MAGISTERARBEIT / MASTER’S THESIS

Titel der Magisterarbeit / Title of the Master‘s Thesis

„Comparing Classes of Volatility Models“

verfasst von / submitted by

Alexander Leo Kuhn, BSc

angestrebter akademischer Grad / in partial fulfilment of the requirements for the degree of

Magister der Sozial- und Wirtschaftswissenschaften (Mag. rer. soc. oec.)

Wien, 2020 / Vienna 2020

Studienkennzahl lt. Studienblatt / degree programme code as it appears on the student record sheet:	UA 066 951
Studienrichtung lt. Studienblatt / degree programme as it appears on the student record sheet:	Magisterstudium Statistik
Betreut von / Supervisor:	ao. Univ.-Prof. Dipl.-Ing. Dr. Erhard Reschenhofer

Contents

List of Figures

List of Tables

Chapter 1

Introduction

In economic and statistical literature, there are two major approaches when it comes to the estimation of latent variance and volatility: the Frequentist approach, where model estimation is usually straightforward and computationally easy to achieve, and the Bayesian approach, where the volatility of a stochastic process (like stock prices or exchange rates over time) in itself follows a random distribution. This twofold view of estimation methodology is not limited to the field of econometrics, e.g. interval estimation and testing procedures (Hogg, McKean, and Craig 2013) or classification methods (Hastie, Tibshirani, and Friedman 2009) all have similar distinctions.

From a computational point of view, Bayesian models are much more intense, as the parameter estimates and their posterior distributions are generated by thousands (or even millions) of draws from a Gibbs sampler or a more sophisticated algorithm (e.g. Metropolis-Hastings). The goal of this thesis is to determine, whether the increase in complexity and calculation length can be offset by more accurate predictions with respect to in-sample and out-of-sample errors in expanding or rolling window forecasts. For this purpose, the time-series of the S&P 500 Index, the EUR/USD exchange rate and the Gold spot price will be examined. To further evaluate the robustness in extremely volatile scenarios, the author chose the price series of the cryptocurrency Bitcoin (BTC), which contains daily log-returns as high as ± 0.4. The specifications and intricacies of each considered model, a broad comparison of methodology and motivation behind each, as well as the results regarding model accuracy, predictive quality and estimation complexity will be presented. The thesis will be concluded with a review to discuss the suitability of each model in different scenarios.

Chapter 2

Notation and Model Specification

While the ecosystem of econometric models is much larger, three main models will be considered: HAR (*heterogeneous autoregression*), GARCH (*generalized autoregressive conditional heteroscedasticity*) and SV (*stochastic volatility*). They differ widely in the level of theoretical knowledge necessary to derive the model equations as well as in the complexity of estimation. Furthermore, they represent the two (statistical) universes introduced in Chapter 1.

2.1 Return and Volatility

To determine how volatile a time-series is, we have to consider the logarithmic returns r_t. These are the changes in price P_t at time t relative to the price P_{t-1} at a previous time $t-1$ (Brockwell and Davis 2016). The log-returns can be calculated on a yearly, monthly, daily, or (if sufficient data is available) even intra-daily level. The volatility σ_t would then be the conditional standard deviation of r_t, given all past returns, i.e.

$$\sigma_t = \sqrt{\mathrm{Var}\left(r_t | \mathcal{F}_{t-1}\right)}$$

where $\mathcal{F}_{t-1}$ denotes the σ-algebra induced by $r_{t-1}, r_{t-2}, ...$ As the conditional standard deviation is not directly observable, the absolute daily log-returns will serve as the proxy to fit the HAR model.[1]

[1] In the unlikely (but possible) case, that r_t is exactly or very close to 0, the author added an offset of 10^{-13} to r_t in order to avoid computational problems.

While we are usually more interested in the actual return for tomorrow, the nature of stochastic processes is the impossibility of predicting even the direction of the price series. Nevertheless there are approaches that try to determine the magnitude of change in price. This can in turn be used to determine, whether a certain investment A yields more or less expected risk than investment B.

2.2 Model Specification

2.2.1 HAR

The HAR or heterogeneous autoregression model was first introduced in (Corsi 2007). Consider the following equation for the volatility of the next day $t + 1$

$$\sigma_{t+1} = \beta_0 + \beta_1 \, \sigma_t^{(d)} + \beta_2 \, \sigma_t^{(w)} + \beta_3 \, \sigma_t^{(m)} + \varepsilon_t \tag{2.1}$$

where $\sigma_t^{(d)}$ denotes the volatility of the current day and $\sigma_t^{(w)}$, $\sigma_t^{(m)}$ denote the **average** volatility over the past week and month respectively, i.e.

$$\sigma_t^{(w)} = \frac{1}{5} \sum_{i=0}^{5} \sigma_{t-i}$$

and

$$\sigma_t^{(m)} = \frac{1}{21} \sum_{i=0}^{21} \sigma_{t-i}$$

This corresponds to 5 trading days per week and 21 trading days per month. The choice of referencing the model coefficients as β is to emphasize the linear regression framework the model is built on. The only assumption w.r. to the distribution of ε_t is that it has zero mean and is truncated on the left to ensure the positivity of volatility (Corsi 2007). To estimate the HAR model we simply use $|r_t|$ as a proxy for the unobserved σ_t in (2.1), create the covariate vectors for each time window (day, week, month) and calculate the ordinary least squares (OLS) estimate for the coefficient vector $\boldsymbol{\beta}$.

The OLS approach makes the model easy to estimate, yields many varieties of inference and enables straightforward interpretation and forecasting. Many extensions exist, most notably (Corsi and Renò 2009) introducing leverage and jumps, as well as (Bollerslev, Patton, and Quaedvlieg 2016) extending HAR to include measurement errors of intraday volatility.

