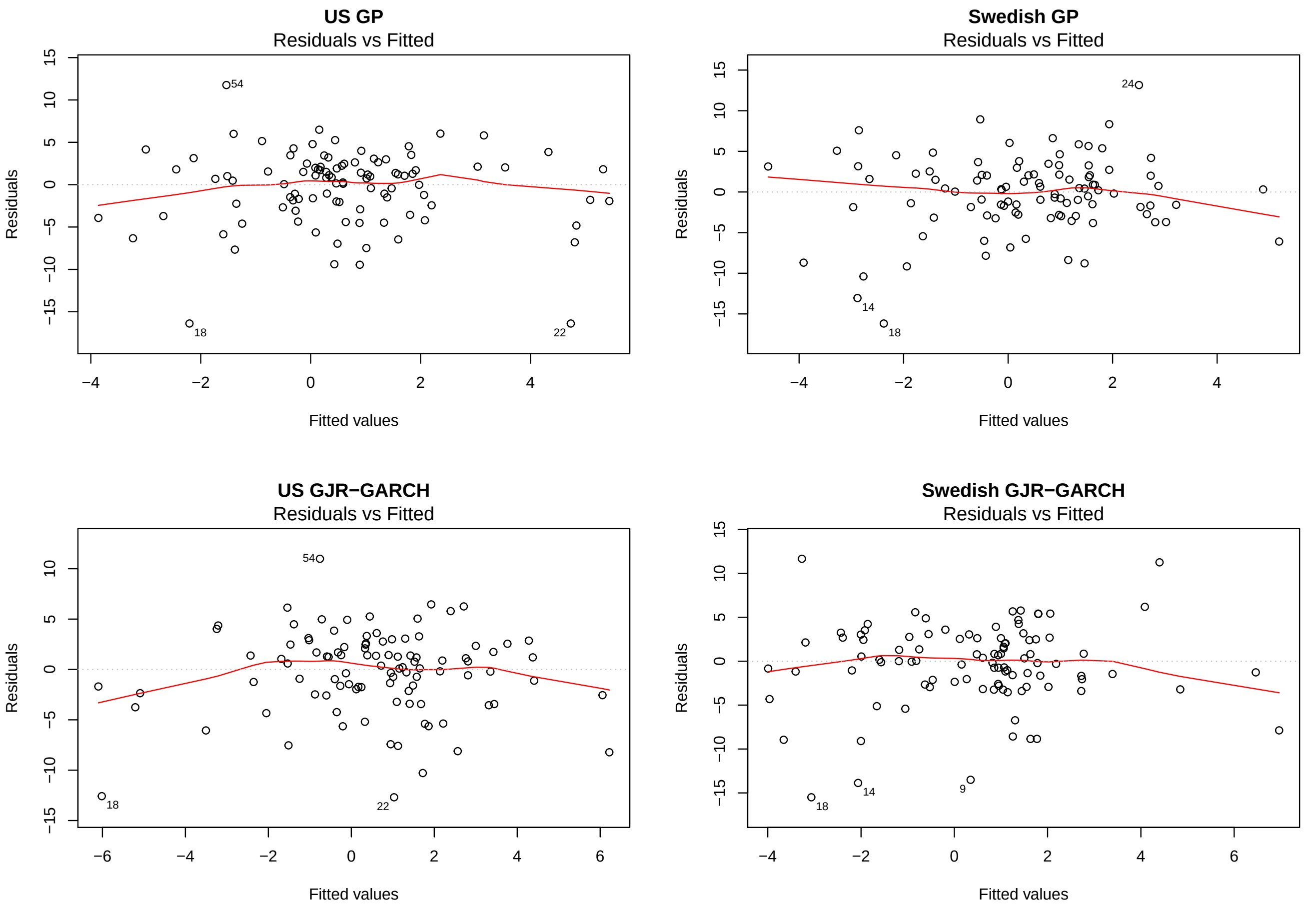

Figure 6: Diagnostic plots of the regression residuals vs the fitted values of the dependent variables, including red trend lines, for the GP and GJR-GARCH in the US and Swedish market respectively.

