Performance of Stochastic Volatility and GARCH Models in Different Market Regimes

Erik Lundgren & Felix Viitanen

January 2022

Supervisors:
Krzysztof Podgórski, Lund University
Alessandro Mercuri, Nordea

STAH11: Bachelor Thesis in Statistics
15 ECTS

Lund University
School of Economics and Management
Department of Statistics

Abstract

Reliable methods for estimating financial return volatility are crucial in many areas of trading and investing. Two such frameworks, the GARCH and SV, have been of particular interest to academics and practitioners alike. The GARCH model describes the variance of the current innovation as a function of the actual sizes of the previous innovations. In contrast, the stochastic volatility model describes volatility as a latent variable following a stochastic process. This thesis attempts to extend the research conducted by Lopes and Polson (2010) by analyzing the performance of the Gaussian GARCH(1,1) and basic SV model on the SP500 and OMXS30 before and during the endogenous credit crisis, as well as before and during the exogenous COVID-19 pandemic. The results indicate that the SV model consistently fits the data better than the GARCH model on all data sets, while the fit for both models became worse during the periods of market stress, and even more so for the pandemic. In regards to the volatility estimation performance, the GARCH model tends to be better for periods with low volatility, while the performance is similar in highly volatile climates. Finally, the pandemic appeared to be the stress event that had the largest negative impact on the model validation.

Keywords: Stochastic volatility, Bayesian econometrics, Volatility, Endogenous shocks, Exogenous shocks, Market stress, Model comparison, SP500, OMXS30.

Contents

1 Introduction

Volatility and its dynamics play an important role for financial market participants, having implications for derivative pricing, credit spreads, portfolio management, risk management, and other financial and economic issues. By way of illustration, volatility forecasts are used as inputs for market and credit risk management systems (including those employed for determining banks' economic and regulatory capital requirements), as well as risk measures in many asset-pricing models and in formulas for options pricing. Therefore, Aydemir (2002) explains that in order to adequately manage financial risks, the availability of reliable volatility estimates and forecasts are of crucial importance. With this motivation in mind, significant efforts have been put into volatility modeling and forecasting. However, a complication arises as realizations of return volatility are fundamentally latent, not directly observable, as opposed to the realized financial returns.

Many efforts to model volatility during periods of market stress have been mainly focused on the US capital markets, with less research focused on the Swedish capital market. Furthermore, less attention has been paid to comparing episode of market stress in light of the nature of the shock.

1.1 Volatility modeling

As a way to think about volatility forecasting, one may invert option pricing formulas to determined implied volatilities over a fixed time period. For instance, the Black-Scholes formula for options pricing is perhaps the most extensively used formula by industry practitioners, even in the scenario when the underlying assumptions of the model are violated. Despite this widespread use, Black-Scholes is not without material shortcomings, as it assumes an underlying constant volatility during the lifespan of the derivative, which is unaffected by changes in the underlying securities price levels (Ghysels, E. and Harvey,

A. C. and Renault, E., 1996).

An alternative approach is to invoke strong parametric assumptions through the GARCH or SV modeling framework. ARCH modeling was introduced in 1982 by Engle, and generalized by Bollerslev in 1986, and is central to time series modeling for heteroskedastic data. The class of stochastic volatility (SV) models, presented by Taylor in 1986, can be seen as an alternative of the deterministic volatility modeling within the GARCH framework. The SV models treat the underlying volatility as an autoregressive process. Based on the reformulation of volatility in the SV model, a crucial part of the modeling exercise is employing an appropriate estimation procedure. For the GARCH model maximum likelihood estimation (MLE) methods are primarily used for estimating the parameters in the model. For the SV model, there are other possibilities one may consider. Most commonly, a Bayesian context is provided for the estimation of the parameters which are computed with Markov chain Monte Carlo (MCMC) methods.

1.2 Previous studies

Various studies comparing the relative performance of GARCH and SV models have been conducted. Some researchers claim that there is enough evidence in support for the superiority of the SV model in terms of misspecification diagnostics, goodness of fit, and forecasting performance, see Kastner (2019) for instance. However, there are plenty examples in the literature of GARCH models giving researchers superior results in terms of the above.

Kim, Shephard, and Chib (1998) conducted the first proper comparison of the GARCH and the SV model. Three models were examined, namely the Gaussian GARCH, the t-GARCH and the SV model on the daily observations of weekday close exchange rate for the GBP/USD. Consequently. The models were compared via likelihood ratio (LR) testing, as well as through the deployment of a Bayes factor. While the LR-test gave strong evidence against the Gaussian GARCH, the statistics slightly supported the t-GARCH over the SV model. However, the authors mention that the t-GARCH is less parsimonious than the SV model, and thus argued that they fit the data equally well. The results from analyzing the Bayes factor mirrors that of the LR-test. The SV model outperformed the Gaussian GARCH, however, the t-GARCH appears to be the best fit for the data. The authors argued that the SV model suffers from the fact that they are using the simplest specification of the model, and that performance can be improved by implementing extensions which are discussed in the latter sections of their paper.

One such extension to the SV approach, namely the class of stochastic volatility with jumps (SVJ) models, was examined by Lopes and Polson (2010), who analyzed and monitored the credit crisis of the late 2000s using particle filtering (PF) methods. They performed sequential estimates of volatility for the Standard & Poor 500 (SP500), Nasdaq-100 (NDX100) and the Financial Select Sector SPDR Fund (XLF) indices during the early parts of the crisis. Consequently, the goal of their research was to compare volatility estimates from the SV, SVJ and the GARCH model with the implied market volatility as expressed by the CBOE Volatility Index (VIX) and the CBOE Nasdaq Volatility Index (VXN) calculated from options prices. A number of empirical results were found in the study. For instance, tracking volatility becomes increasingly difficult during periods of high market stress, as opposed to low-volatility periods. Furthermore, including the jump component into the SV model can result in drastic changes to the volatility estimates. Finally, the SVJ model is concluded to perform significantly better in times of high market stress in comparison to the basic SV and the GARCH(1,1) model.

A further direct comparison of the relative performance of the Gaussian GARCH(1,1) and the basic SV model can be found in Allen and McAleer (2020), where they examined the models over ten years of daily data from the The Financial Times Stock Exchange (FTSE) index. Moreover, they also put the volatility estimates in relation to a simple historical volatility model (HISVOL). The relative performance of the models was explored through an ordinary least squares (OLS) regression, as well as through quantile regression analysis. The objective of their research was to address the sparsity in the literature, as Granger and Poon (2005) had previously noted that there were an insufficient number of SV studies providing a comparison to the GARCH and HISVOL models. While both the papers by Kim, Shephard, and Chib (1998) and Lopes and Polson (2010) argue for the SV modeling framework providing relatively superior performance, the results obtained by Allen and McAleer (2020) point in the opposite direction. When the model fit was measured by the adjusted R^2 values, the GARCH model seems to outperform the SV model slightly by the OLS prediction. However, neither the GARCH or the SV model were able to outperform the simple HISVOL model.

1.3 Formulation of the research problem

The aim of this paper is to extend the research conducted by Lopes and Polson (2010) by analyzing the how the Gaussian GARCH(1,1) model and the basic SV model react to two different types of market shocks. Similar to the original paper, we investigate the credit crisis, which can be classified as an endogenous shock to the market. In addition to this, we also include the COVID-19 pandemic in the analysis, which instead can be classified as an exogenous shock. Another important extension of the original paper is the inclusion of the Swedish market as a point of comparison to the US market. The volatility estimation and goodness of fit research are performed before and during the endogenous shocks caused by the global credit crisis of the late 2000s, as well as before and during the exogenous shocks caused by the COVID-19 pandemic. We therefore seek to answer the following questions:

- Is there a difference in the model specification, goodness of fit and volatility estimation performance of the two models depending on if the market shocks originate from exogenous or endogenous events?

- How will the relative performance of the two models differ across markets (Sweden and US) and market regime ('normal' and 'stressed')?

The rest of the paper has the following structure. In the next section, descriptions of the realized volatility measure, the GARCH approach and the stochastic volatility approach are given, followed by the estimation methods employed. Then, the data used for the modeling exercise is presented, followed by the results. Thereafter, we present the conclusions of the study.

2 Volatility modeling

2.1 GARCH(p,q) model

Autoregressive conditional heteroskedasticity (ARCH) models, which follow the deterministic conditional volatility framework since volatility at time t is ultimately determined by the given previous values, were introduced by Engle (1982) with the idea of incorporating all the past error terms. The ARCH model was generalized into the GARCH model by Bollerslev (1986), to include lagged term conditional volatility. As a general idea, Allen (2020) explains that the GARCH model predicts that the best predictor of volatility is captured by the past realizations, daily log returns and the previous determinations of volatility. Since its introduction, the GARCH approach has experienced considerable application to real-word problems within financial and economic time series, with the Gaussian GARCH$(1, 1)$ model perhaps seeing the most widespread usage.