2.2.2 GARCH

The GARCH or generalized autoregressive conditional heteroscedasticity model uses a more complex construction (Hansen and Lunde 2005). It is defined by two model equations, one for the return r_t and one for the underlying volatility process:

$$r_t = \mu + \sigma_t \, \varepsilon_t, \quad \varepsilon_t \sim \mathcal{N}(0, 1) \tag{2.2}$$

$$\sigma_t^2 = \omega + \sum_{i=1}^{p} \alpha_i \, r_{t-i}^2 + \sum_{j=1}^{q} \beta_j \, \sigma_{t-j}^2 \tag{2.3}$$

Therefore, the variance σ_t^2 only depends on the conditional density $f(r_t \,|\, \mathcal{F}_{t-1})$ where $\mathcal{F}_{t-1}$ denotes the σ-algebra induced by $r_1, ..., r_{t-1}$. As (Hansen and Lunde 2005) have shown, from the frequentist universe, the GARCH(1,1) model yields the best results considering model complexity. The model equation for GARCH(1,1) is given by

$$\sigma_t^2 = \omega + \alpha_1 \, r_{t-1}^2 + \beta_1 \, \sigma_{t-1}^2 \tag{2.4}$$

Recursive substitution yields

$$\sigma_t^2 = \omega + \alpha_1 \, r_{t-1}^2 + \beta_1 \, (\omega + \alpha_1 \, r_{t-2}^2 + \beta_1 \, \sigma_{t-2}^2) = ... = \tilde{\omega} + \sum_{i=1}^{t-1} \tilde{\alpha} \, r_{t-i}^2 + \tilde{\beta} \, \sigma_1^2 \tag{2.5}$$

Note that plugging in the result from (2.5) into (2.2) implies an ARMA(1,1) model for the squared returns. The interpretation of the 4 parameters in the GARCH(1,1) model is straightforward:

- μ represents the expected value of r_t and is often assumed to be 0.
- ω is the baseline value for the latent volatility process
- "α_1 measures the extent to which a volatility shock today feeds trough into next period's volatility" (Campbell, Lo, and MacKinley 1996)
- "$(\alpha_1 + \beta_1)$ measures the rate at which this effect dies out over time" (Campbell, Lo, and MacKinley 1996) - also known as persistence

In order to estimate the parameter vector $\theta = (\mu, \omega, \alpha_1, \beta_1)'$, we have to consider the joint density over all observations/returns: $f_\theta\,(r_1, ..., r_{T-1}, r_T)$. We can decompose the density like this:

$$f_\theta\,(r_1, ..., r_{T-1}, r_T) = \frac{f_\theta\,(r_1, ..., r_{T-1}, r_T)}{f_\theta\,(r_1, ..., r_{T-2}, r_{T-1})} \frac{f_\theta\,(r_1, ..., r_{T-2}, r_{T-1})}{f_\theta\,(r_1, ..., r_{T-3}, r_{T-2})} ... f_\theta\,(r_1)$$

where each fraction represents the conditional density, i.e.

$$\frac{f_\theta\left(r_1,...,r_{T-1},r_T\right)}{f_\theta\left(r_1,...,r_{T-2},r_{T-1}\right)} = f_\theta\left(r_T|r_1,...,r_{T-1}\right) = f_\theta\left(r_T \mid \mathcal{F}_{T-1}\right)$$

Substitution yields

$$f_\theta\left(r_1,...,r_{T-1},r_T\right) = f_\theta\left(r_T|\mathcal{F}_{T-1}\right) f_\theta\left(r_{T-1}|\mathcal{F}_{T-2}\right) ... f_\theta\left(r_1\right) = \prod_{t=1}^{T} f_\theta\left(r_t|\mathcal{F}_{t-1}\right)$$

Now we can maximize the likelihood w.r.t. θ using the log-likelihood (in the Gaussian case)

$$l\left(\theta\right) = \sum_{t=1}^{T} f_\theta\left(r_t|\mathcal{F}_{t-1}\right) \propto -\frac{1}{2}\sum_{t=1}^{T}\left(\log\sigma_t^2 + \frac{\left(r_t-\mu\right)^2}{\sigma_t^2}\right)$$

This can be done recursively given an initial value of σ_1^2, e.g. the unconditional variance $\frac{\omega}{1-\alpha_1-\beta_1}$ or the sample return variance.

The standard GARCH model can be extended in many different ways, accounting for asymmetric response, leverage, fat-tailed conditional densities and multivariate scenarios (Hansen and Lunde 2005).

2.2.3 SV

The SV or stochastic volatility model offers a radically different approach for describing the latent volatility process. Represented in an hierarchical structure, the returns r_t are assumed to be normally distributed, conditioned on the unobserved latent volatility process h_t. This means that each return has its own level of variance. This unobserved latent volatility process is itself normally distributed, conditioned on the previous latent volatility h_{t-1}, the level of log-variance μ, the persistence of log-variance ϕ and the volatility of log-variance σ_η. Formally, the model can be written down in three equations:

$$r_t \mid h_t \sim \mathcal{N}\left(0, \exp h_t\right) \tag{2.6}$$

$$h_t \mid h_{t-1}, \mu, \phi, \sigma_\eta \sim \mathcal{N}\left(\mu + \phi\left(h_{t-1}-\mu\right), \sigma_\eta^2\right) \tag{2.7}$$

$$h_0 \mid \mu, \phi, \sigma_\eta \sim \mathcal{N}\left(\mu, \sigma_\eta^2/\left(1-\phi^2\right)\right) \tag{2.8}$$

As SV is a Bayesian model, we need to specify prior distribution for each entry in the parameter vector $\theta = \left(\mu, \phi, \sigma_\eta\right)'$. The initial values are randomly drawn from those distributions. A good choice for the priors is crucial, as the priors influence the final posterior draws. An overly uninformative prior (e.g. $\mathcal{U}\left[-10000, 10000\right]$) will be strongly overruled by the data, resulting in an overfit model. If the chosen prior is too strong (e.g. $\mathcal{N}\left(0, 1\right)$) it might dominate the data and the posterior would be strongly biased. The specific choice of priors for this thesis will be discussed in the next chapter.

