DELFT UNIVERSITY OF TECHNOLOGY

FACULTY EEMSC

Thesis for the degree of MSc in Applied Mathematics in specialization of Financial Engineering

Modelling stochastic volatility via factor models: An asymmetric approach

By:

Tim Vogel

4486293

Supervised by:

Dr. N PAROLYA

Thesis committee:

Prof. Dr. A PAPAPANTOLEON

Dr. N PAROLYA

To be defended publicly at: October 19, 2022

ABSTRACT

In this thesis, a factor model which estimates multivariate time series is extended to include an asymmetric relation between the returns of assets and the volatility of said assets. The model proposed in this thesis uses the classical factor model, with univariate logarithmic volatility equations to model the factors as well as the asset innovations. The volatility equations for the factors are extended to contain an asymmetric relationship with the factor returns of the day before. In this thesis, a method to estimate this asymmetric model is developed, the method of estimation mainly relies upon MCMC methods. A method to estimate the logarithmic likelihood for the model is provided as well. This method uses a particle filter to estimate the distribution of the volatility. Using the logarithmic likelihood, it is shown that the asymmetry in the data is identifiable, by comparing the likelihood of the model to the likelihood of the classical factor model, as well as to the likelihood of a factor model with a jump extension. Finally, the model is tested on a real data set of daily returns where its effectiveness is again compared to the classical model and the jump model.

PREFACE

Dear reader,

This report is written as part of my thesis project for the master of Applied Mathematics at the Delft University of Technology, for the Financial Engineering specialization.

Looking back on the time I worked on this project, I look back on a time in which I have learned a lot on a variety of topics (some of these being: Monte Carlo Markov Chains, stochastic volatility, multi-factor models, Kalman filters, particle filters, and many more). I can categorize the duration of the master thesis into several instructive stages. During the first stage, the focus was on understanding the literature and recreating existing models. This process, although it took a bit longer than I hoped, was a very fruitful period of increasing my understanding of the topic. It is a good feeling that I can now understand a paper instantly, from which I know that it would have taken me a long time to understand less than a year ago. During the second stage, the focus shifted to developing an extension, although many extensions can be found, I elected to propose an extension that can be observed in market data. Finally, during the last stage, the model is tested. During these tests, I often found I missed an important detail when setting up the model. This is a result of working with quite a large model. Unfortunately, most of the errors are learned the hard way, when analyzing the results after a week's worth of simulation. At the time, that could be quite frustrating. However, in the end, when I could see that the model was able to beat some of the existing models, the effort proved to be worth it,

During this project, I have had frequent support from my supervisor, Dr. Nestor Parolya, who I would like to thank. Especially for handing me this project and our meetings, in which we discussed the problems I encountered and in which he often pushed me in the right direction. I also would like to thank Prof. Dr. Antonis Papapantoleon for taking the time for being on the thesis committee and evaluating my performance. Finally, I would like to thank the TU Delft HPC Cluster team, who allowed me to do many more and much bigger simulations than I would be able to do on my own laptop.

Sincerely,

Tim Vogel

Delft, October 11, 2022

CONTENTS

INTRODUCTION

1.1. MODELING VOLATILITY

Volatility is a key concept in finance, and its modeling is of great importance in many fields as option pricing, for example. Currently, two main approaches to modeling volatility can be considered dominant: (i) generalized auto-regressive conditional heteroskedasticity (GARCH) modeling and (ii) stochastic volatility (SV) modeling.

GARCH models [Bollerslev, 1986] are a type of statistical model used to predict future volatility in financial markets. The model is based on the assumption that past volatility is a good predictor of future volatility. A GARCH(p,q) model, assumes that the current volatility depends on the past q returns and the p last volatilities.

SV models assume that the volatility of asset prices fluctuates over time as well. The modeling, however, is taken in a different direction: the volatility itself is assumed to follow a certain distribution. When this distribution is known, sampling from it can give an indication of future volatility, which can, for example, be utilized by risk managers.

Both GARCH models, as well as SV models, are generally considered to be better than models with constant volatility. Since they can more accurately capture the observed changes in volatility over time. When comparing GARCH models to SV models, SV models are generally better able to capture the time-varying volatility of a time series, while GARCH models are better at capturing the auto-correlation of a time series. An additional advantage of GARCH models is that they are generally easier to estimate.

There are some reasons why it is good to capture the time-varying volatility of a time series. For example, a volatile stock is typically seen as a riskier investment than a less volatile stock. Moreover, capturing the time-varying volatility of a time series can help to improve the forecasting distribution of given time series. For this reason, in this thesis SV models will be the focal point.

1.2. MOVING TO MULTIVARIATE APPROACHES

In some areas, such as risk management or portfolio analysis, univariate models are not sufficient. In these areas, models in which assets are modeled together are necessary. In [Chib et al., 2006] a Multivariate Stochastic Volatility (MSV) factor model is proposed. In their paper, they successfully display a framework for high-dimensional MSV models. The attractiveness of MSV models can be seen, because they can capture the cross-correlations among different assets' volatilities, and they can be used to study the dynamics of volatility clusters, not only within a time series but also across different assets.

By using factors, they are able to scale back the number of parameters necessary to estimate the MSV model from order p^2 to order pk, where p is the amount of assets in the model, and k is the number of factors chosen. Since k is necessarily smaller than p, and often by a factor of 5 or more, a huge improvement in the number of parameters to estimate is seen.

For extreme events, [Chib et al., 2006] use a jump model extension (MSVJ). This extension explains the extremes quite well, however, it introduces two new problems: (i) it introduces a lot of new parameters and latent variables, and (ii) for medium to small-sized time series, the model heavily depends on the prior. It is shown that about 7500 data points (30-40 years of daily returns) are necessary to estimate the parameters sufficiently accurate.

1.3. ASYMMETRY IN SV MODELS

It is generally accepted that there is a relation between asset returns and the volatility of said asset, as stated by [Christie, 1982]: "Historically a variance/stock price relation is part of market folklore, the usual claim being that the relation is a negative one." A possible explanation given is the leverage effect within listed companies, but another possible explanation can be found in traders (both retail and institutional) trading with leverage, where liquidations and stop-losses will definitely increase volatility in worst-case scenarios.

Some work on modeling asymmetry in both GARCH models, as well as SV models has already been done, for example by: [Nelson, 1991], [Glosten et al., 1993], [Harvey and Shephard, 1996], [Asai and McAleer, 2011], or [Tsiotas, 2012]. However, asymmetry in MSV models has been the topic of research significantly less. Only a few papers can be found on this topic, one example is [Asai and McAleer, 2006]. Moving from MSV models to MSV factor models, none can be found.