Following the original paper by Bollerslev (1986), the GARCH(p, q) model is given by

$$r_t = \sigma_t \kappa_t, \tag{2.1}$$

$$\sigma_t^2 = \omega + \sum_{i=1}^{p} \alpha_i r_{t-i}^2 + \sum_{j=1}^{q} \beta_j \sigma_{t-j}^2, \tag{2.2}$$

where r_t is a discrete-time stochastic variable describing financial returns. The model describes σ_t^2, the conditional variance at time t, as a function of the previous squared sample returns r_{t-1}, and the previous conditional variance σ_{t-1}^2. Here p denotes the order of the GARCH terms σ^2 which is captured by α, q denotes the order of the ARCH terms r^2 which is captured by β_j, and ω is some constant. Setting $p = q = 1$ gives the GARCH$(1, 1)$ model deployed in this paper. Furthermore, in order to ensure that σ_t^2 is non-negative and stationary, the following set of conditions are imposed

$$\omega > 0,\ \alpha_i \geq 0,\ \beta_j \geq 0 \text{ and } \alpha_i + \beta_j < 1.$$

The procedure of calculating the residuals from the GARCH model is relatively straight forward. The residuals are the difference between actual and predicted values in the conditional mean in equation (2.1) and can mathematically be presented as $\hat{\kappa}_t = \frac{r_t}{\sigma_t}$. Finally, $\{\kappa_t\}$ is assumed to be a set of iid standard normal random variables, and Ψ_{t-1} is an information set at time $t-1$. Therefore, the conditional distribution of r_t follows the normal distribution

$$r_t | \Psi_{t-1} \sim \mathcal{N}(0, \sigma_t^2). \tag{2.3}$$

2.2 GARCH model estimation: MLE

A common method for estimating the parameters in the GARCH model is via MLE methods. Based on the GARCH model presented in section 2.1, this method will be applied for estimating the parameter vector $\boldsymbol{\theta} = (\omega, \alpha, \beta)$.

The MLE procedure is now as follows. Given the time-series $T = \{1, 2, ..., n\}$, we can write the conditional densities as

$$f(r_1, r_2, ..., r_n; \boldsymbol{\theta}) = f(r_1; \boldsymbol{\theta}) f(r_2 | r_1; \boldsymbol{\theta}) ... f(r_n | r_1, r_2, ..., r_{n-1}; \boldsymbol{\theta}),$$

where $\boldsymbol{\theta}$ is a vector of parameters. Recall from equation (2.3) that the conditional distribution is $\mathcal{N}(0, \sigma_t^2)$, thus the likelihood function is given by

$$\mathcal{L}(\boldsymbol{\theta} | r_1, r_2, ..., r_n) = \prod_{t=1}^{n} \frac{1}{\sigma_t \sqrt{2\pi}} \exp\left(-\frac{1}{2} \left(\frac{r_t}{\sigma_t} \right)^2 \right).$$

Seeing as it is simpler to work with summations as opposed to products, hence the logarithm is applied to $\mathcal{L}(\boldsymbol{\theta} | r_1, r_2, ..., r_n)$, giving us the log-likelihood

$$L(\boldsymbol{\theta} | r_1, r_2, ..., r_n) = -\frac{n}{2} \ln(2\pi) - \frac{1}{2} \sum_{t=1}^{n} \ln(\sigma_t^2) - \frac{1}{2} \sum_{t=1}^{n} \frac{r_t^2}{\sigma_t^2}.$$

Finally, the maximum likelihood estimate $\hat{\boldsymbol{\theta}}_{ML}$ is the value of $\boldsymbol{\theta}$ that maximises the log-likelihood. The maximum likelihood estimate is thus given by

$$\hat{\boldsymbol{\theta}}_{ML} = \underset{\boldsymbol{\theta}}{\operatorname{argmax}}\, L(\boldsymbol{\theta} | r_1, r_2, ..., r_n).$$

This method for estimating the parameters was performed in the R language by using the package **fGarch**, developed by Wuertz et al. (2019). Parameter estimates from the GARCH model is presented in section 4.1 and 4.2.

2.3 Stochastic volatility (SV) model

Stochastic volatility, as described by Andersen and Benzoni (2009), refers to models in which return variation dynamics are subject to unobserved random shocks, resulting in the volatility being treated as a latent variable. The SV model was introduced in the seminal work of Taylor (2008) and offers an alternative to the GARCH framework by assuming that the conditional variance follows a stochastic process.

As data is typically observed in discrete time, a discrete model is presented as given by Kim, Shephard, and Chib (1998), for a demeaned time series over $T = \{1, ..., n\}$, equally spaced points

$$r_t = \exp(h_t/2)\varepsilon_t, \tag{2.4}$$

$$h_t = \mu + \phi(h_{t-1} - \mu) + \sigma\eta_t, \tag{2.5}$$

$$h_t \sim N(\mu, \sigma^2/(1 - \phi^2)), \tag{2.6}$$

$$\gamma(\varepsilon_t, \eta_t) = \gamma(\varepsilon_t, \varepsilon_{t-k}) = \gamma(\eta_t, \eta_{t-k}) = 0, \tag{2.7}$$

with r_t describing financial returns at time t, and the latent variable h_t being the log volatility assumed to follow a stationary AR(1) process given in equation (2.5). One can interpret μ and σ as the level of volatility and the volatility of h_t respectively. The parameter ϕ is the persistence parameter of the volatility and to ensure stationarity, $|\phi| < 1$ needs to be fulfilled. Further, ε_t and η_t are standard normal white noise sequences. Thus, the parameter vector is given by independent prior distributions and is denoted by $\boldsymbol{\theta} = (\mu, \sigma^2, \phi)$ as specified by Kastner and Frühwirth-Schnatter (2014).

Due to the presence of the latent variable h_t, how the standardized residuals are obtained is not as obvious as for the GARCH model. The general methodology is presented by Durham (2007) where *generalized residuals* are constructed by using the outputs from the estimation method. Once constructed, the estimated log returns from equation (2.4) is consequently rearranged, and together with the estimated log volatility from equation (2.5), the standardized residuals are calculated as $\hat{\varepsilon}_t = \exp(-\hat{h}_t/2)r_t$.

2.4 SV model estimation: MCMC

As opposed to the GARCH model, the likelihood of the SV model requires integration over the T-dimensional volatility vector, $p(y|\boldsymbol{\theta}) = \int p(y|\boldsymbol{\theta}, h)p(h|\boldsymbol{\theta})$ which is often mathematically intractable in practise (Jacquier and Polson, 2011). Instead, MCMC methods applied to SV model estimation, introduced by Albert and Chib (1993), are currently seen as the standard method of numerical integration used for computing the posterior densities in the Bayesian estimation framework. The estimation method for the SV model in this paper was performed with the R package **stochvol**, developed by Hosszejni and Kastner (2021).

MCMC methods are sampling techniques allowing the user, given a Bayesian setting, to characterise a posterior probability distribution by randomly sampling from said distribution. The MCMC approach, as presented by Rizzo (2019), is to construct a Markov chain of the desired distribution and to record states of the chain in order to obtain samples from said distribution. The chain is finally run long enough for it to reach convergence with the desired distribution. Van Ravenzwaaij, Cassey, and Brown (2018) explain that the strength of this method stems from the fact that it can be used to sample from a distribution by only knowing how to calculate densities for different samples. For an in depth and increasingly technical introduction to MCMC procedures can be found in Gilks (1996) or Tierney (1994), and for MCMC methods with application to financial and econometric time-series, see Johannes and Polson (2009).

The technique is applied to the SV model for estimating the parameter vector $\boldsymbol{\theta} = (\mu, \sigma^2, \phi)$, and the latent log-volatility $\boldsymbol{h} = (h_1, ..., h_n)$ in the SV model. In the SV modeling framework, $\boldsymbol{\theta}$ is assumed to be a random variable while $\boldsymbol{h}$ is assumed to be stochastic. The specific MCMC algorithm applied to the estimation problem is a Metropolis-Hastings with added components, as discussed by Kastner and Frühwirth-Schnatter (2014). The two key component of the specific algorithm is the "all without a loop" (AWOL) feature, which significantly reduces correlation of the draws, and the ancillarity-sufficiency interweaving strategy (ASIS) which exploits the fact that sampling efficiency improves substantially when considering a non-centered version of the SV model.