Chapter 3

Data and Methodology

In this chapter the choice of financial time-series will be discussed, while introducing the methods of the underlying `R` packages to fit the models as well as explaining the choice of priors in the SV model. The main goal was to create a general comparison in multiple scenarios, measure the influence of increased volatility and especially the suspected (illicit) human intervention in the Bitcoin peak at the end of 2017.

3.1 Financial Data

Apart from the Bitcoin (BTC) time-series which is interesting to study for multiple reasons, the author used other time-series data that is well established in the literature. The S&P 500 Index (SPX) was used in (Corsi 2007) to measure the effectiveness of the HAR model. (Corsi 2007) also used the EUR/CHF value pair which was actively influenced by the Swiss national bank intervention in early 2015. Therefore the EUR/USD exchange rate was used. Another time-series of particular interest due to the large role it plays in international financial markets and institutions is the gold price (XAU). The time-series of SPX, EUR/USD and XAU will be referred to as the "standard time-series" to emphasize the difference in behaviour.

The considered time frame is 10/2006 through 05/2018 for SPX, EUR/USD and XAU available in daily closing prices (3000 observations each) from the financial platform of Reuters[1]. For BTC the records start in 07/2014, yielding 1374 observations (due to BTC being traded around the clock).

[1] `https://www.reuters.com/finance`

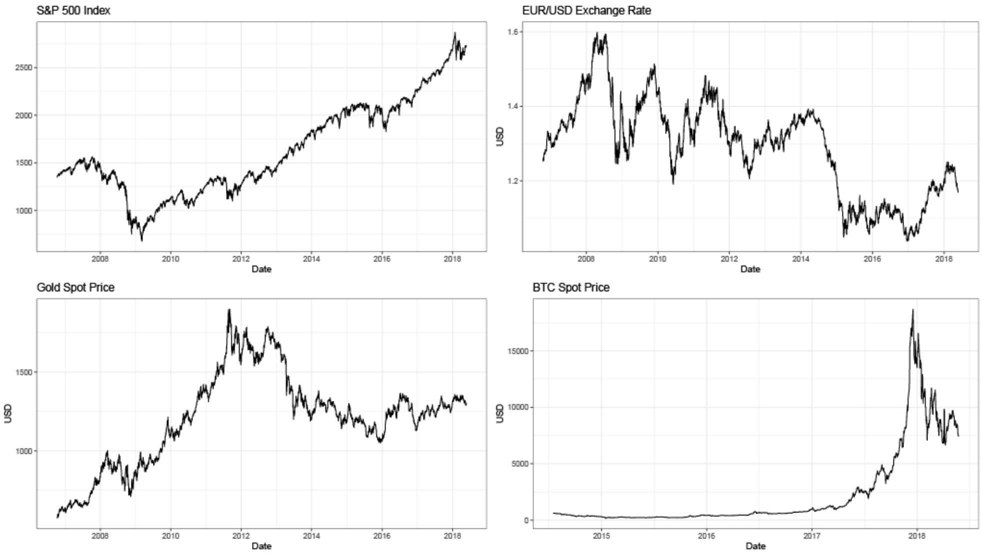

Figure 3.1: Price series for SPX, EUR/USD, XAU and BTC

In Figure 3.1 we can already see the different behaviour of the prices over time, most notably the peak of BTC in late 2017. To compare the different magnitudes of volatility/variance we need to check the logarithmic return series in Figure 3.2.

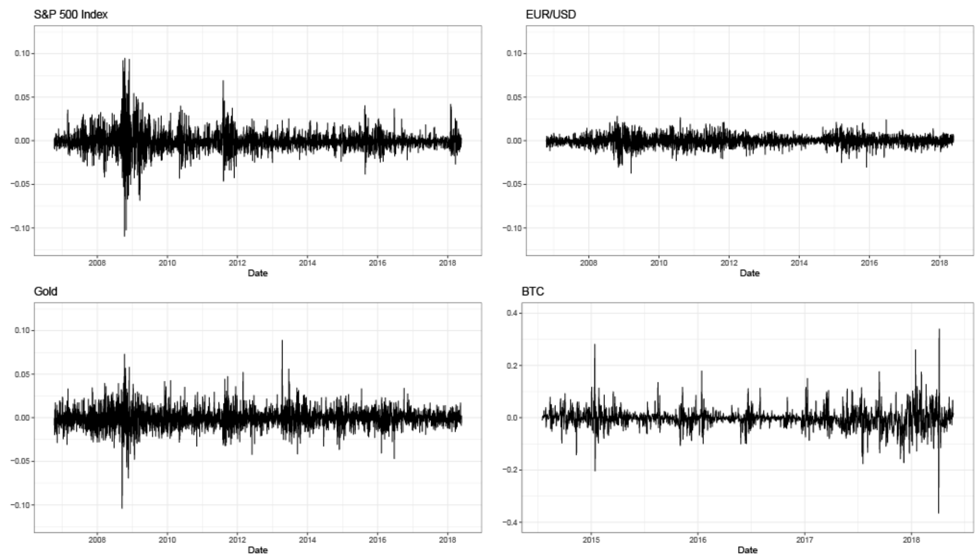

Figure 3.2: Logarithmic return series for SPX, EUR/USD, XAU and BTC

Note that the scale of the returns is different by a factor of 4 between the standard time-series and BTC.