In this thesis, the aim is to extend the factor model from [Chib et al., 2006] to include the asymmetric relation between asset returns and volatility. Moreover, the goal is to model this asymmetry within the common factors, making the market asymmetric as a whole and not on a per-asset basis. The main goal is to create an extension that performs better than the classical factor model, and that can compete with the more complex jump model.

The thesis is organized in the following way: In Section 2 the main topics frequently returning within this thesis are explained. In Section 3, it is shown that asymmetry in the data is more common in indices than within stocks, an indication that the market is asymmetric altogether instead of on a per asset basis. Next, in Section 4 the literature surrounding asymmetric models is reviewed and the selection process of the model is shown. In section 5 the model is fully specified and a way to estimate its parameters is given. In Section 6 a way to compare different models (including the model developed in this thesis) against each other is explained. In Section 7 it is shown that the parameters of the model can be estimated and that the model can be distinguished from the classical factor model as well as the factor model with jump extension. In Section 8 the model is tested on real-world equity returns, and finally, in Section 9 the results are discussed.

BACKGROUND

2.1. STOCHASTIC VOLATILITY

In this section, some frequently returning topics in this paper are covered, starting with stochastic volatility (SV).

Using an SV model, it is assumed that the return of an asset, y_t, can be modelled by:

$$y_t \sim N(\mu_y, \exp(h_t)) \tag{1}$$

With $t \in \{1, \ldots, n\}$. Without loss of generality, it can be assumed that $\mu_y = 0$. Within this equation, h_t can be considered the stochastic logarithmic volatility. For a more general introduction to Stochastic Volatility, visit chapter 5 of [Ghysels et al., 1996]. Often (as in [Chib et al., 2002]), h_t is assumed to be an AR(1) process:

$$h_t - \mu_h = \mu_h + \phi_h(h_{t-1} - \mu_h) + v_t \tag{2}$$

Where $v_t \sim N(0, \sigma_h^2)$. In this case, h depends on three latent variables: μ_h, ϕ_h, and σ_h. Each of these variables can be interpreted:

- μ_h is the long-term average of the logarithmic volatility.
- ϕ_h is the rate at which the logarithmic volatility returns to its long-term average. When ϕ_h is close to 0, h_t will return to μ_h relatively fast. On the other hand, when ϕ_h is close to 1, h_t tends to drift away from the mean more often and for longer periods.
- σ_h is the volatility of the noise added to the log-volatility h.

A few paths of h_t and the corresponding y_t are shown in Figure 1. In these pictures, it is visible what the volatility does. In times of higher volatility, returns are distributed more widely. It can also be seen that when modeling with equation 2, volatility tends to cluster.

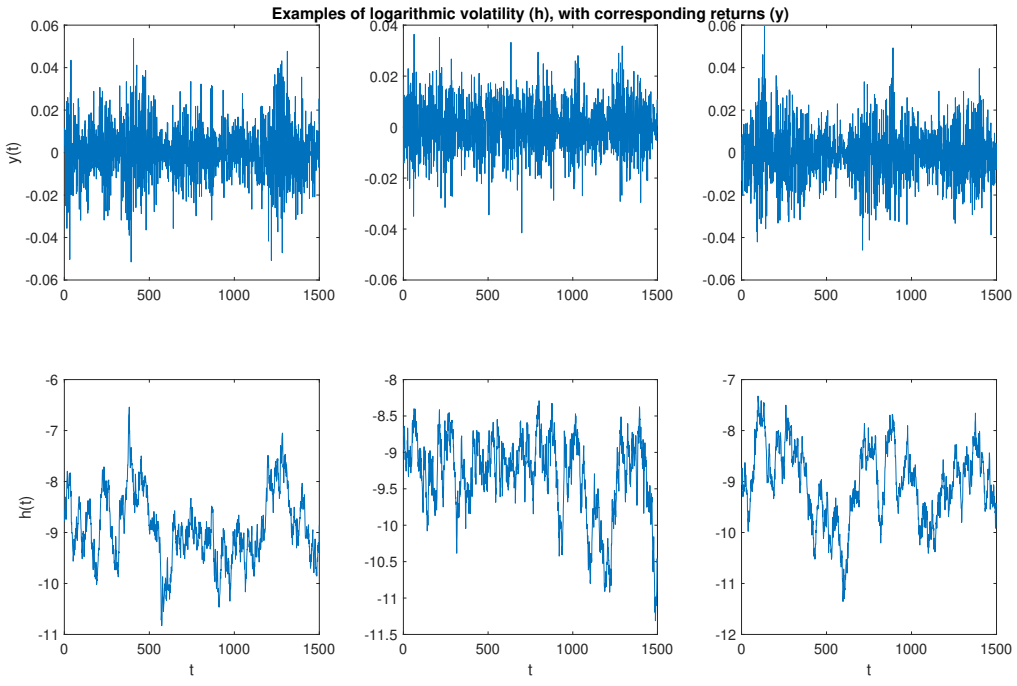

Figure 1: Three examples of logarithmic volatility with corresponding returns.

2.2. FACTOR MODELS

In fields where multiple assets should be modeled simultaneously, such as in portfolio or risk management, it does not suffice to use the model in 2.1 for each asset independently. Instead, a factor model, in which various assets get simulated simultaneously can be used [Chib et al., 2006]:

$$y_t = B f_t + u_t \tag{3}$$

Where $y_t = \{y_{1t}, y_{2t}, \ldots, y_{pt}\}$ are the p observations at time t, and again $t \in \{1, \ldots, n\}$. B is the factor loading matrix of size p by k, with the restrictions $b_{ii} = 1$ for $i \leq k$ and $b_{ij} = 0$ for $j > i$. So B contains ones on the main diagonal, zeros above this diagonal and free elements below the main diagonal. In equation 4, an example of B is provided, in this example $\beta_{i,j}$ are the free elements of B.

$$B = \begin{bmatrix} 1 & & & \\ \beta_{11} & 1 & & \\ \beta_{12} & \beta_{22} & \ddots & \\ \vdots & \vdots & & 1 \\ \beta_{1,k+1} & \beta_{2,k+1} & \ldots & \beta_{k,k+1} \\ \vdots & \vdots & & \vdots \\ \beta_{1p} & \beta_{2p} & \ldots & \beta_{k,p} \end{bmatrix} \tag{4}$$

Continuing the explanation of equation 3, $f_t = \{f_{1t}, f_{2t}, \ldots, f_{kt}\}$ are the unobserved factors. And finally, $u_t = \{u_{1t}, u_{2t}, \ldots, u_{pt}\}$ are the innovations. For the innovations, the following model is used:

$$u_{jt} \sim N(0, \exp(h_{jt})) \tag{5}$$

Where h_{jt} is the log-volatility following equation 2 (or another logarithmic volatility model). For the factors, a similar equation is defined:

$$f_{jt} \sim N(0, \exp(h_{p+j,t})) \tag{6}$$

So both u_{jt} as well as f_{jt} are functions of the log-volatility h_{jt}, however since h_{jt} is modelled independently for each j, no correlations from this part should be expected.