In order to perform the estimation via Bayesian methods, specification of a prior distribution for the parameter vector, $p(\mu, \sigma^2, \phi)$, is needed. Kastner and Frühwirth-Schnatter (2014) equip the level μ with the usual normal prior $\mu \sim \mathcal{N}(b_\mu, B_\mu)$. The prior for the persistence parameter ϕ is chosen as $(\phi + 1)/2 \sim \mathcal{B}(a_0, b_0)$, in similar fashion to Kim, Shephard, and Chib (1998). Finally, the volatility σ of the log volatility h_t is chosen as $\pm\sqrt{\sigma^2} \sim \mathcal{N}(0, B_\sigma)$. The specification of the priors garners careful attention. A common strategy for μ, the level of log-volatility, is specifying the relatively vague prior, $\mu \sim \mathcal{N}(0, 100)$. The persistence parameter is set to the default option as given in **stochvol**, $(\phi + 1)/2 \sim \mathcal{B}(5, 1.15)$. The default option in the package is also applied to σ, setting the prior to $\pm\sqrt{\sigma^2} \sim \mathcal{N}(0, 1)$.

After having specified the prior distribution, we run the MCMC sampler, and the joint posterior distribution is thereby obtained. The parameter estimations from the SV model is presented in section 4.1 and 4.2.

2.5 Volatility estimation

Measuring a models ability to produce accurate estimations of volatility is fundamental when conducting inference on model performance. In order to validate a models ability to provide accurate estimates a proxy for volatility is required. A variety of proxies to estimate the true volatility have been presented. Following Hansen and Lunde (2006), an estimate $\hat{\sigma}_t^2$ of the volatility at day t can be obtained by sampling the log-price process at several occasions during the day t and from these several observations evaluate the intra-day variability of returns. For theoretical motivations, examples of this technique can be found in Barndorff-Nielsen and Shephard (2002), and Andersen, Bollerslev, et al. (2003).

Poon and Granger (2003) explain that the accuracy of the estimates can consequently be assessed through mean squared error (MSE), root mean squared error (RMSE), and mean absolute error (MAE), commonly used in financial time series. In this paper we use these measures as a goodness of fit and estimating ability assessment. Let σ_t^2 denote the volatility estimates provided by the GARCH and SV model so that the estimation errors are given by

$$MSE = \frac{1}{n}\sum_{t=1}^{n}(\sigma_t^2 - \hat{\sigma}_t^2)^2,$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{t=1}^{n}(\sigma_t^2 - \hat{\sigma}_t^2)^2},$$

$$MAE = \frac{1}{n}\sum_{t=1}^{n}|\sigma_t^2 - \hat{\sigma}_t^2|.$$

3 A study of volatility in the OMSX30 and SP500 data

3.1 Data

The data sets consist of daily closing prices from the OMXS30, a Swedish stock market index, and Standard Poor 500 (SP500), a US stock market index, from 2002-01-01 to 2022-11-01. Bloomberg was used as the data source for every data set studied in this thesis. Both data sets span from 2002-01-01 to 2022-11-01 and to perform the parameter estimation of the GARCH and SV model, the data is demeaned utilizing the demean function provided in is the daily log returns are calculated on the raw data as

$$r_t = 100\left[\ln(P_t) - \ln(P_{t-1})\right], \tag{3.1}$$

where P_t denoted the returns at time t. Following the methodology presented by Lopes and Polson (2010), the sample period is divided into four sub-periods. Two four-year periods leading up to each crisis, and two one year periods during the first year of each crises. The two four-year periods were used to evaluate performance of the models in periods of relativity low volatility, and subsequently the following one year periods for each stress episode, highlighted in Figure 3.1 and Figure 3.2. The green is indicating the four-year periods consisting of 1225 observations covering the period 2002-01-02 to 2006-12-29, and 978 observations covering the period 2016-01-12 to 2019-12-29 respectively for both indices. The red instead indicates the one-year periods consisting of 243 observations covering the period 2007-01-02 to 2007-12-29, and 245 observations covering the period 2020-01-02 to 2020-12-29, respectively for both indices.

3.1.1 SP500

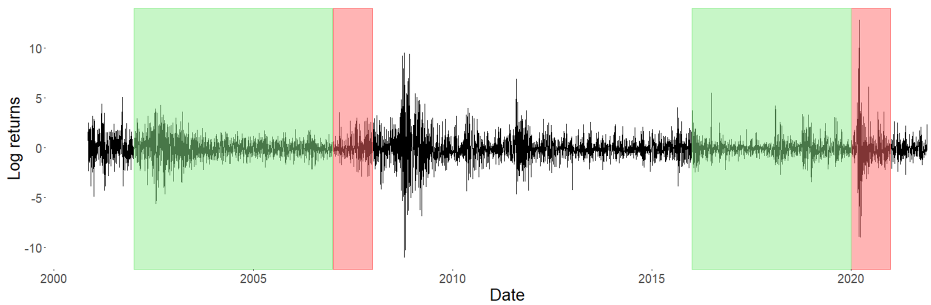


Figure 3.1: Daily logarithmic rate of returns used for the sample statistics for the SP500 from 2002-01-01 to 2022-11-01 with the green areas indicating relatively stable market conditions and the red areas market stress periods.

The SP500 is a US stock market index consisting of 500 large-cap US equities, weighted by float-adjusted market capitalization, and is considered a proxy for the US stock market. The index will be weighted on specific dates but the rebalancing schedule can be changed at the discretion of Standard & Poor, the index provider. Summary statistics of the logarithmic daily return data is presented in Table 3.1 and graphics of the logarithmic daily return data can be found in Figure 3.1 An interesting feature of the plots is that volatility clustering is clearly present in the data.

Sample size	Min	Max	Mean	St.dev	Skewness	Kurtosis
5296	-0.109	0.128	0.000	0.007	0.418	11.635

Table 3.1: Summary statistics for the daily log returns of the SP500 index from 2002-01-01 to 2022-11-01.

3.1.2 OMXS30

The OMXS30 is a Swedish stock market index consisting of the 30 most traded stocks, weighted by each stocks market capitalization, and is rebalanced biannually. Summary statistics of the logarithmic daily return data is presented in Table 3.2 and a plot of the logarithmic daily return data can be found in Figure 3.2, the highlighted periods are the same as for the US data in Figure 3.1. For the Swedish data, volatility clustering is clearly present as well.

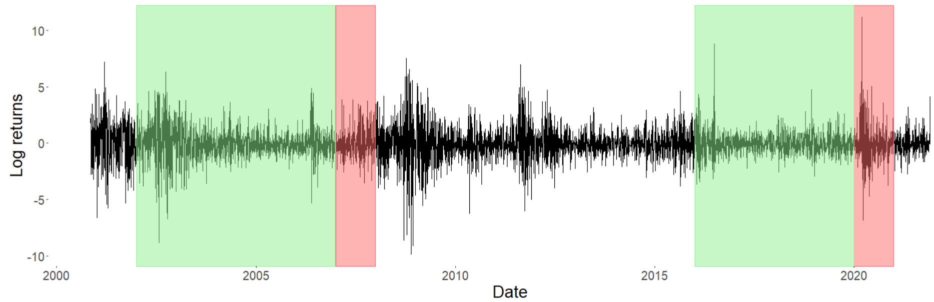


Figure 3.2: Daily logarithmic rate of returns used for the sample statistics for the OMXS30 data from 2002-01-01 to 2022-11-01 with the green areas indicating relatively stable market conditions and the red areas market stress periods.

Sample size	Min	Max	Mean	St.dev	Skewness	Kurtosis
5287	-0.099	0.112	0.000	0.014	0.102	4.761

Table 3.2: Summary statistics for the daily log returns of the OMXS30 index from 2002-01-01 to 2022-11-01.

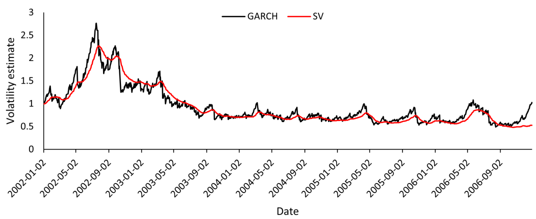
3.2 Results

This section presents the results from the modeling procedure, covering parameter estimates, residual analysis and volatility estimation errors. When comparing the misspecification diagnostics of the models during different sub-samples, indices and amount of market stress, we are mainly looking at if the models are in line with the previously specified conditions, that is stationarity and normal white noise residuals. The parameter estimates are used to check if the models are stationary, the Ljung-Box and ARCH LM test are used to investigate if the residuals are white noise, and the QQ-plots are used to investigate if the errors are normal. Finally, in addition to model fit, the estimation performance of the models are also an important point of comparison for how good the models are in relation to one another for the different sub-samples. To test this ability, we use the MSE, RMSE and MAE, error metrics where smaller errors indicate a better goodness of fit and volatility estimation ability.