3.2 Estimation Period and Comparison

To have consistent numbers for comparison, the author opted to estimate all three models in different ways:

- Rolling windows of size 250, 500, 750 and 1000 with one-day-ahead out-of-sample forecast
- Expanding windows with initial sizes of 250, 500, 750 and 1000 with one-day-ahead out-of-sample forecast

To compare and benchmark the different models with respect to their out-of-sample forecast, the author chose the method of predictive likelihoods established in (Geweke and Amisano 2010). As opposed to more traditional measures such as the mean absolute error (MAE) and mean squared error (MSE), the predictive likelihood takes the variance into account. By construction, the assumed distribution for HAR and GARCH is normal with mean 0 and standard deviation σ_t. Therefore it is easy to calculate the predictive likelihood using the predicted volatility $\hat{\sigma}_t$:

$$PL\left(\hat{\sigma}_t\right) = f\left(r_t\right) \tag{3.1}$$

where f denotes the probability density function of a normal distribution with mean 0 and standard deviation $\hat{\sigma}_t$. Trivially, the total likelihood of n predictions by any model A (e.g. from an expanding window forecast mentioned above) can be represented as the sum of predictive log-likelihoods:

$$\log PL_A = \log \prod_{t=1}^{n} PL\left(\hat{\sigma}_t\right) = \sum_{t=1}^{n} \log PL\left(\hat{\sigma}_t\right)$$

While it is relatively easy to extract the predictive likelihood from HAR and GARCH models it is not a trivial task to derive this property in Bayesian models. The following remark is based on the paper (Geweke and Amisano 2010) and enables the use of a similar approach, where Equation (3.1) again uses the probability density function of a normal distribution with mean 0 but standard deviation $\left(\exp \hat{h}_t\right)$ as defined in Equation (2.6). The estimate for $\hat{h}_t$ is the median of the posterior distribution of h_t.

3.2.1 Remark: Predictive Likelihood in Bayesian Models

To justify the use of predictive likelihoods the author wants to recite the important implications from (Geweke and Amisano 2010): The one-step-ahead predictive likelihood of the return r_t estimated by some volatility model A, which can be evaluated only at time t or later, is the real number

$$\begin{aligned} PL_A(t) &= p(r_t \mid r_1, ..., r_{t-1}, A) \\ &= \int_{\theta_A} p(r_t \mid r_1, ..., r_{t-1}, \theta_A, A) \times p(r_t \mid r_1, ..., r_{t-1}, A) \, \mathrm{d}\theta_A \end{aligned}$$

In most time-series models the evaluation of $p(r_t \mid r_1, ..., r_{t-1}, \theta_A, A)$ through the posterior distribution is straightforward, leading to the following approximation:

$$M^{-1} \sum_{m=1}^{M} p\left(r_t \mid r_1, ..., r_{t-1}, \theta_{A,t-1}^{(m)}, A\right)$$

using an ergodic sequence $\{\theta_{A,t-1}^{(m)}\}$ with M draws from a posterior simulator. For the data set $\mathbf{r}_T$, the marginal likelihood of the model A is

$$p(\mathbf{r}_T \mid A) = \prod_{t=1}^{T} p(r_t \mid r_1, ..., r_{t-1}, A)$$

implying the additive decomposition

$$\log p(\mathbf{r}_T \mid A) = \sum_{t=1}^{T} \log PL_A(t)$$

3.3 Choice of Prior Distributions for Stochastic Volatility

As mentioned earlier, the choice of the proper prior distributions is crucial, especially regarding the balance of information and bias. For stochastic volatility models the priors are well established (Kastner 2016):

- $\mu \sim \mathcal{N}(b_\mu, B_\mu)$
- $(\phi + 1)/2 \sim \mathcal{B}(a_0, b_0)$
- $\sigma_\eta^2 \sim \mathcal{G}(1/2, 1/2)$

For the level of log-variance μ an almost flat normal distribution with $b_\mu = 0$ and $B_\mu = 10000$ was chosen. For the persistence of log-variance ϕ, normalized as seen above, a Beta distribution was used to make sure the parameter stays between -1 and 1. The shape parameters for the Beta distribution $a_0 = 5$ and $b_0 = 1.5$ imply mean of 0.77 and a standard deviation of 0.15, leading to very little mass for negative values of ϕ. The volatility of log-variance σ_η has a Gamma prior, which is a very common choice for dispersion parameters, with shape $\alpha, \beta = 0.5$. The prior density functions are shown in Figure 3.3.

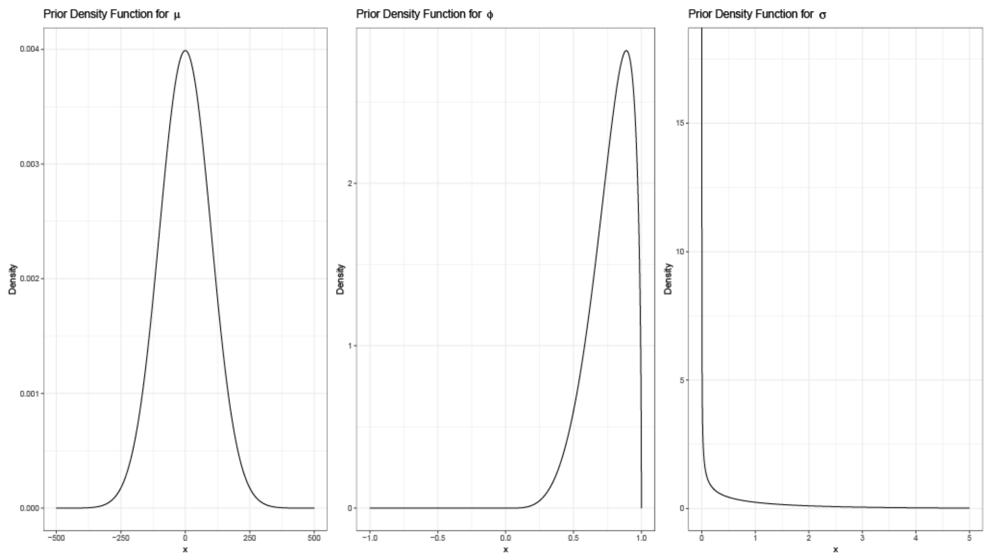

Figure 3.3: Prior density functions for μ, ϕ and σ

Chapter 4

Results

4.1 Model Output and In-Sample Estimation

The in-sample fit was obtained by fitting a single model to each time-series using all observations. This was done in order to build an intuition for the way the models differ and to visualize those differences. This section will use the S&P 500 Index time-series as an illustration.