Summarizing, for each time step, this model simulates a given amount of factors. Using the factor loading matrix, it then projects these factors upon all the assets. Finally, it adds individual noise terms to all the assets.

As can be imagined, there are a lot of possibilities within this category of models. For example, the normal distributed innovations in equation 5 can be replaced by different distributions (e.g. a student t distribution), a jump component can be added to equation 3, or the logarithmic volatility in equation 2 can be modelled in a different way, as will be done in this paper.

2.3. MCMC SAMPLING

The method used in this paper for estimating the model's parameters is called Markov Chain Monte Carlo sampling, or MCMC sampling. Using MCMC sampling, the posterior distribution is estimated by simulating samples. These samples are used to create a posterior distribution of the parameters used in the model. The way this is accomplished is by constructing an irreducible and aperiodic Markov Chain, whose invariant distribution is the target posterior distribution.

The Monte Carlo part comes into play, by drawing from the Markov Chain multiple times. By doing this, an estimate of the posterior distribution is obtained. However, due to the size of the parameters and latent variables of the model, it becomes infeasible to simulate the entire parameter space at the same time. Instead, a blocking scheme should be constructed. As discussed in [Chib and Greenberg, 1995], it is preferred to group certain latent parameters and variables. For example, in the factor model, from Section 2.2, the following blocking scheme could be used:

$$[B|Y_n, \{h_{j.}\}, f_t] \qquad [f_t|Y_n, B, \{h_{j.}\}] \qquad [\Theta|Y_n, B, f_t] \qquad [h_{jt}|y, B, f_t, \Theta, \{h_{js}\}(s \neq t)]$$

Where $Y_n = [y_1, \ldots, y_n]$ and $h_{j.}$ is notation for the collection $(h_{j1}, \ldots, h_{jn})$. That is, simulate the factor loading matrix, then the factors, then the parameters for the log-volatility and last the log-volatility one time step each block. Each sampling conditioned on the current values of all the other parameters and variables.

In [Chib and Greenberg, 1995] it is explained that faster mixing occurs if all $\{h_{j.}\}$ get sampled simultaneously. Moreover, in [Chib et al., 2006] it is shown that it is more efficient to sample B and $\{f_t\}$ in one block as well. The reason for this is that they only appear in product form in the model. A preferred blocking scheme for the factor model would therefore be:

$$[B, f_t | Y_n, \{h_{j.}\}] \qquad\qquad [\Theta, \{h_{j.}\} | Y_n, B, f_t]$$

Using draws from this sampler, the distribution of B and θ can be found.

2.4. Metropolis-Hastings Algorithm

One recurring step in the MCMC algorithm is the Metropolis-Hastings algorithm [Chib and Greenberg, 1995]. The Metropolis-Hastings algorithm (MH Algorithm), is a useful and intuitive tool to simulate samples from multivariate distributions. The method is used to sample from a posterior distribution during an MCMC method.

The goal of the MH algorithm is to sample from a target distribution $\pi(\theta|y) \propto \pi(\theta) f(y|\theta)$, given a data-set y. Since an MCMC method is used, the current state of θ, θ_c is known as well. Using the MH algorithm, a transition kernel $p(\theta_c, \theta^*)$ is required. In [Chib and Greenberg, 1995] it is explained that a sufficient condition for such a function is the reversibility condition (i.e. $p(\theta_c, \theta^*) = p(\theta^*, \theta_c)$).

First, a candidate θ^* is drawn from a candidate generating density $q(\theta_c, \theta^*)$ (which may depend on the current state of the algorithm as well as the data set). It may be the case that:

$$\pi(\theta_c) q(\theta_c, \theta^*) > \pi(\theta^*) q(\theta^*, \theta_c) \tag{7}$$

To comply with the reversibility condition, the left-hand side should be equal to the right-hand side. To accomplish this, the acceptance probability is introduced:

$$\alpha(\theta_c, \theta^*) = \min\left(1, \frac{\pi(\theta^*) q(\theta^*, \theta_c)}{\pi(\theta_c) q(\theta_c, \theta^*)}\right) \tag{8}$$

Now with probability $\alpha(\theta, \theta^*)$, the candidate gets accepted, and with probability $1 - \alpha(\theta, \theta^*)$ it is rejected. In this case, the current value of θ (θ_c) is selected as the next value for θ in the Markov Chain. Similarly, if the inequality of equation 7 is the other way around:

$$\alpha(\theta^*, \theta_c) = \min\left(1, \frac{\pi(\theta_c) q(\theta_c, \theta^*)}{\pi(\theta^*) q(\theta^*, \theta_c)}\right) \tag{9}$$

This means that the following transition kernel, always complies with the reversibility condition:

$$p(\theta, \theta^*) = q(\theta, \theta^*) \alpha(\theta, \theta^*) \tag{10}$$

Below the MH algorithm can be found in pseudo-code:

1. Start iteration i with θ_i as current value.
2. Generate θ^* from $q(\theta_i, \theta^*)$.
3. Calculate $\alpha(\theta_i, \theta^*)$.
4. Generate u from $U(0,1)$.
5. If $u < \alpha(\theta_i, \theta^*)$, $\theta_{i+1} = \theta^*$
6. Otherwise: $\theta_{i+1} = \theta_i$
7. Repeat for $i = i + 1$

2.5. KALMAN FILTERS

One important part of the sampling algorithm when sampling the log-volatility equations, is the Kalman Filter. Consider the following state-space model:

$$\mathbf{x}_t = \phi \mathbf{x}_{t-1} + \Upsilon \mathbf{u}_t + \mathbf{w}_t \tag{11}$$

$$\mathbf{y}_t = A_t \mathbf{x}_t + \Gamma \mathbf{u}_t + \mathbf{v}_t \tag{12}$$

With $\mathbf{x}_t$ the $k \times 1$ state vector. $\mathbf{x}_t$ is not observed directly. However, the transformed $p \times 1$ vector $\mathbf{y}_t$ is observed. A_t is the $p \times k$ observation matrix. ϕ is an $k \times k$ auto-regressive matrix, $\mathbf{u}_t$ is a $r \times 1$ vector of inputs, with Υ ($k \times r$) and Γ ($p \times r$) the matrices which define how $\mathbf{x}_t$ and $\mathbf{y}_t$ react to the input. $\mathbf{w}_t$ and $\mathbf{v}_t$ are the state and observation noise, with variance matrices respectively Q ($k \times k$) and R ($p \times p$).