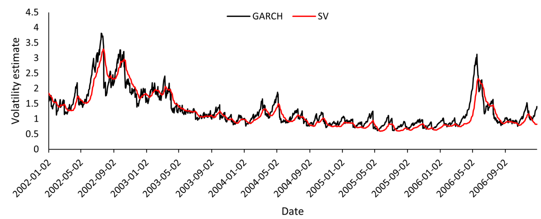
3.2.1 Volatility estimates for normal market regimes

Volatility estimates from the SV and GARCH models for the periods of relatively low market stress are presented in Figure 3.3. These sub-samples reflect relatively stable market conditions, and in all four series the two models appear to follow each other closely.

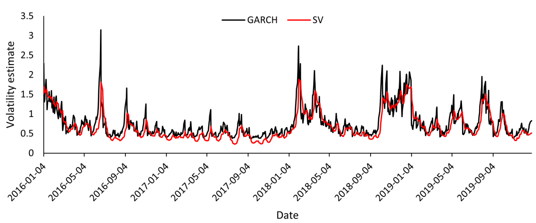
However, the volatility estimates from the GARCH model are on average higher than that of the SV model for all four series. This is particularly evident in both the SP500 and OMXS30 series leading up to the pandemic, where the estimates produced by the GARCH model are almost twice as high at certain points in time. Moreover, the SV model seems to be leading the GARCH model. It is therefore likely that the SV model adjusts to changes in volatility at an earlier point and needs less estimations to expand the estimates, which is seen to be more prevalent for the sub-samples leading up to the pandemic. It is thus likely that the volatility estimates adjust faster for the SV model.



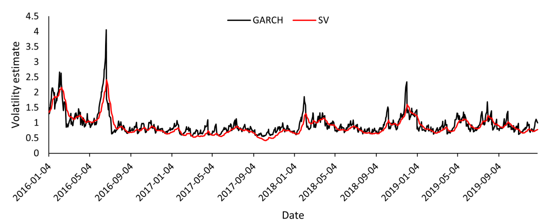
(a) SP500, before credit crisis



(b) OMXS30, before credit crisis



(c) SP500, before pandemic



(d) OMXS30, before pandemic

Figure 3.3: Estimated volatilities derived from σ_t^2 and $exp(h_t)$ for the GARCH and SV model respectively when applied on the SP500 and OMXS30 data sets. The sub-samples consists of 1224 observations over the period 2002-01-02 to 2006-12-29, capturing the period leading up to the credit crisis. Additionally, the sub-samples also consists of 977 observation over the period 2016-01-02 to 2019-12-29, covering the period leading up to the pandemic.

3.2.2 Parameter estimates for normal market regimes

SP500 data

Table 3.1 presents the parameter estimates for the models applied to the SP500 data set for the sub-samples with relatively low market volatility. For the SV model, we find that the persistence parameter ϕ is able to satisfy its constraint for both sub-samples on the US data. The model is therefore concluded to be stationary for both sub-samples leading up to market shocks. Recall from section 3.3 that the restriction imposed is $|\phi| < 1$.

The parameter constraints for the GARCH model, as given in section 3.1, are also met for both sub-samples on the US data. The constant ω, the ARCH-effect α_1 and the GARCH-effect β_1 are all greater than zero. Meanwhile, $\alpha_1+\beta_1 = 0.281+0.693 = 0.974 < 1$ before the credit crisis, and $\alpha_1 + \beta_1 = 0.067 + 0.924 = 0.991 < 1$ before the pandemic. This confirms that the GARCH model is stationary for both sub-samples. Moreover, the ARCH-effect α_1 measures to which extent past residuals effect the current volatility, while the GARCH-effect β_1 instead measures the effect past volatility has on current volatility. We can see that the GARCH-effect is greater than the ARCH-effect during both sub-samples, implying that past volatility has the largest influence on the current state of volatility in the GARCH model.

		SP500, before credit crisis		SP500, before pandemic	
Model	Parameter	Mean	St.dev	Mean	St.dev
SV	μ	-0.461	0.762	-0.985	0.203
	ϕ	0.995	0.003	0.932	0.019
	σ^2	0.007	0.003	0.167	0.041
GARCH	ω	0.036	0.003	0.007	0.008
	α_1	0.281	0.012	0.067	0.041
	β_1	0.693	0.013	0.924	0.036

Table 3.3: Parameter estimates for the two models applied to the SP500 data set for the periods leading up to market stress. The first estimates are based on 1224 observations over the period 2002-01-02 to 2006-12-29, covering the period leading up to the credit crisis of the late 2000s. Additionally, the second estimates are based on 977 observations over the period 2016-01-02 to 2019-12-29, covering the periods leading up to the COVID-19 pandemic.

OMXS30 data

Table 3.4 presents the parameter estimates for the models applied to the OMXS30 data set for the sub-samples leading up to the market shocks. For the SV model, we again find that the persistence parameter is able to satisfy its constraints for both sub-samples, that is $|\phi| < 1$. The SV model is thus also concluded to be stationary when applying the Swedish data.

For the GARCH model we see that the parameter constraints are met for both sub-samples on the Swedish data as well. Constant ω, the ARCH-effect α_1 and GARCH-effect β_1 are all greater than zero which is in line with the constraints. Further, $\alpha_1 + \beta_1 = 0.102 + 0.888 = 0.990 < 1$ for the sub-samples leading up to the credit crisis, and $\alpha_1 + \beta_1 = 0.177 + 0.766 = 0.943 < 1$ for the the sub-samples leading up to the pandemic. This implies that the model is stationary during both periods. Furthermore, in line with the results from the US data, the GARCH effect has a larger impact on the current volatility than the ARCH effect for the Swedish data.

		OMXS30, before credit crisis		OMXS30, before pandemic	
Model	Parameter	Mean	St.dev	Mean	St.dev
SV	μ	0.253	0.421	-0.305	0.197
	ϕ	0.985	0.007	0.959	0.016
	σ^2	0.025	0.009	0.043	0.015
GARCH	ω	0.023	0.009	0.058	0.020
	α_1	0.102	0.021	0.177	0.035
	β_1	0.888	0.021	0.766	0.046

Table 3.4: Parameter estimates for the two models applied to the OMXS30 data set before the periods of high market stress. The first estimates are based on 1224 observations over the period 2002-01-02 to 2006-12-29, covering the period leading up to the credit crisis of the late 2000s. Additionally, the second estimates are based on 977 observations over the period 2016-01-02 to 2019-12-29, covering the periods leading up to the COVID-19 pandemic.

3.2.3 Residuals for normal market regimes

SP500 data

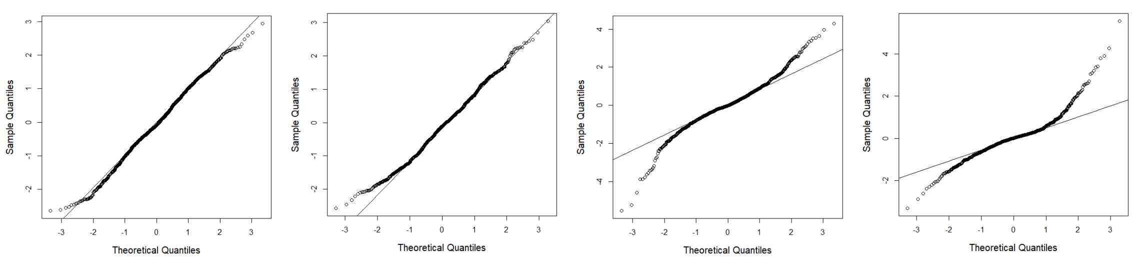
In Table 3.5 the Ljung-Box and ARCH LM tests for the sub-samples of relatively low market volatility in the SP500 data set are presented in order to test if serial correlation or ARCH-effect is present in the residuals. While both models exhibit an ARCH-effect during the periods leading up to the pandemic, no such effect is present for the SV model for the sub-samples leading up to the credit crisis. Therefore, the SV model may be a better fit when using the US data considering that an ARCH-effect is present in the residuals for all data sets for GARCH model. This implies that the GARCH model exhibits autocorrelation, which violates the condition that the residuals are white noise. The GARCH model is thus not a correct model for the data. Meanwhile, the residuals from the US data before the pandemic indicate the ARCH effect for the SV model, implying that this model may not be appropriate for the specific data either.