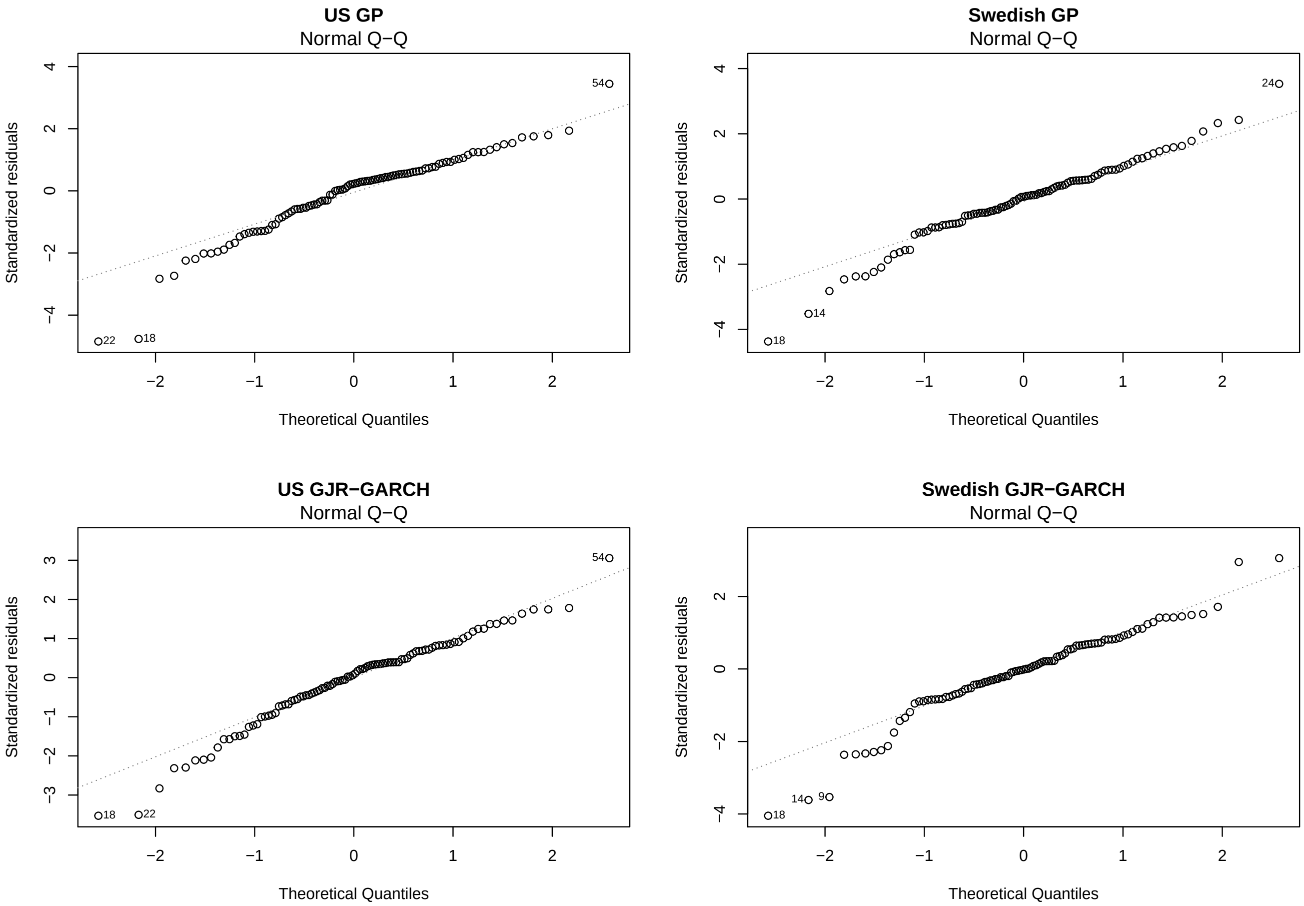

Figure 7: Diagnostic plots of the regression residuals vs an $\mathcal{N}(0,1)$ distribution for the GP and GJR-GARCH in the US and Swedish market respectively. A clear linear trend line would indicate that the residuals follow a normal distribution.