Next, the Kalman Filter is presented. The filter takes the values $y_1, \ldots, y_t$ as input and gives the best estimate for x_t. This, in turn, can be exploited to compute certain likelihoods. For some starting values $x_0^0 = \mu$ and $P_0^0 = \Sigma_0$, the Kalman filter starts with the "Model update equations":

$$\mathbf{x}_t^{t-1} = \phi \mathbf{x}_{t-1}^{t-1} + \Upsilon \mathbf{u}_t \tag{13}$$

$$P_t^{t-1} = \phi P_{t-1}^{t-1} \phi' + Q \tag{14}$$

Where $\mathbf{x}_t^{t-1} = E[x_t | y_1, \ldots, y_{t-1}]$, $\mathbf{x}_t^t = E[x_t | y_1, \ldots, y_t]$. $P_t^{t-1} = E[(\mathbf{x}_t - \mathbf{x}_t^{t-1})(\mathbf{x}_t - \mathbf{x}_t^{t-1})']$ and $P_t^t = \text{Cov}(x_t | y_1, \ldots, y_t)$. Next are the "Data correction equations":

$$K_t = P_t^{t-1} A_t' (A_t P_t^{t-1} A_t' + R)^{-1} \tag{15}$$

$$\mathbf{x}_t^t = \mathbf{x}_t^{t-1} + K_t (\mathbf{y}_t - A_t \mathbf{x}_t^{t-1} - \Gamma \mathbf{u}_t) \tag{16}$$

$$P_t^t = (I - K_t A_t) P_t^{t-1} \tag{17}$$

Here K_t is called the Kalman gain. Two important products of the Kalman Filter are the innovation and its variance:

$$\epsilon_t = \mathbf{y}_t - A_t \mathbf{x}_t^{t-1} - \Gamma \mathbf{u}_t \tag{18}$$

$$\Sigma_t = A_t P_t^{t-1} A_t' + R \tag{19}$$

These quantities can be used to produce the log-likelihood, l, which is an important value when sampling.

$$l = -\frac{1}{2} \sum_{t=1}^{n} \log |\Sigma_t| - \frac{1}{2} \sum_{t=1}^{n} \epsilon_t' \Sigma_t^{-1} \epsilon_t \tag{20}$$

A proof of the Kalman Filter, together with a chapter-length review of theory and various applications can be found in Chapter 6 of [Shumway and Stoffer, 2000].

KALMAN FILTER APPLIED TO STOCHASTIC VOLATILITY MODELS.

Since the log-volatilities in the factor model (equation 3) all operate independently, they can be separated per series. The state equation (equation 11) is just equation 2. The space equation (equation 12) can be determined by:

$$\begin{aligned} u_t &\sim N(0, \exp(h_t)) \\ u_t &\sim \exp(\frac{h_t}{2}) N(0,1) \\ u_t^2 &\sim \exp(h_t) \chi_1^2 \\ \log(u_t^2) &\sim h_t + \log(\chi_1^2) \\ z_t &= h_t + v_t \end{aligned} \tag{21}$$

Where $z_t = \log(u_t^2)$ and v_t is a log chi-squared random variable. Since the Kalman Filter requires the random variable to be Gaussian, a seven-component mixture of Gaussians is used to approximate the log chi-squared random variable, all parameters from these Gaussians can be found in Table 1 (taken from [Chib et al., 2002]). The equation to estimate the log chi-squared using these parameters becomes:

$$\log(\chi_1^2) \approx \sum_{i=1}^{7} q_i N(m_i, v_i^2) \tag{22}$$

Component (s_{tj})	Probability (q_i)	Mean (m_i)	Variance (v_i^2)
1	0.00730	-11.40039	5.79596
2	0.10556	-5.24321	2.61369
3	0.00002	-9.83726	5.17950
4	0.04395	1.50746	0.16735
5	0.34001	-0.65098	0.64009
6	0.24566	0.52478	0.34023
7	0.25750	-2.35859	1.26261

Table 1: The parameters of a mixture of seven Gaussian distributions to approximate a log chi-squared distribution.

The Kalman Filter can now be implemented (the subscript j is dropped for the ease of notation):

$$h_t^{t-1} = \mu + \phi(h_{t-1}^{t-1} - \mu) \tag{23}$$
$$P_t^{t-1} = \phi^2 P_{t-1}^{t-1} + \sigma^2 \tag{24}$$
$$K_t = \frac{P_t^{t-1}}{P_t^{t-1} + v_{s_t}^2} \tag{25}$$
$$h_t^t = h_t^{t-1} + K_t(z_t - h_t^{t-1} - m_{s_t}) \tag{26}$$
$$P_t^t = (1 - K_t)P_t^{t-1} \tag{27}$$

And the quantities to determine the likelihood become:

$$\epsilon_t = z_t - h_t^{t-1} - m_{s_t} \tag{28}$$
$$\Sigma_t = P_t^{t-1} + v_{s_t}^2 \tag{29}$$

2.6. Extended Kalman Filters

Extended Kalman Filters (EKFs) can be used when (i) the equations are non-linear or (ii) the random variables are non-Gaussian. However, as discovered within this thesis, the theory surrounding EKFs is not as wide covering as the theory surrounding KFs.

In this section, the main differences between the KF and the EKF will be highlighted. First the equation $f(x)$ is defined such that:

$$\mathbf{x}_t = f(\mathbf{x}_{t-1}, \mathbf{u}_t) + \mathbf{w}_t \tag{30}$$

This equation should replace equation 11 in the state space model. That is, instead of only allowing for linear transformations of $\mathbf{x}_t$ and $\mathbf{u}_t$, any transformation is allowed.

The Kalman Filter, in turn, needs to be updated as well. This can be done by defining new "model update equations":

$$\mathbf{x}_t^{t-1} = f(\mathbf{x}_{t-1}^{t-1}, \mathbf{u}_t) \tag{31}$$
$$P_t^{t-1} = \left(\frac{\delta f}{\delta x}(\mathbf{x}_{t-1}^{t-1}, \mathbf{u}_t)\right)^2 P_{t-1}^{t-1} + \sigma^2 \tag{32}$$

The data correction equations remain the same.