The subfigures of Figure 3.4 give the QQ-plots of the standardized residuals for the two models during the low volatility periods on the SP500 data set. The standardized residuals for the SV model are similar for both sub-samples and follow the line to high degree, consequently making them appear to be Gaussian. Therefore, it may be concluded that the errors are normal white noise, and in line with the specified assumptions. We may however be sceptical towards the period before the pandemic, as the SV model was unable

Model	Data set	Ljung-Box$_{10}$	ARCH LM
SV	SP500, before credit crisis	10.013 (0.439)	16.181 (0.09457)
SV	SP500, before pandemic	8.304 (0.599)	59.948 (3.707e−09)***
GARCH	SP500, before credit crisis	16.178 (0.094)	329.13 (2.2e−16)***
GARCH	SP500, before pandemic	17.951 (0.055)	78.534 (9.722e−13)***

Table 3.5: Results of the the Ljung-Box test using up to 10 lags and the ARCH LM test for the standardized residuals of the two models during the two periods of relatively low volatility applied on the SP500 data set. Test statistics are presented with their respective p-values in the parentheses.

to pass the ARCH LM test. For the GARCH model however, is immediately evident that the standardized residuals of the model deviates substantially from the line during both periods. From the tail deviations, it can be seen that the errors exhibit heavy tails. In turn, this implies that the model does not seem not pass misspecification tests. With this in mind, it is not likely that the GARCH model is appropriate for explaining the data and needs to be extended in order to perform well. One such extension would be the inclusion of a jump component, as implemented by Lopes and Polson (2010) for the SV model. This extension could possibly allow the model to capture the outliers and heavy tail behavior.



(a) SV, before credit crisis

(b) SV, before pandemic

(d) GARCH, before credit crisis

(b) GARCH, before pandemic

Figure 3.4: QQ-plots of the standardized residuals, which are estimated by $\hat{\kappa}_t = r_t/\sigma_t$ for the GARCH and $\varepsilon_t = exp(-h_t/2)r_t$ for the SV model, during the high volatility period on the SP500 data set. Theoretical quantiles are plotted against the x-axis, and the sampled quantiles are plotted against the y-axis.

OMXS30 data

The corresponding Ljung-Box and ARCH LM tests for the OMXS30 data set are presented in Table 3.6. The SV model seems to be more strongly favored during the periods with low volatility for the Swedish data, compared to the US data, as it passes both tests for both sub-samples. Meanwhile, the GARCH model does not manage to pass the Ljung-Box test for up to ten lags during the period leading up to the pandemic, implying serial correlation is present in the residuals. Additionally, the GARCH model does not pass the ARCH LM tests for either sub-sample pointing towards GARCH model is not suitable for the Swedish data.

Model	Data set	Ljung-Box$_{10}$	ARCH LM
SV	OMXS30, before credit crisis	3.306 (0.973)	17.339 (0.06719)
SV	OMXS30, before pandemic	7.694 (0.658)	23.161 (0.01017)***
GARCH	OMXS30, before credit crisis	12.63 (0.245)	203.44 ($2.2e{-16}$)***
GARCH	OMXS30, before pandemic	28.282 (0.001)***	57.695 ($9.866e{-09}$)***

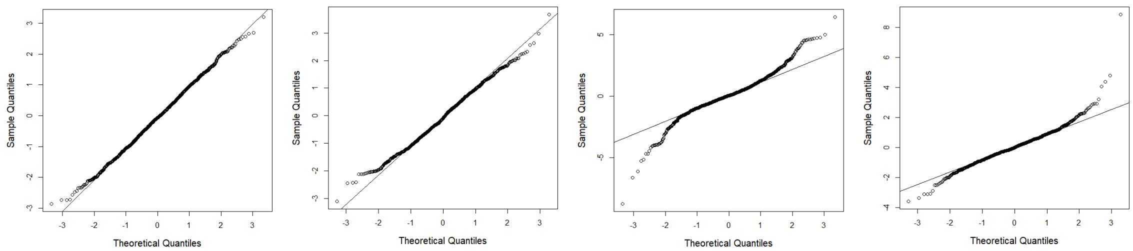
Table 3.6: Results of the the Ljung-Box test using up to 10 lags and the ARCH LM test for the standardized residuals of the two models during the two periods of relatively low volatility applied on the OMXS30 data set. Test statistics are presented with their respective p-values in the parentheses.

The QQ-plots of the standardized residuals for the two models during the low volatility periods applied to the Swedish data is presented in the subfigures of Figure 3.5. The QQ-plots appear to behave in similar fashion as when applied to the US data. The SV model seems to operate within the specified assumptions of the model, that is Gaussian errors. Meanwhile, the GARCH model is again moving outside of this assumption. Therefore, the SV model seems to fit the Swedish data well while the GARCH model again may need to be extended in order to pass the misspecification tests. To conclude, while the fit of the SV model is better for the Swedish sub-sample used, the fit of the GARCH model is better for the US sub-sample. However, the fit of the SV model appears to be better than the GARCH model for both sub-samples.

3.2.4 Model assessment during normal market regimes

SP500 data

The goodness of fit metrics MSE, RMSE and MAE for the two models and series during stable market periods applied to the SP500 data set are presented in Table 3.7. Observing both models, the series leading up to the COVID-19 pandemic yields the best goodness of fit according to the error metrics, pointing towards this series being easiest to estimate. Contrasting the two models, we observe that the assessment for the SV model is more accurate for the sub-sample leading up to the pandemic, while the estimated values from the GARCH model are more accurate for the sub-sample leading up to the credit crisis.



(a) SV, before credit crisis **(b) SV, before pandemic** **(c) GARCH, before credit crisis** **(d) GARCH, before pandemic**

Figure 3.5: QQ-plots of the standardized residuals, which are estimated by $\hat{\kappa}_t = r_t/\sigma_t$ for the GARCH and $\varepsilon_t = exp(-h_t/2)r_t$ for the SV model, during the high volatility period on the SP500 data set. Theoretical quantiles are plotted against the x-axis, and the sampled quantiles are plotted against the y-axis.

Model	Data set	MSE	RMSE	MAE
SV	SP500, before credit crisis	0.866	0.931	0.600
SV	SP500, before pandemic	**0.268+**	**0.517+**	**0.347+**
GARCH	SP500, before credit crisis	0.437+	0.661+	0.517+
GARCH	SP500, before pandemic	**0.397**	**0.630**	**0.471**

Table 3.7: The MSE, RMSE and MAE metrics for the two periods of relatively low market stress applied to the SP500 data set. Emboldening highlights the smallest estimation errors when comparing the performance of the two data sets for the different models alone. The plus sign (+) highlights the smallest estimation errors when comparing the performance of the two models for the two different data sets.

Moreover, based on the fit of the models where we found the GARCH model to be misspecified and moving outside of its assumptions, the volatility estimation performance of the GARCH model is surprising. While this can be interpreted as a definitive results showcasing the superiority of the GARCH model, it is likely that this conclusion is misleading as the GARCH model produces volatility estimates that consistently will be closer to the observed market volatility relative to the SV model. As seen previously, the SV model is quicker to react than the GARCH model. A possibly more reasonable interpretation of the results above is that the SV models provides better insight into in which direction the volatility is heading and the GARCH model provides better estimates of observed values.

OMXS30 data

The volatility estimation metrics MSE, RMSE and MAE for the two models and series during stable market periods applied to the OMXS30 data set are presented in Table 3.8. Similarly to the US data, the period leading up to the pandemic yields the lowest estimation errors. A striking difference from the US data however, is that the GARCH model provides superior volatility estimation errors for both sub-samples.

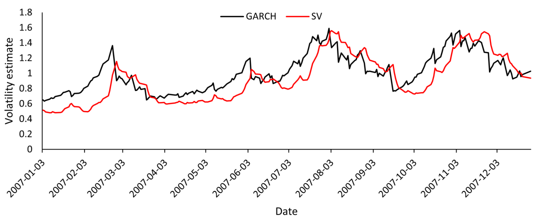
Model	Data set	MSE	RMSE	MAE
SV	OMXS30, before credit crisis	3.84	1.96	1.42
SV	OMXS30, before pandemic	**0.475**	**0.689**	**0.465** +
GARCH	OMXS30, before credit crisis	0.917 +	0.958+	0.730 +
GARCH	OMXS30, before pandemic	**0.464+**	**0.681+**	**0.514**

Table 3.8: The MSE, RMSE and MAE metrics before for the two periods of relatively low market stress applied to the OMXS30 data set. Emboldening highlights the smallest estimation errors when comparing the performance of the two data sets for the different models alone. The plus sign (+) highlights the smallest estimation errors when comparing the performance of the two models for the two different data sets.