4.1.1 HAR

	Estimate	Std. Error	t value	Pr(> |t|)
β_0	0.0010	0.0003	3.84	0.0001
β_1	-0.1168	0.0215	-5.42	0.0000
β_2	0.4628	0.0506	9.14	0.0000
β_3	0.5303	0.0510	10.40	0.0000

Table 4.1: HAR model coefficients for SPX

As explained before, the coefficients represent the OLS estimate of the average volatility for 1, 5 and 22 days respectively. The value of β_0 suggest a very low ground level of volatility, while due to non-orthogonal regressors (the daily, weekly and monthly log-returns are not independent) there is no meaningful interpretation for their size or sign. Following (Corsi 2007) it can be shown (this was done for the S&P 500 among others) that the HAR model manages to capture the empirically observed long-memory (or persistence) of volatility.

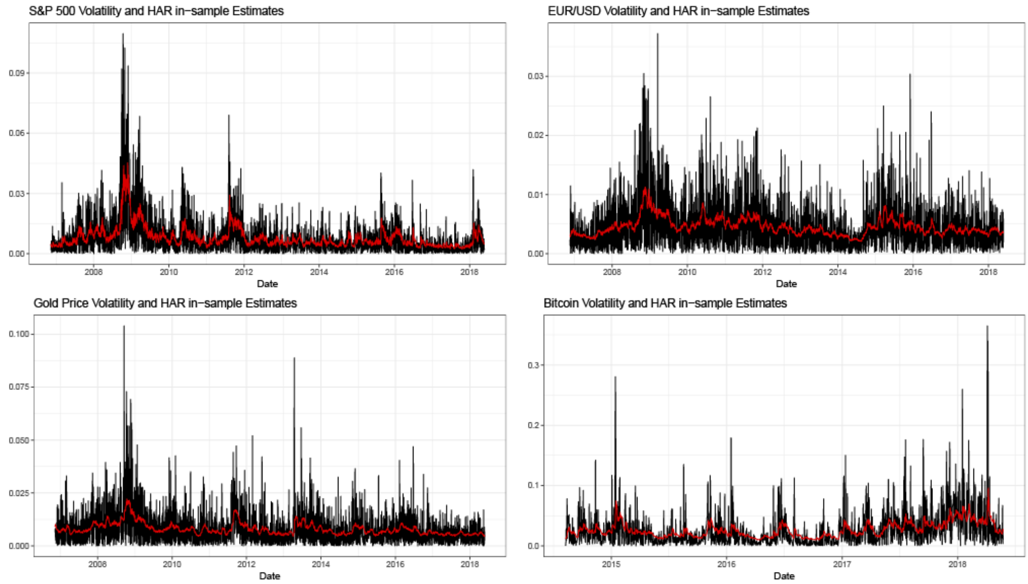

Figure 4.1: Volatility and in-sample HAR fit for SPX, EUR/USD, XAU and BTC

The comparison of the volatility proxy (black) and HAR in-sample estimates (red) in Figure 4.1 reveals, what intuitively was clear from the beginning: by the nature of averages over longer periods and the effect size, HAR generally underestimates the volatility but factors in major shocks of the underlying time-series for a reasonable amount of time, resembling a clear picture of time stretches with different levels of volatility.

4.1.2 GARCH

	Estimate	Std. Error	t value	Pr(> |t|)
μ	0.0063	0.0006	10.7362	0.0000
AR(1)	0.9811	0.0054	181.6528	0.0000
MA(1)	-0.8947	0.0143	-62.4819	0.0000
ω	0.0000	0.0000	2.0323	0.0421
α_1	0.1204	0.0129	9.3571	0.0000
β_1	0.8547	0.0133	64.4369	0.0000

Table 4.2: GARCH model coefficients for SPX

A notable difference of the GARCH estimates in comparison to HAR is the higher baseline of volatility ($\mu = 0.0063$ as compared to $\beta_0 = 0.001$). The persistence parameter ($\alpha_1 + \beta_1$) is very high (0.9751) which is clearly visible in the in-sample estimates in Figure 4.2.

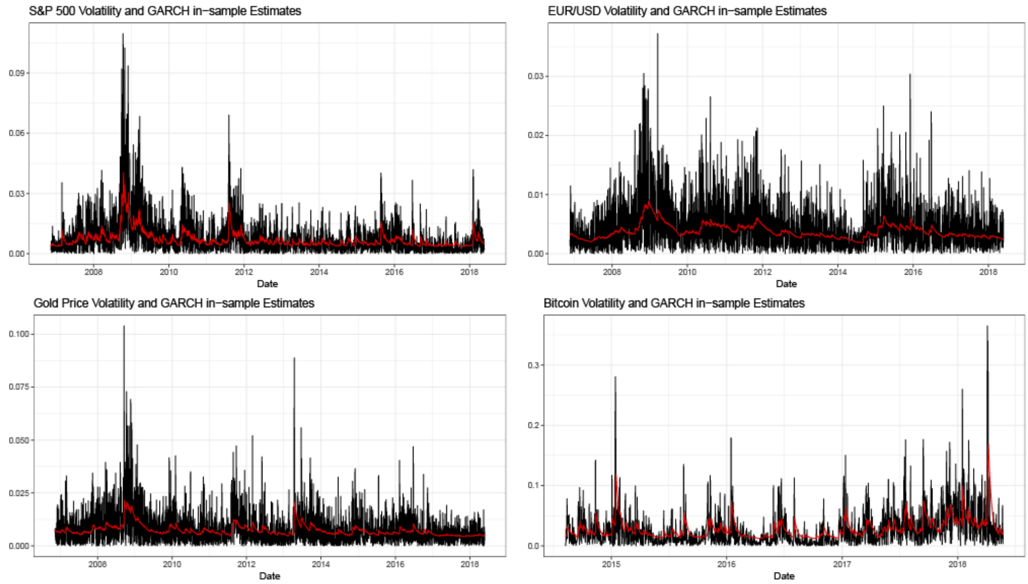

Figure 4.2: Volatility and in-sample GARCH fit for SPX, EUR/USD, XAU and BTC

The in-sample estimates (red) are smoother and react much quicker to large shocks, best seen in the Gold time-series, where the the GARCH estimates have sharp peaks. Due to the size of the persistence parameter, it also takes a while for large shocks to "wash out", leading to the typical (almost) vertical and slowly decaying peaks.