Some problems using the extended Kalman filter can be encountered. Unlike its linear version (the KF), the EKF is generally not the optimal estimator. Additionally, because it linearizes the function f from equation 30, the filter may diverge if the initial estimate is too far off. Another problem using EKFs is that the sampled variance tends to underestimate the real variance [Huang et al., 2008]. Despite these disadvantages, the EKF is still a well-performing algorithm. For example, it is used in state estimation in applications such as navigation or GPS systems [Wan, 2006].

ASYMMETRY IN DATA

3.1. ASYMMETRY IN INDICES

In this section, the asymmetric relation between asset returns and volatility within markets will be analyzed. First an analysis of the asymmetry in index returns is made. In Section 3.2 a similar analysis is made on stock returns.

The following indices (including abbreviations) are selected:

- S&P 500 Index (SP500)
- UK 100 Index (UK100)
- DAX Index (DAX)
- Nikkei 225 Index (Nikkei)

This way, indices from three different regions in the world are analyzed: Asia, North America, and Europe. Please note that all four indices have a different base currency. In this analysis, all prices will be analyzed in the US Dollar.

The indices are analyzed daily, starting from 01/01/2011 up until 01/01/2021. First, the mean gets removed from the daily returns. Thereafter, the data is split up into two groups:

$$Y_t^+ = I_{[y_{t-1}>0]} \tag{33}$$

$$Y_t^- = I_{[y_{t-1}<0]} \tag{34}$$

Where y_t is the asset return. These variables indicate whether the return of the day before is either positive or negative. Next, the mean absolute return of returns after a positive day $\bar{r}^+$ and the mean absolute return of returns after a negative day $\bar{r}^-$ are defined:

$$\bar{r}^+ = \frac{1}{\sum_{t=1}^n Y_t^+} \sum_{t=1}^n |y_t| Y_t^+ \tag{35}$$

$$\bar{r}^- = \frac{1}{\sum_{t=1}^n Y_t^-} \sum_{t=1}^n |y_t| Y_t^- \tag{36}$$

The quantity of interest, is the ratio between these two:

$$R = \frac{\bar{r}^-}{\bar{r}^+} \tag{37}$$

Where $R > 1$ indicates that absolute returns are typically higher the day after a negative return, and $R < 1$ indicates the inverse. The uncertainty in R is also measured:

$$\sigma_- = \frac{1}{\sqrt{\sum_{t=1}^n Y_t^-}} \sum_{t=1}^n Y_t^- \left(|y_t| Y_t^- - \bar{r}^-\right)^2 \tag{38}$$

$$\sigma_+ = \frac{1}{\sqrt{\sum_{t=1}^n Y_t^+}} \sum_{t=1}^n Y_t^+ \left(|y_t| Y_t^+ - \bar{r}^+\right)^2 \tag{39}$$

$$\sigma_R = \frac{1}{\bar{r}^+} \left(\sigma_- + R\sigma_+\right) \tag{40}$$

The results of equation 35, 36 and 37 for the selected indices are displayed in table 2[1].

In table 2 it can be seen that R is about $2\sigma_R$ above 1 for all indices. This is the first indication that there is some kind of asymmetry between returns after positive and after negative asset returns the day before.

[1] If the indices are not corrected for currency and would be expressed in their base currency, the effect would become even more visible. The values for R would be 1.188, 1.212, and 1.184 for UK100, DAX, and N225 respectively.

Index	$\bar{r}^+$	$\bar{r}^-$	R	σ_R
SP500	0.0065	0.0079	1.219	0.089
UK100	0.0083	0.0093	1.131	0.068
DAX	0.0097	0.0114	1.180	0.071
N225	0.0085	0.0097	1.136	0.069

Table 2: The mean absolute return for days after a positive return and the mean absolute return for days after a negative return for four selected indices. The ratio between the two and the standard deviation of the ratio is displayed as well.

Next, a KS-test is performed [Dodge, 2008]. For this test, the following data-sets are defined:

$$r_t^+ = |y_t| Y_t^+ \tag{41}$$

$$r_t^- = |y_t| Y_t^- \tag{42}$$

Where all zero contributions will be deleted from each set. In Figure 2 the empirical distribution functions for r^+ and r^- are shown, and in table 3 the results of the KS-test are displayed.

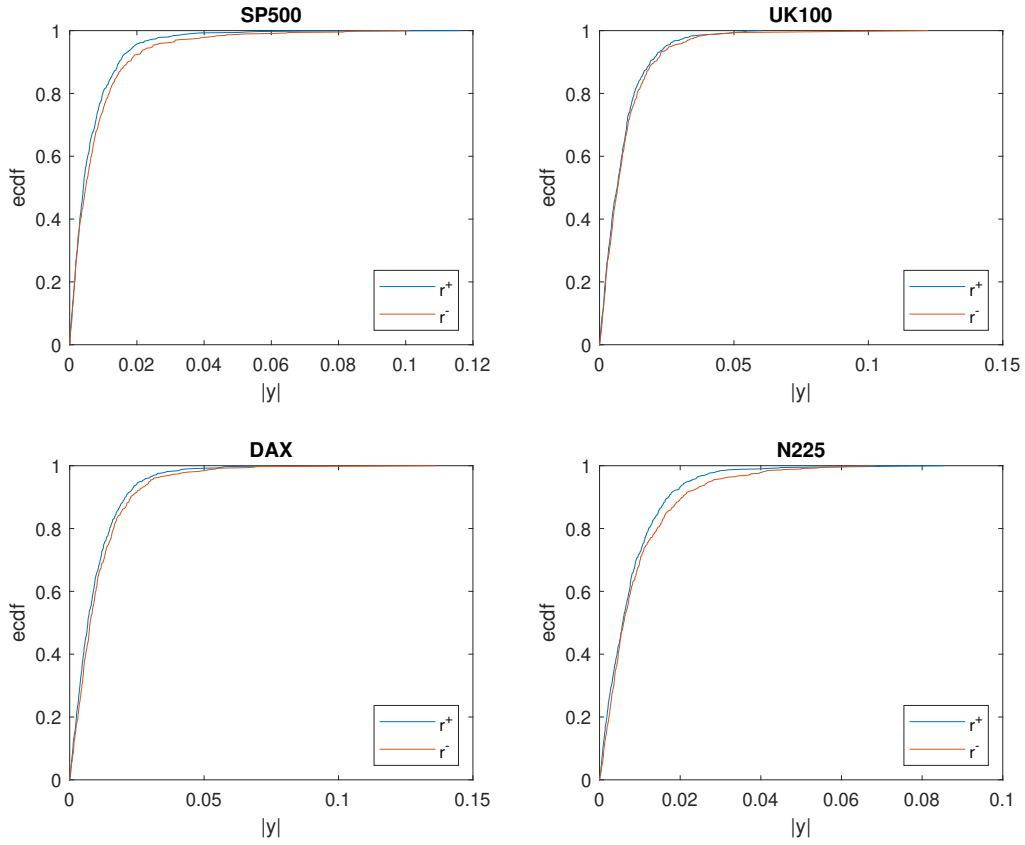

Figure 2: The empirical cumulative distribution functions for the absolute returns the day after a positive return and the day after a negative return.