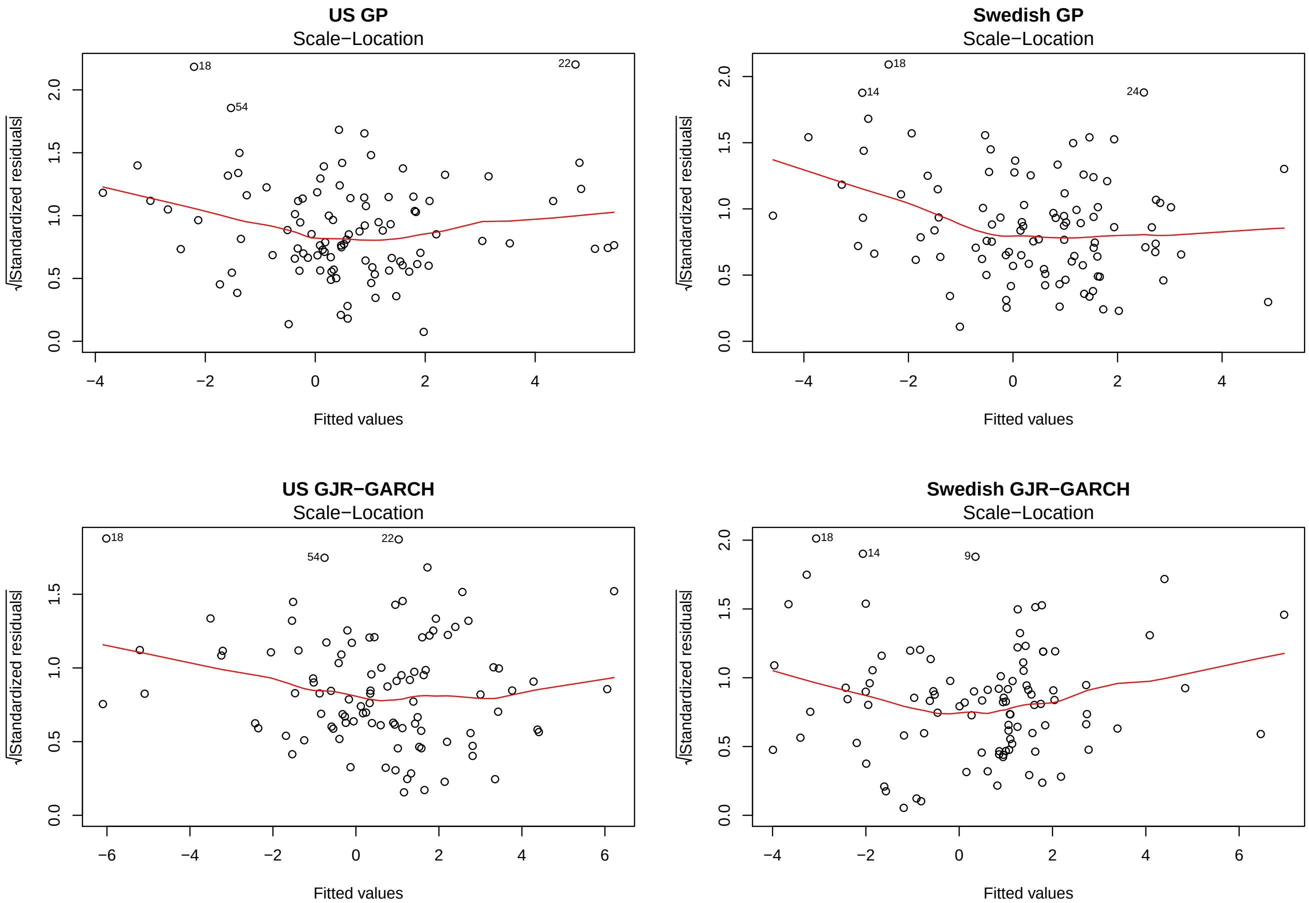

Figure 8: Scale-Location plots for the GP and GJR-GARCH in the US and Swedish market respectively. Homoscedasticity would appear as a random scattering around a horizontal line.