4.1.3 SV

	Mean	SD	5%	Median	95%
μ	-9.5377	0.2179	-9.8947	-9.5390	-9.1832
ϕ	0.9800	0.0050	0.9714	0.9802	0.9877
σ	0.2206	0.0206	0.1873	0.2200	0.2561
$\exp(\mu/2)$	0.0085	0.0009	0.0071	0.0085	0.0101
σ^2	0.0491	0.0092	0.0351	0.0484	0.0656

Note that the estimate of μ is on the logarithmic scale. The "true" value for the volatility baseline is given by $\exp(\mu/2)$ and is even higher than the GARCH baseline. The same holds for the persistence parameter, which is exceptionally high with $\phi = 0.98$.

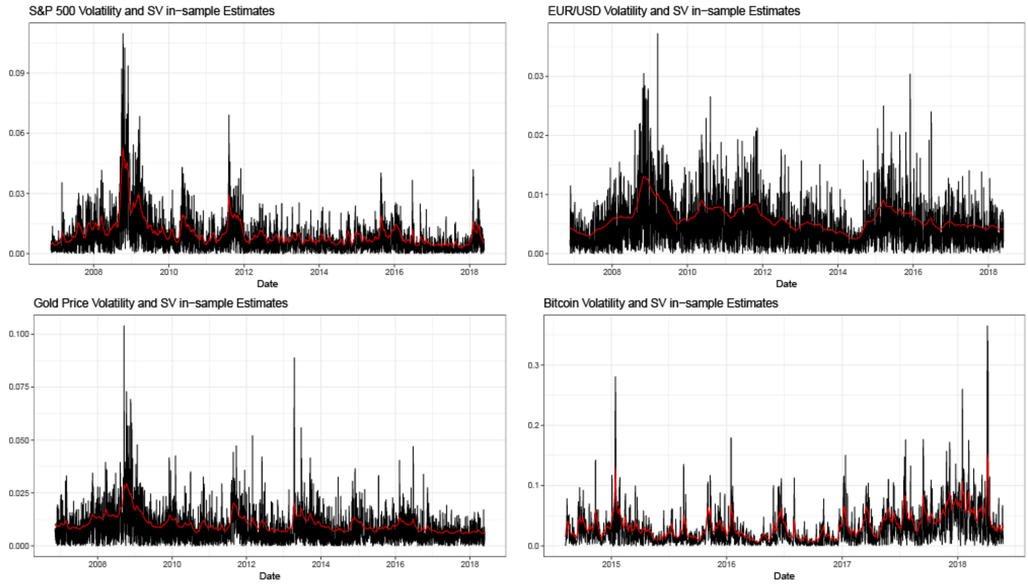

Figure 4.3: Volatility and in-sample SV fit for SPX, EUR/USD, XAU and BTC

The parameters are easily verified by looking at the in-sample estimates in Figure 4.3: the level of predicted volatility (red) is much higher and shows clear stretches of low and high volatility periods. It also decays much faster than the GARCH estimates. Intuitively, the SV model gives the best in-sample fit while also estimating a probability distribution for each estimate (and therefore, also for future predictions). This can be extended to risk estimation, interval estimation and also the use of either the mean or median for prediction.

4.2 Out-of-sample Prediction

As the results are very similar for the expanding and rolling window estimates, only the expanding window with an initial window size of 1000 observations will be considered. This has two reasons: firstly, because one would usually use as much of the available data as possible, to give the model time to reduce the bias caused by random outliers. Secondly, because the initial window size of 1000 performed considerably better when compared to other window sizes. As explained and derived in Section 3.2, the predictive likelihood was used to measure the model fit to the volatility proxy. Detailed results for all window types and sizes can be found in Table 4.3 and Table 4.4.

		GARCH		HAR		SV	
Instrument	Window	Expand	Roll	Expand	Roll	Expand	Roll
SPX	250	9353	9356	8541	8893	10142	10288
SPX	500	8569	8560	7821	7998	9270	9365
SPX	750	7913	7893	7460	7653	8531	8600
SPX	1000	7072	7062	6662	6829	7630	7675
EUR/USD	250	11019	10804	10619	10643	11821	11786
EUR/USD	500	9980	9758	9649	9575	10727	10640
EUR/USD	750	9145	9031	9013	8953	9792	9755
EUR/USD	1000	8197	8150	8066	8037	8753	8743
XAU	250	9518	9371	9231	9296	10220	10403
XAU	500	8676	8567	8465	8519	9342	9456
XAU	750	7954	7894	7808	7862	8539	8616
XAU	1000	7113	7069	6976	7044	7636	7679
BTC	250	2600	2596	2423	2506	3054	3072
BTC	500	1980	1944	1835	1851	2334	2337
BTC	750	1350	1334	1240	1231	1609	1610
BTC	1000	696	685	630	612	857	853