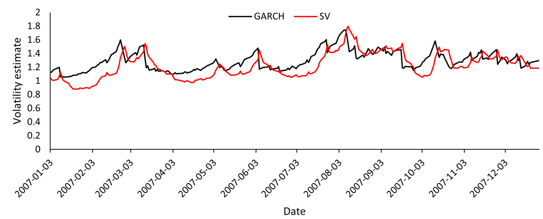
Additionally, comparing the results from the two indices, the models perform better on the US data. This is expected, considering the different structures of the two indices. As the OMXS30 is a market weighted price index that is re-balanced twice and consists of the 30 most actively traded stocks on the Stockholm stock exchange it will be inherently more volatile than the SP500 index, which can be seen by comparing Figure 3.2 and Table 3.2 with Figure 3.1 and Table 3.1, where kurtosis is of particular interest. Furthermore, for stocks to be included in the SP500 the underlying companies are required to meet certain criteria, for example financial stability and market capitalization minimum among others. Meanwhile, the OMXS30 series before the credit crisis yields the highest errors, similarly implying that this series is the most difficult for volatility estimation.

3.2.5 Volatility estimates for stressed market regimes

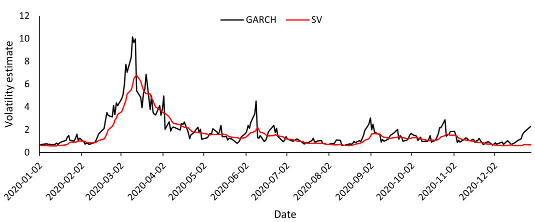
During the periods characterized by market stress, some notable differences between the estimates of the two models are observed in Figure 3.6. For one, the volatility estimates of the two models during these periods no longer seem to follow each other as closely compared to the results during periods of relatively low volatility. In other words, the difference between the models become more obvious as the data is characterised by market stress. One such difference is that the GARCH model no longer exclusively estimates higher volatility than the SV model. This is observed during the credit crisis, series (a) and series (b), where the volatility estimates from the SV model on average are the highest. In direct contrast, for the pandemic, series (c) and (d), the estimates tend to be higher on average for the GARCH model. Furthermore, there is a larger difference in how much the SV model leads during periods of market stress. This is predominantly observed during the credit crisis where the SV model appears to react significantly faster than the GARCH model to changes in the levels of volatility. While this is still observable during the pandemic, the difference is not as prevalent.



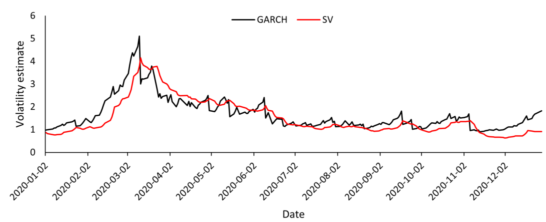
(a) SP500, during credit crisis



(b) OMXS30, during credit crisis



(c) SP500, during pandemic



(d) OMXS30, during pandemic

Figure 3.6: Estimated volatilities derived from σ_t^2 and $exp(h_t)$ for the GARCH and SV model respectivily when applied on the SP500 and OMXS30 data sets. The sub-samples consist of 243 observations over the period 2007-01-02 to 2007-12-29, capturing the credit crisis. Additionally, the sub-samples consist of 245 observation over the period 2020-01-02 to 2020-12-29, capturing the COVID-19 pandemic.

3.2.6 Parameter estimates for stressed market regimes

SP500 data

Parameter estimates for the models during the periods of high market volatility applied to the SP500 data set are presented in Table 3.9. The constraint specified for the persistence parameter ϕ in SV model is met during both periods, seeing as $|\phi| < 1$. Stationarity for the SV model is therefore confirmed.

However, a clear difference from the other presented parameter estimates can be observed in the GARCH model, where stationarity is not met for both sub-samples. The

Model	Parameter	SP500, during credit crisis Mean	St.dev	SP500, during pandemic Mean	St.dev
SV	μ	-0.304	0.458	0.212	1.062
	ϕ	0.921	0.058	0.961	0.021
	σ^2	0.119	0.089	0.145	0.052
GARCH	ω	0.029	0.023	0.134	0.068
	α_1	0.086	0.034	0.501	0.124
	β_1	0.884	0.042	0.552	0.082

Table 3.9: Parameter estimates for the two models applied to the SP500 data set for the periods of high market stress. The first sub-samples consist of 243 observations over the period 2007-01-02 to 2007-12-29, capturing the credit crisis. Additionally, the second sub-samples also consist of 245 observation over the period 2020-01-02 to 2020-12-29, capturing the COVID-19 pandemic.

imposed constraints of ω, α_1 and β_1 being greater than zero are met for both sub-samples. However, while $\alpha_1 + \beta_1 = 0.086 + 0.884 = 0.970 < 1$ during the credit crisis, $\alpha_1 + \beta_1 = 0.501 + 0.552 = 1.053 > 1$ during the pandemic. This implies that the model is not stationary during the pandemic when fitting the model to the US data, which is a problematic consequence of the pandemic. This will theoretically lead to the GARCH model producing unreliable and spurious results as well as poor understanding and estimates. Futhermore, seeing as β_1 is greater than α_1, the GARCH-effect seems to be greater than the ARCH effect during the pandemic, in line with the previous parameter estimates.

OMXS30 data

Table 3.10 presents the parameter estimates for the two models during the periods of high market stress applied on the OMXS30 data set. The parameter constraint for the SV model is again met during both periods considering that $|\phi| < 1$. The GARCH model also meets the imposed parameter constraints needed to ensure stationarity for both sub-samples. The constant ω, α_1 and β_1 are all greater than zero. Further, $\alpha_1 + \beta_1 = 0.059 + 0.856 = 0.915 < 1$ during the credit crisis, and $\alpha_1 + \beta_1 = 0.147 + 0.816 = 0.963 < 1$ during the pandemic. Seeing as β_1 is greater than α_1 for both indices, the GARCH effect has a larger impact on the current volatility than the ARCH effect, for both sub-samples on the Swedish data.

3.2.7 Residuals for stressed market regimes

SP500 data

The Ljung-Box and ARCH LM tests for up to ten lags during periods of high volatility applied to the SP500 data set are presented in Table 3.11. Unsurprisingly, the results imply

Model	Parameter	OMXS30, during credit crisis Mean	St.dev	OMXS30, during pandemic Mean	St.dev
SV	μ	0.376	0.196	0.443	0.763
	ϕ	0.761	0.224	0.967	0.028
	σ^2	0.126	0.094	0.105	0.055
GARCH	ω	0.140	0.122	0.120	0.048
	α_1	0.059	0.041	0.147	0.041
	β_1	0.856	0.093	0.816	0.040

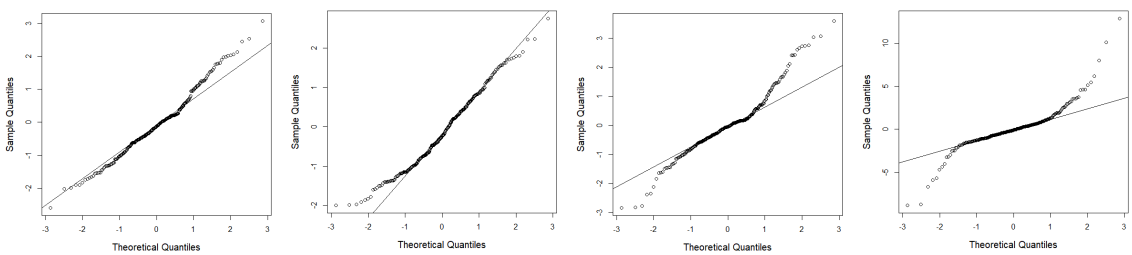
Table 3.10: Parameter estimates for the two models applied to the OMXS30 data set for the periods of high market stress. The first sub-samples consist of 243 observations over the period 2007-01-02 to 2007-12-29, capturing the credit crisis. Additionally, the second sub-samples consist of 245 observation over the period 2020-01-02 to 2020-12-29, capturing the COVID-19 pandemic.

that the ARCH-effect and serial correlation in the residuals is more prevalent for the SV model for the data characterised by market stress as opposed to during the stable market conditions. In other words, the results suggest that the SV model has a higher tendency for misspecification as market stress is prevalent in the data. Therefore, the specification of the models can be questioned for both indices during the pandemic, as they either do not manage to pass the Ljung-Box test or the ARCH LM test. The residuals of the GARCH model also become worse during market stress, seeing as the model fails the Ljung-Box test during the pandemic. This suggests that both models have problems with fitting the data correctly during periods of high volatility on the US sub-samples, and the assumption of white noise errors is highly questioned.

Model	Data set	Ljung-Box$_{10}$	ARCH LM
SV	SP500, during credit crisis	14.552 (0.149)	23.097 (0.010)**
SV	SP500, during pandemic	24.07 (0.007)**	17.713 (0.060)
GARCH	SP500, during credit crisis	17.287 (0.068)	29.339 (0.001)***
GARCH	SP500, during pandemic	146.82 ($< 2.2e{-}16$)***	99.078 ($< 2.2e{-}16$)***

Table 3.11: Results of the the Ljung-Box test using up to 10 lags and the ARCH LM test for the standardized residuals of the two different models for the applied to the SP500 data set. Test statistics are presented with their respective p-values in the parentheses.