As can be seen in Table 3, only for the UK100 Index, based on the KS-test one is not able to reject the hypothesis that r^+ and r^- come from the same distribution at the $\alpha = 0.05$ level. In Figure 2, it can be seen that the empirical distribution of r^- tends to be distributed higher for all four indices.

Index	p	h
SP500	0.0003	1
UK100	0.0562	0
DAX	0.0153	1
N225	0.0128	1

Table 3: The results of the KS test, comparing the distributions r^+ and r^-. Displayed are the p-values, as well as h, which indicates whether the hypothesis that r^+ and r^- come from the same distribution can be rejected at the $\alpha = 0.05$ level.

3.2. ASYMMETRY IN STOCKS

For returns on stocks, a similar analysis is performed. This time, some stocks from the SP500 are selected. In Section 3.1, the asymmetry was the most present in this index. Two top 10 ranked stocks (ranked by weight), two stocks ranked around 100 and two stocks ranked around 250 are selected at random.

- Apple Inc. (AAPL)
- Microsoft Corporation (MSFT)
- Northrop Grumman Corp. (NOC)
- Progressive Corporation (PGR)
- First Republic Bank (FRC)
- Keurig Dr Pepper Inc. (KDP)

The same analysis as was performed to the indices is now performed to the returns of the stocks. The results are shown in Table 4.

Stock	$\bar{r}^+$	$\bar{r}^-$	R	σ_R
AAPL	0.0116	0.0132	1.140	0.067
MSFT	0.0103	0.0115	1.111	0.069
NOC	0.0094	0.0103	1.094	0.065
PGR	0.0090	0.0096	1.071	0.064
FRC	0.0112	0.0120	1.067	0.066
KDP	0.0087	0.0091	1.046	0.072

Table 4: The mean absolute return for days after a positive return and the mean absolute return for days after a negative return for six selected stocks. The ratio between the two and the standard deviation of the ratio is displayed as well.

In Table 4, it can be seen that for the higher-ranked stocks (AAPL and MSFT), the asymmetry is still visible, albeit less significant than for the indices. For lower ranked stocks, there still are some signs of asymmetry, however, this time the value of R is closer to 1 (the no asymmetry case), in most cases the distance from R to 1 is slightly above σ_R

The KS-test has been performed again as well, the results can be found in Figure 3 and Table 5.

Index	p	h
AAPL	0.0025	1
MSFT	0.0389	1
NOC	0.0417	1
PGR	0.6594	0
FRC	0.3056	0
KDP	0.7006	0

Table 5: The results of the KS-test, comparing the distributions r^+ and r^-. Displayed are the p-values, as well as h, which indicates whether the hypothesis that r^+ and r^- come from the same distribution can be rejected at the $\alpha = 0.05$ level.

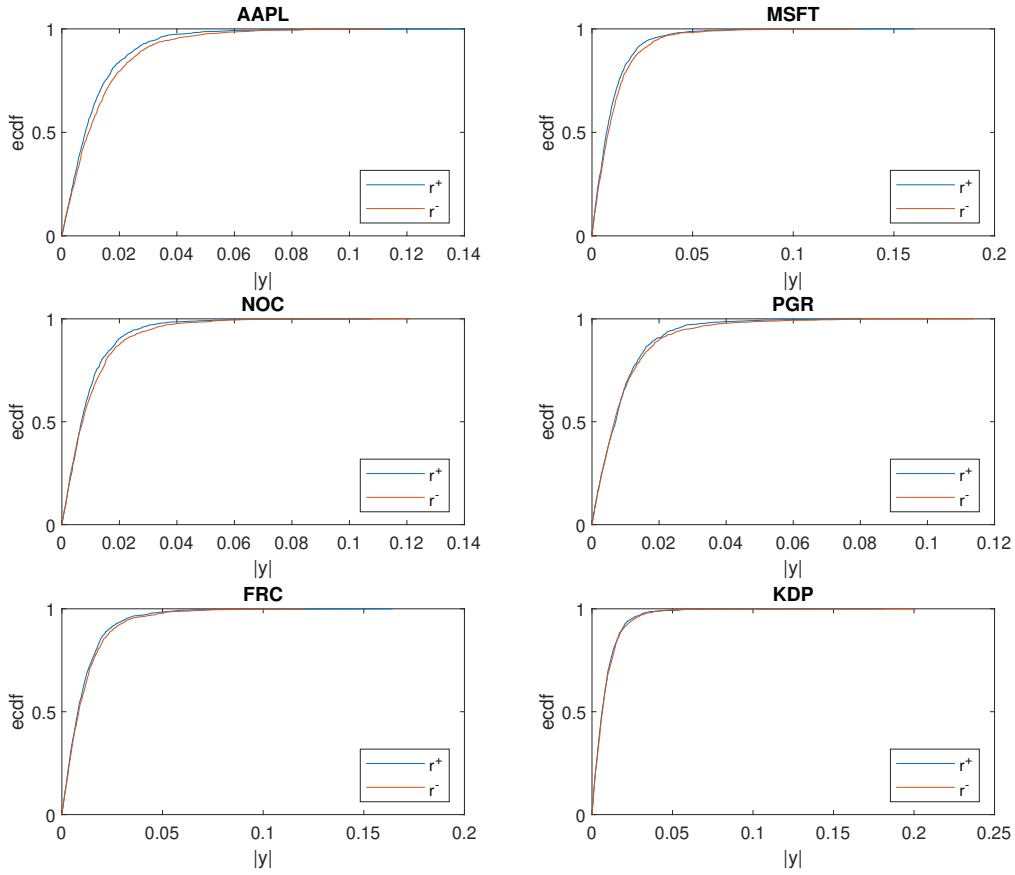

Figure 3: The empirical cumulative distribution functions for the absolute returns the day after a positive return and the day after a negative return.