Table 4.3: Total predictive log-likelihood

		GARCH		HAR		SV	
Instrument	Window	Expand	Roll	Expand	Roll	Expand	Roll
SPX	250	3.4963	3.4975	3.1930	3.3244	3.7915	3.8460
SPX	500	3.5335	3.5300	3.2250	3.2980	3.8226	3.8618
SPX	750	3.6382	3.6290	3.4301	3.5186	3.9223	3.9542
SPX	1000	3.6735	3.6687	3.4607	3.5476	3.9636	3.9869
EUR/USD	250	3.9650	3.8878	3.8213	3.8298	4.2539	4.2412
EUR/USD	500	3.9462	3.8585	3.8155	3.7863	4.2414	4.2073
EUR/USD	750	4.0127	3.9626	3.9549	3.9284	4.2968	4.2802
EUR/USD	1000	4.0400	4.0167	3.9756	3.9612	4.3140	4.3090
XAU	250	3.4276	3.3743	3.3240	3.3475	3.6802	3.7460
XAU	500	3.4333	3.3901	3.3499	3.3713	3.6969	3.7419
XAU	750	3.4932	3.4670	3.4290	3.4528	3.7499	3.7838
XAU	1000	3.5093	3.4875	3.4413	3.4749	3.7673	3.7886
BTC	250	2.3135	2.3097	2.1553	2.2296	2.7175	2.7330
BTC	500	2.2652	2.2238	2.1001	2.1174	2.6708	2.6740
BTC	750	2.1630	2.1380	1.9864	1.9727	2.5789	2.5808
BTC	1000	1.8616	1.8311	1.6840	1.6374	2.2917	2.2812

Table 4.4: Predictive log-likelihood adjusted for different period lengths

The adjusted log-likelihood takes into account that smaller window sizes yielded more prediction, leading to larger total log-likelihood. In essence, this is equivalent to the average log-likelihood per prediction. To compare the performance of the different models easily, which is the central goal of this thesis, the cumulative log-likelihood for the expanding window with initial size of 1000 are visualized in Figure 4.4.

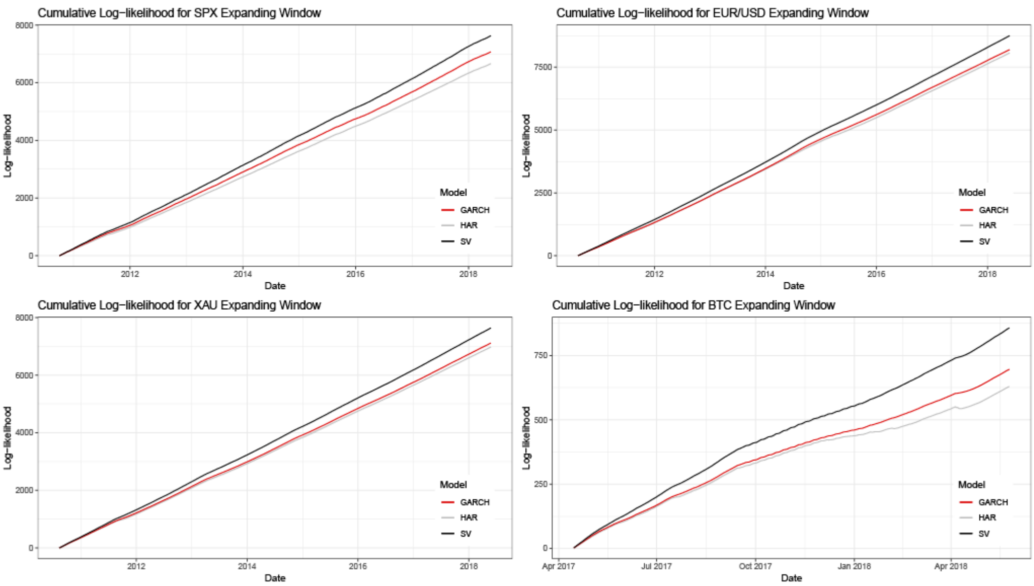

Figure 4.4: Cumulated predictive log-likelihood for expanding windows (n=1000)

It is easy to see that the SV model outperforms GARCH and HAR by some margin. It is also interesting that GARCH outperforms HAR considerably in more volatile scenarios (S&P 500 and Bitcoin price), while they are quite close in the less volatile time-series (EUR/USD exchange rate and Gold price). Furthermore, while the increase in total log-likelihood happens almost uniformly over the time-series for S&P 500, EUR/USD and Gold, it fluctuates strongly for Bitcoins. As mentioned, the large shocks in the Bitcoin time-series are of magnitude 0.3-0.4, therefore it would be surprising, if any model could predict those kind of outliers in a reliable way. This is also highlighted in the fact that the predictions of the HAR model even take on negative log-likelihood values, placing them particularly far away from the observed values. If we compare the actual predictions by model, as seen in Figure 4.5, we see a similar picture to the in-sample estimates.

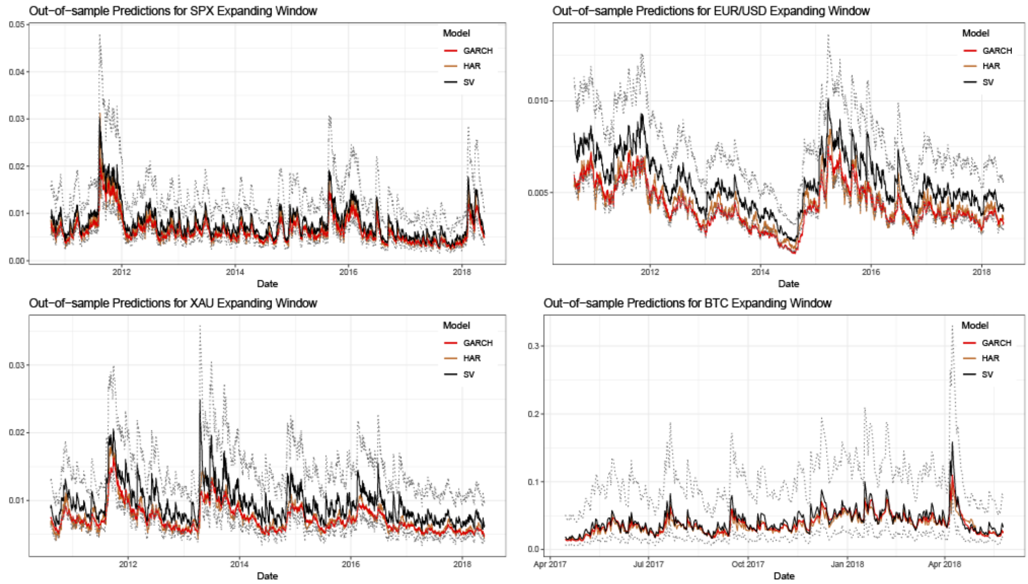

Figure 4.5: Predicted volatility for expanding windows (n=1000)