QQ-plots for the models during the periods of high market stress is presented in the subfigures of Figure 3.7. It is clear that the standardized residuals deviate from the QQ-line to a larger extent during market stress. This is also in line with the results when testing for white noise errors. For the endogenous credit crisis, the SV model appears to be more heavily tailed towards the right. In direct contrast, during the exogenous pandemic, the model instead appears to be more heavily skewed towards the left. This appears to speak for a jump component possibly being a viable extension to the model during endogenous events. However, the extent of the deviation is in no way drastic and the assumption of normal errors is seemingly still met. In contrast to the QQ-plots of the SV model which communicate normal errors, the GARCH model still tends to produce heavy tails. This is in similar fashion to the results from the QQ-plots during periods of low market volatility. Therefore, the GARCH model proves to show a degree of misspecification for the US data used in this study.



(a) SV, during credit crisis **(a) SV, during pandemic** **(a) GARCH, during credit crisis** **(a) GARCH, during pandemic**

Figure 3.7: QQ-plots of the standardized residuals, which are estimated by $\hat{\kappa}_t = r_t/\sigma_t$ for the GARCH and $\varepsilon_t = exp(-h_t/2)r_t$ for the SV model, during the high volatility period on the SP500 data set. Theoretical quantiles are plotted against the x-axis, and the sampled quantiles are plotted against the y-axis.

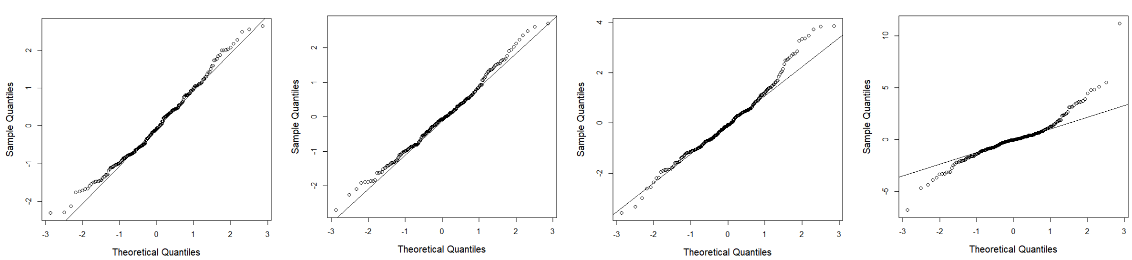
OMXS30 data

The respective Ljung-Box and ARCH LM tests for the OMXS30 data set is presented in Table 3.12. The residuals from the SV model passes both tests during the credit crisis, confirming that the errors are white noise. This may however not be necessarily true during the pandemic, as the model is not able to pass the test for ARCH-effect and therefore shows a degree of misspecification. Therefore, the SV model should be treated with skepticism during the endogenous events on the Swedish data. Unexpectedly however, the errors from the GARCH model perform slightly better when market stress is a characteristic in the data, compared to the results obtained during stable markets conditions. In the results for the previous section of low volatility periods, all the sub-samples failed the ARCH LM test. During the credit crisis on the Swedish data however, contradictory to ones intuition, the GARCH model passes both of the tests, which indicates that the GARCH model should be correctly specified for this sub-sample.

Model	Data set	Ljung-Box$_{10}$	ARCH LM
SV	OMXS30, during credit crisis	6.819 (0.742)	5.168 (0.879)
SV	OMXS30, during pandemic	10.564 (0.392)	23.373 (0.009)**
GARCH	OMXS30, during credit crisis	6.1758 (0.800)	6.7353 (0.750)
GARCH	OMXS30, during pandemic	12.335 (0.263)	39.813 (1.828e−05)***

Table 3.12: Results of the the Ljung-Box test using up to 10 lags and the ARCH LM test for the standardized residuals of the two different models applied to the OMXS30 data set. Test statistics are presented with their respective p-values in the parentheses.

The respective QQ-plots for the models are presented in the subfigures of Figure 3.8. For the SV model, it is still clear that the errors appear to be normal and therefore meet the assumption of normal white noise errors. Moreover, observe that for the GARCH model, the QQ-plot for the credit crisis also seems to fit this assumption quite well. This is in line with the results when testing for white noise, where the GARCH model passed both the Ljung-Box and the ARCH LM test. Therefore, one may conclude that the GARCH model fits the data well on the Swedish data during the credit crisis but not for the other sub-samples.



(a) SV, during credit crisis **(a) SV, during pandemic** **(a) GARCH, during credit crisis** **(a) GARCH, during pandemic**

Figure 3.8: QQ-plots of the standardized residuals, which are estimated by $\hat{\kappa}_t = r_t/\sigma_t$ for the GARCH and $\varepsilon_t = exp(-h_t/2)r_t$ for the SV model, during the high volatility period on the SP500 data set. Theoretical quantiles are plotted against the x-axis, and the sampled quantiles are plotted against the y-axis.

3.2.8 Model assessment during stressed market regimes

SP500 data

The volatility estimation errors for the periods of market stress applied to the SP500 data set are presented in Table 3.13. A difference compared to the periods of stable markets is that the sub-samples during the pandemic have the highest errors for both models, whereas the opposite was true before the data was characterised by market stress. We also see that both models are able to produce the best estimates during the credit crisis.

Model	Data set	MSE	RMSE	MAE
SV	SP500, during credit crisis	**0.476+**	**0.690+**	**0.533+**
SV	SP500, during pandemic	51.0	7.14	2.89
GARCH	SP500, during credit crisis	**0.571**	**0.756**	**0.627**
GARCH	SP500, during pandemic	2.14+	1.46+	0.980+

Table 3.13: The MSE, RMSE and MAE error metrics during the periods of market stress applied to the SP500 sub-sample. Emboldening highlights the smallest estimation errors when comparing the performance of the two data sets for the different models alone. The plus sign (+) highlights the smallest estimation errors and the best goodness of fit when comparing the performance of the two models for the two different data sets.

During the periods before market stress, the SV model outperformed the GARCH model during the period leading up to the pandemic. During the periods of market stress, the SV model now outperforms the GARCH model during the credit crisis. This implied that in the context of the SP500, the SV model may be better suited for volatility estimation during endogenous shocks while the GARCH model may be superior for estimation volatility during exogenous events.

OMXS30 data

The volatility estimation errors for the periods of market stress applied to the OMXS30 data set are presented in Table 3.14. While the GARCH model still shows a greater goodness of fit than the SV model overall in the Swedish sub-samples, the SV model proves superior for an increasing number of estimation errors as opposed to before the shocks. This shows that the relative performance of the SV model in relation the GARCH model increases with increasing market stress for the Swedish sub-samples. Similar to the US sub-samples, the SV model tends to produce better estimation errors during the credit crisis while the GARCH model instead produces the best estimates during the pandemic. This strengthens the conclusion of the SV model being the appropriate choice during endogenous events,

Model	Data set	MSE	RMSE	MAE
SV	OMXS30, during credit crisis	**0.724+**	**0.851+**	**0.724**
SV	OMXS30, during pandemic	10.3	3.21	1.92
GARCH	OMXS30, during credit crisis	**0.751**	**0.867**	**0.720+**
GARCH	OMXS30, during pandemic	1.79+	1.34+	1.03+

Table 3.14: The MSE, RMSE and MAE error metrics during the periods of market stress applied to the OMXS30 data set. Bold numbers represent the smallest, and thus best, volatility estimation errors. Emboldening highlights the smallest estimation errors when comparing the performance of the two data sets for the different models alone. The plus sign (+) highlights the smallest estimation errors when comparing the performance of the two models for the two different data sets.

while the GARCH model is the best choice during exogenous events given the data we are working with.

4 Conclusion

The aim of this paper has been to expand on the research of Lopes and Polson (2010) by assessing the SV and GARCH models before and during two different periods of market stress, one the result of an endogenous shock, the other induced by an exogenous shock on for both the SP500 and the OMXS30. For the discrete-time SV model we considered, we modelled the volatility as a latent variable following a stochastic process, specifically we chose an autoregressive process of order one as this is a common choice in the literature. This contrasts the GARCH model, where volatility is instead modelled with the deterministic conditional volatility framework. Moreover, another difference in the models is how the estimation methods were performed. For the SV model, considered a Bayesian estimation framework and deployed an MCMC sampler in order to estimate the parameters, and is generally more complicated as opposed to the MLE for the GARCH model. The specific algorithm we used for the estimation problem of the SV model was a variant of Metropolis-Hastings with some added components, provided by the **stochvol** package in R. This is yet another key difference to the paper by Lopes and Polson (2010), as they used PF methods, as opposed to the specific MCMC algorithm.