In Figure 3, it is visible that, with the exception of AAPL, the distribution of r^+ and r^- are more similar to each other, compared with Figure 2. This is confirmed in Table 5, where is shown that, based on the KS-test, one cannot reject that the distribution of r^+ and r^- are different three out of six times. Also note that the p-values for PGR, FRC, and KDP are not even close to being rejected (for any reasonable value of α).

3.3. Integrating results in the factor model

In Section 3.1 it was shown that there are indications that indices show an asymmetric relation between volatility and returns. More specifically, following a negative return, typically a bigger absolute return is seen the day after. This phenomenon might also be present in stocks, however, to a much lesser extent, as shown in sector 3.2.

Since indices are a weighted average of many stocks, some of the common behaviour of the stocks is explained by the index, while each singular stock also has an idiosyncratic piece of its returns. When looking at the factor model equation (explained in Section 2.2), the factors (f_t) can be linked to this common behaviour, while the innovations (u_t) can be compared to the idiosyncratic return.

The way the conclusion from the two previous sections is integrated into the factor models, is by making the factors asymmetric while keeping the innovations non-asymmetric. This way, (i) the indices will become asymmetric, and (ii) individual stocks might also display asymmetry in returns, depending on the values of the factor loadings in the factor loading matrix (B) (i.e. large factor loadings will lead to more asymmetry in the singular asset returns, dampened out by their non-asymmetric innovations).

LITERATURE REVIEW AND MODEL SELECTION

4.1. EXISTING ASYMMETRIC MODELS

In this section, the selection of the model used in this paper will be explained. First the existing asymmetric models in the literature are reviewed. Next, in Sections 4.2 until 4.4 the process of selecting the model is explained.

The first category of asymmetric models looked into are the asymmetric GARCH models. A nice overview of asymmetric GARCH models can be found in [Chen et al., 2019]. Here they explore models from [Nelson, 1991], [Glosten et al., 1993], and [Ding et al., 1993]. Below these models will be listed. All models will make use of the same asset return equation:

$$y_t = \epsilon_t \tag{43}$$

$$\epsilon_t = \sqrt{h_t}\, v_t \tag{44}$$

With v_t a normally distributed random variable. The first asymmetric GARCH model considered, is the Glosten-Jagannathan-Runkle GARCH model (GJR-GARCH) otherwise known as the Threshold GARCH model. The reason it is known as the threshold Garch model is that it makes use of a threshold variable:

$$d_t = 1_{\epsilon_t < 0} \tag{45}$$

This variable indicates a negative return at time t. Making use of this variable, the following model for the volatility is used:

$$h_t = \mu + \sum_{j=1}^{m} \phi_i h_{t-i} + \sum_{i=1}^{q} \alpha_i \epsilon_{t-i}^2 + \gamma \epsilon_{t-1}^2 d_{t-1} \tag{46}$$

Here γ denotes the asymmetry parameter. As will be the case in all the models to follow.

Another way to asymmetrically model the volatility is the Asymmetric Power GARCH (APARCH) model. Instead of making use of an indicator, this model uses the absolute value to determine whether a return is negative or positive.

$$h_t^{\frac{\delta}{2}} = \mu + \sum_{j=1}^{m} \phi_i h_{t-1}^{\frac{\delta}{2}} + \sum_{i=1}^{q} \alpha_i (|\epsilon_{t-i}| - \gamma \epsilon_{t-i})^{\delta} \tag{47}$$

In this model, δ is the power term parameter. Note that for $\delta = 2$ and $\gamma = \pm 1$ the APARCH model and the GJR-GARCH model are the same.

The last GARCH model considered in this thesis is the Exponential GARCH (EGARCH) model. This model has an extra desired property, which is that the volatility is modeled at the logarithmic level:

$$\log(h_t) = \mu + \sum_{j=1}^{m} \phi_i \log(h_{t-1}) + \sum_{i=1}^{q} \frac{\alpha_i}{\sqrt{h_{t-1}}} (|\epsilon_{t-i}| + \gamma \epsilon_{t-i}) \tag{48}$$

Next, the existing asymmetric SV models are analyzed. In [Asai et al., 2006], an overview of some of the developed asymmetric SV models can be found. The univariate "Leverage Effect" model was first proposed by [Harvey and Shephard, 1996], with later adaptations or applications by among others: [Tsiotas, 2012] and [Asai and McAleer, 2011].

The leverage model is defined in the following way:

$$y_t = \exp\left(\frac{h_t}{2}\right) \epsilon_t \tag{49}$$

$$h_{t+1} = \mu + \phi \log(h_t) + \sigma \eta_t \tag{50}$$

Where $\mathrm{Cor}(\eta_t, \epsilon_t) = \gamma$. This model adds a correlation between the return y_t and the increase in logarithmic volatility h_{t+1}. When γ is negative, this model adds the asymmetric relation observed in Section 3.

[Danielsson, 1998], [Asai and McAleer, 2006] and [Chan et al., 2005] considered the multivariate asymmetric extension to the model discussed above.

$$y_t = H\epsilon_t \tag{51}$$

$$H_t = \text{Diag}\left(\exp\left(\frac{h_{1t}}{2}\right), \ldots, \exp\left(\frac{h_{pt}}{2}\right)\right) \tag{52}$$

$$h_{jt} = \mu_j + \phi_j h_{j,t-1} + \eta_t \tag{53}$$

$$\begin{bmatrix}\epsilon_t \\ \eta_t\end{bmatrix} = N\left(\begin{bmatrix}0\\0\end{bmatrix}, \begin{bmatrix}P_\epsilon & L \\ L & \Sigma_\eta\end{bmatrix}\right) \tag{54}$$

Within this model, P_ϵ is the covariance matrix between the various assets, and Σ_η is a positive definite covariance matrix. $L = \text{Diag}(\gamma_1, \ldots, \gamma_p)$ is the matrix that determines the asymmetry. Again for negative values of γ_i the relation observed in Section 3 is obtained.

Although the models above can't be copied one-to-one into the asymmetric factor model, they can be used for inspiration and as a starting point. The model from equation 51 until equation 54 is closest to the AMSV factor model. However, due to (i) the factors and innovations all being considered independent in the factor model, and (ii) the factors scaling with h, this model is not ideal. The EGARCH model from equation 48 has some advantages the factor model could make use of. One of these advantages is that the obtained values for ϵ are re-scaled by h. This makes sure that the model is stable since a larger return goes together with higher volatility. Another advantage is that the EGARCH model is univariate, this way no dependence relations between volatility or returns across different assets are added. Incorporating this model into the factor model, all correlations between assets are still modeled using only the factors.