The SV model predicts the highest values, while GARCH and HAR are not far apart. Figure 4.5 also contains the 2.5% and 97.5% quantiles of the SV predictions as calculated from the conditional posterior distribution. It is interesting to note that in the Bitcoin time-series, where the SV predictions have the largest variance, all the models have a similar trajectory and both GARCH and HAR are well within the quantiles of the SV posterior distribution.

Chapter 5

Conclusion

We saw that the frequentist models (HAR and GARCH) are very close in results for low-volatility time-series, while GARCH performs considerably better in high-volatility time-series. To some extent, this was expected, as (Hansen and Lunde 2005) have shown. The more important comparison is between GARCH and SV. Here the results are twofold. Firstly, judging by predictive quality, the SV clearly has an edge. One of the reasons is certainly the complete absence of assumptions regarding the distribution of log-returns. While GARCH in general assumes them to be normally distributed, this rarely holds, as log-returns of financial time-series usually have much heavier tails in their empirical distribution. Therefore, SV allows the distribution to take any shape, which has the side effect that one can either use the mean or the median for prediction, as they usually differ (SV often produces asymmetric and leptokurtic posterior distribution). This fact also explains, why the SV predictions have higher peaks and are able to reproduce major shocks more reliably (at least in contrast to GARCH). Secondly, and this is the major downside of the SV model, the computational cost differs by a magnitude of ~ 100. Especially in the expanding window estimate, the growing number of observations led to a runtime of around 2 minutes for GARCH, while it took over 3 hours for the SV estimate. This is due to the fact that to obtain the predictive distribution of the last observation in the S&P 500 Index time-series, the MCMC algorithm has to generate 11000×3000 draws. Note that all the models were run on a Notebook with 8 cores, using 7 of them to parallelize the estimation. This makes SV infeasible for real-time or intra-day estimation. Nevertheless, it is very powerful for modelling the in-sample distribution of log-returns and predictions on a daily or weekly level. The most interesting question for further research would be the feasibility of SV using high-frequency data, where more information about the intra-day volatility levels is provided. If the rise in computational power can lead to shorter estimation durations, SV might very well be the next big thing for firms and investment funds in the financial market.

References

Bollerslev, T., A. Patton, and R. Quaedvlieg (2016). "Exploiting the errors: A simple approach for improved volatility forecasting". In: *Journal of Econometrics* 192.1, pp. 1–18.

Brockwell, P. and R. Davis (2016). *Introduction to Time Series and Forecasting.* 3rd ed.

Campbell, J., A. Lo, and A. MacKinley (1996). *The Econometrics of Financial Markets.* 2nd ed. Chap. 12, p. 483.

Corsi, F. (2007). "A Simple Approximative Long-Memory Model of Realized Volatility". In: *Journal of Financial Econometrics* 7.2, pp. 174–196.

Corsi, F. and R. Renò (2009). "HAR volatility modelling with heterogeneous leverage and jumps". In:

Geweke, J. and G. Amisano (2010). "Comparing and evaluating Bayesian predictive distributions of asset returns". In: *International Journal of Forecasting* 26, pp. 216–230.

Hansen, P. and A. Lunde (2005). "A forecast comparison of volatility models: does anything beat a GARCH(1,1)?" In: *Journal of Applied Econometrics* 20, pp. 873–889.

Hastie, T., R. Tibshirani, and J. Friedman (2009). *The Elements of Statistical Learning.* 2nd ed. Chap. 6.

Hogg, R., J. McKean, and A. Craig (2013). *Introduction to Mathematical Statistics.* 7th ed. Chap. 11.

Kastner, G. (2016). "Dealing with Stochastic Volatility in Time Series Using the R Package stochvol". In: *Journal of Statistical Software* 69, pp. 1–30.

Appendix A

Summary

The goal of this thesis was to give an insight into two different worlds of statistical volatility estimation: the frequentist framework, on which the easily estimable and interpretable HAR and GARCH models are build, and the bayesian SV model, which is much more difficult to calculate. Furthermore, those three models were tested on real-world financial time-series with different characteristics regarding the predictive accuracy as well as the length and complexity of estimation. The measurement of predictive accuracy follows the less common use of predictive likelihoods rather than the traditional deployment of MSE or MAE. In the conclusion, accuracy and complexity are weighed up in order to examine the suitability of each model in different scenarios.

Appendix B

Zusammenfassung der Arbeit

Das Ziel dieser Arbeit war es, einen Einblick in zwei unterschiedliche Welten der statistischen Volatilitätsschätzung zu geben. Einerseits das frequentistische Framework, auf dem die HAR und GARCH Modelle basieren, die einfach zu schätzen und zu interpretieren sind. Andererseits das bayesianische SV Modell, das per Konstruktion eine sehr viel aufwändigere Schätzmethode erfordert. Des Weiteren wurden diese 3 Modelle anhand von realen Wirtschaftsdaten aus unterschiedlichen Zeitreihen getestet und sowohl die Genauigkeit der Prognose, als auch die Dauer und der Aufwand der Schätzung, gegenüberstellt. Um die Genauigkeit zu vergleichen, wurde die weniger verbreitete "predictive likelihood" verwendet, im Gegensatz zu traditionellen Maßen wir MSE oder MAE. Das Ergebnis wägt diese beiden Eigenschaften gegeneinander ab und umschreibt Szenarien, in denen das Eine oder das Andere Modell besser geeignet ist.