The empirical results shows that the SV model appears to fit the specifications for the data better than the GARCH model for all data sets that were tested. This is predominantly evident in the QQ-plots where the SV model follows the QQ-line more closely than the GARCH model, indicating that the SV model operates within the specified assumptions, that is normal errors, while the GARCH model does not. Moreover, the QQ-plot for the SV model also did experience much change after market stress was a characteristic in the data, while the QQ-plot for the GARCH model still did not provide acceptable results for interpreting the errors as normal. Moreover, we could also see that the GARCH model during the pandemic for the SP500 data set did not fulfill stationarity, providing evidence for the GARCH model not being appropriate for the usage during endogenous shocks on the US markets.

In terms of the volatility estimates produced by the two models, and interesting observation was the fact that the SV model leads the GARCH model for all sub-samples examined in the thesis. This implies that the SV model is able to adjust to changing regimes with higher speed, which is of high importance, especially in volatile market regimes.

It was also concluded that an ARCH-effect was on high occasion present in both models with the GARCH model suffered to a higher extent. However, the ARCH effect unexpectedly lessened for the Swedish sub-sample as we examined the stress event, as the OMXS30 during the credit crisis was the only data set where the GARCH model had an acceptable fit. In regards to serial correlation, it was clear that the exogenous shock from the pandemic was the series with the highest serial correlation in the residuals for both models, regardless of index. The serial correlation exhibited by both series cloud be explained by a liquidity spiral event that occurred in the these markets during the pandemic market stress time window. In terms of model fit, we can therefore conclude that the SV model is superior to the GARCH model, and that while the fit is worse during market stress, the exogenous shock caused by the pandemic seems to worsen the fit more drastically than the endogenous shock of the credit crisis.

When comparing the results obtained from the two sub-samples characterised by market stress, it becomes apparent that the market behaviour differs. As the two shocks arose from either endogenous or exogenous events, two possible explanation for the difference in behaviour can be found. Firstly, the market participants knowledge and understanding of what is causing the crisis, and in turn its, implications for the economy is one factor. During the financial crisis market participants presumably were more knowledgeable about the implications and root cause of the crisis, which stands in contrast to the pandemic where the same participants presumable are less knowledgeable about the implications of the virus. Secondly, when a crisis arises from within the financial system, the implications for the markets can be assumed to be greater as the crisis directly relates to the economy and markets. In the case of the pandemic this is not the case as the pandemic is not a result of underlying problems within the financial system or economy, such issues can arise a result of the pandemic but are not directly related to the crisis itself.

In regards to the goodness of fit and estimation performance, the GARCH model is in general concluded to provide slightly better estimates of future volatility for the majority of the sub-samples we examined. Looking at the two indices and estimation errors before the crisis, the GARCH model was clearly favored on the Swedish data, while the models were comparable for the US data. However, once market stress was prevalent in the data, the relative performance of the SV model increased, being comparable on the Swedish data as well. This could indicate that the SV model can potentially be favored during periods of market stress. Furthermore, we also saw that the models volatility estimation abilities in general suffered as market stress was prevalent in the data and even more so during the exogenous pandemic.

We are thus able to conclude that the SV model experienced better result for the validating the model, however the GARCH model is in general able to produce volatility estimates to a similar, and to an even higher degree for the OMXS30 during stable markets. Moreover, during the periods characterised by market stress, both models performance was decreased and even more so during the pandemic for both series. Therefore, we may finally conclude that the models perform better during endogenous shocks as opposed to exogenous shocks for the sub-samples examined in the thesis.

4.1 Further research

The scope of the research can be extended in a myriad of ways. Some examples focusing on how we would be able to draw more specific conclusions are given bellow.

- The thesis hopes to lay forth the groundwork for researching the forecasting performance of the two models during the market regimes we have explored. How do the models compare in terms of forecasting performance?
- Include additional historical market shocks and markets in order to draw more specific conclusions on when the models are appropriate to use. For instance, how would the models react to market shocks to the Russian or Asian financial markets during stable and highly volatile markets?
- Introduce an extension by adding to the diffusion part jumps into the GARCH model, to better capture market turbulence and consequently improve the model fit issues discovered in this paper. Alternatively, one could introduce a gamma variance distribution for the error terms as an alternative to the diffusion with jumps model. Could the extended GARCH model provide a better fit, and consequently fit the data better then the SV model?
- Introduce a change point indication for the selection of when the models should be deployed based on market climate to allow to the strengths of the SV and GARCH models to be combined. At what exact point should we use the SV model and at what exact point should we use the GARCH model?

5 References

Albert, J.H. and Chib, S., 1993. Bayes inference via Gibbs sampling of autoregressive time series subject to Markov mean and variance shifts. *Journal of Business & Economic Statistics*, 11(1), pp.1–15.

Allen, D., 2020. Stochastic Volatility and GARCH: Do Squared End-of-Day Returns Provide Similar Information? *Journal of Risk and Financial Management*, 13(9), p.202.

Allen, D. and McAleer, M., 2020. Do we need stochastic volatility and generalised autoregressive conditional heteroscedasticity? Comparing squared end-of-day returns on FTSE. *Risks*, 8(1), p.12.

Andersen, T. and Benzoni, L., 2009. Stochastic Volatility, Working Paper 2009-04.

Andersen, T., Bollerslev, T., Diebold, F., and Labys, P., 2003. Modeling and forecasting realized volatility. *Econometrica*, 71(2), pp.579–625.

Aydemir, A.B., 2002. *Volatility modelling in finance*. Ed. by K. J. and S. S. 2nd ed. Vol. 1. Elsevier.

Barndorff-Nielsen, O.E. and Shephard, N., 2002. Econometric analysis of realized volatility and its use in estimating stochastic volatility models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 64(2), pp.253–280.

Bollerslev, T., 1986. Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, 31(3), pp.307–327.

Durham, G.B., 2007. SV mixture models with application to S&P 500 index returns. *Journal of Financial Economics*, 85(3), pp.822–856.

Engle, R.F., 1982. Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica*, 50(4), pp.987–1007.

Ghysels, E. and Harvey, A. C. and Renault, E., 1996. 5 Stochastic volatility. *Handbook of Statistics*, 14, pp.119–191.

Gilks, W.R., 1996. Introducing Markov chain monte carlo. *Markov chain Monte Carlo in practice.*

Hansen, P.R. and Lunde, A., 2006. Realized variance and market microstructure noise. *Journal of Business & Economic Statistics*, 24(2), pp.127–161.

Hosszejni, D. and Kastner, G., 2021. *Package 'stochvol'.* (Technical report).

Jacquier, E. and Polson, N.G., 2011. Bayesian Methods in Finance. *The Oxford Handbook of Bayesian Econometrics*, pp.439–512.

Johannes, M. and Polson, N.G., 2009. Markov chain monte carlo. *Handbook of Financial Time Series.* Springer, pp.1001–1013.

Kastner, G., 2019. Dealing with stochastic volatility in time series using the R package stochvol. *arXiv preprint arXiv:1906.12134.*

Kastner, G. and Frühwirth-Schnatter, S., 2014. Ancillarity-sufficiency interweaving strategy (ASIS) for boosting MCMC estimation of stochastic volatility models. *Computational Statistics & Data Analysis*, 76, pp.408–423.

Kim, S., Shephard, N., and Chib, S., 1998. Stochastic volatility: likelihood inference and comparison with ARCH models. *The Review of Economic Studies*, 65(3), pp.361–393.

Lopes, H. and Polson, N.G., 2010. Extracting SP500 and NASDAQ volatility: The credit crisis of 2007-2008. *Handbook of Applied Bayesian Analysis*, pp.319–342.

Poon, S. and Granger, C.W., 2003. Forecasting volatility in financial markets: A review. *Journal of Economic Literature*, 41(2), pp.478–539.

Rizzo, M.L., 2019. *Statistical computing with R.* CRC Press.

Taylor, S.J., 2008. *Modelling financial time series.* 2nd ed. Singapore: World Scientific.

Tierney, L., 1994. Markov chains for exploring posterior distributions. *The Annals of Statistics*, 22(4), pp.1701–1728.

Van Ravenzwaaij, D., Cassey, P., and Brown, S.D., 2018. A simple introduction to Markov Chain Monte–Carlo sampling. *Psychonomic Bulletin & Review*, 25(1), pp.143–154.

Wuertz, D., Setz, T., Chalabi, Y., Boudt, C., Chausse, P., and Miklovac, M., 2019. *FGarch: Rmetrics - autoregressive conditional heteroskedastic modelling.* (Technical report). Technical report, working paper/manual. URL https://CRAN.R-project.org/package=fGarch.