4.2. EGARCH BASED SV MODEL

As discussed in the previous section, the EGARCH model from [Nelson, 1991], altered by [Chen et al., 2019], is taken as a starting point. The model will first be adapted to become in line with the log-volatility equation for the factors from equation 3.

$$h_{jt} = \mu_j + \phi_j(h_{j,t-1} - \mu_j) + \frac{\theta_{j-p}}{\exp(\frac{h_{j,t-1}}{2})}\left(|f_{j-p,t}| - \gamma_{j-p} f_{j-p,t} - \exp(\frac{h_{j,t-1}}{2})\sqrt{\frac{2}{\pi}}\right) + \sigma_j \eta_{jt} \tag{55}$$

Equation 55 is already quite different from the EGARCH model used in [Chen et al., 2019], for example:

- $m = q = 1$, only one lag is regarded.
- ϕ operates on $(h_{t-1} - \mu)$ instead of just h_{t-1}.
- The mean of $|f_{j-p,t-1}|$ is subtracted, so the added term will add an average of zero:

  $E[|f_{j-p,t}| - \gamma f_{j-p,t} - \exp(\frac{h_{j,t-1}}{2})\sqrt{\frac{2}{\pi}}] = 0$. This is also done in [Chen et al., 2019]; however, there it is done in a separate step.

It should also be said, that only the logarithmic volatility equations for the factors change, the equations for the innovations stay the same.

In Figure 4, the returns of various assets using the above model are shown (based on 3 factors). A first observation is that there are periods of higher volatility as well as periods of lower volatility. Another feature of this model is that these periods of high volatility tend to be correlated across assets (see Figure 7 where a comparison is made with the model from Section 4.3).

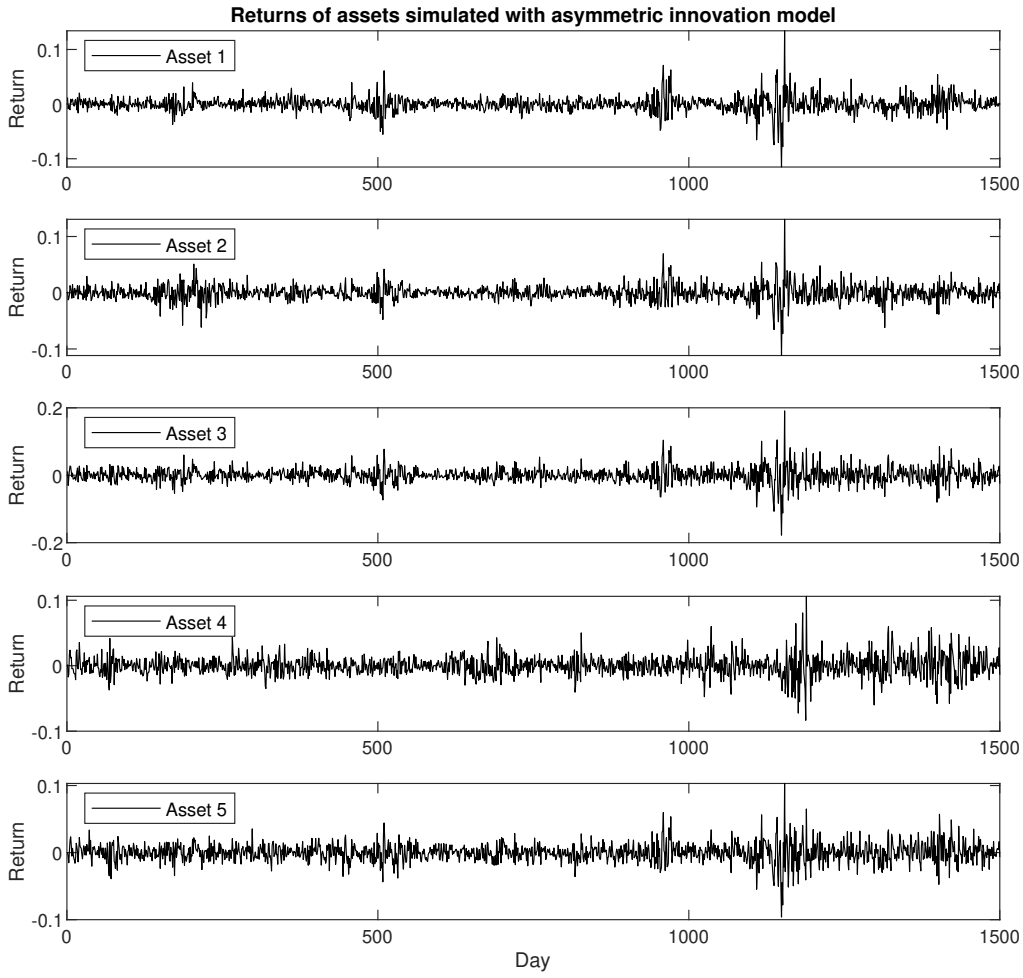

Figure 4: Returns of various assets simulated using the model explained in section 4.2.

Although the simulated data displays the desired behaviour, one major problem arises when trying to estimate the θ and γ parameters of this model. Contrary to the EGARCH model, where the model is a function of the observed returns, in this model the logarithmic volatility is a function of the unobserved factors. Since they are sampled, instead of known, additional uncertainty is added to the model.

To illustrate why this is a problem, Figure 5 is shown. Here, $(\hat{f}_t|y, B, \{h_{j.}\})$ is sampled and the results are compared with the actual factors f_t^*. In the bottom half of Figure 5, it is seen that $\hat{f}_t$ estimates f_t^* quite well. However, when looking at the log-likelihood of $(\theta_j|\{z_{j.}\}, \{s_{j.}\}, \{h_{j.}\}, f_t)$ for an arbitrary factor, it makes a huge difference whether the sampled factors $\hat{f}_j$ or the true factors f_j^* are used. This can be seen in the top half of Figure 5. In this figure, $[\mu_j, \phi_j, \sigma_j, \gamma_{j-p}]$ are fixed at $[\mu_j^*, \phi_j^*, \sigma_j^*, \gamma_{j-p}^*]$, while θ_{j-p} runs from 0 to 0.5.

In the top half of Figure 5, two observations can be made. First of all, using the true factors, the likelihood for all θ is higher. Second, and more importantly, the peak in log-likelihood for the estimated factors is not close to the true value, having an error of about 30%, while for the true factors, the estimate for θ is much closer to the true value (the error is less than 5%). This second problem leads to inaccurate estimates of θ, even when all other parameters are perfectly estimated.

Because of this, more alterations to the existing models will be made. In the next section, a version where the observed returns are used as feedback to h is considered.

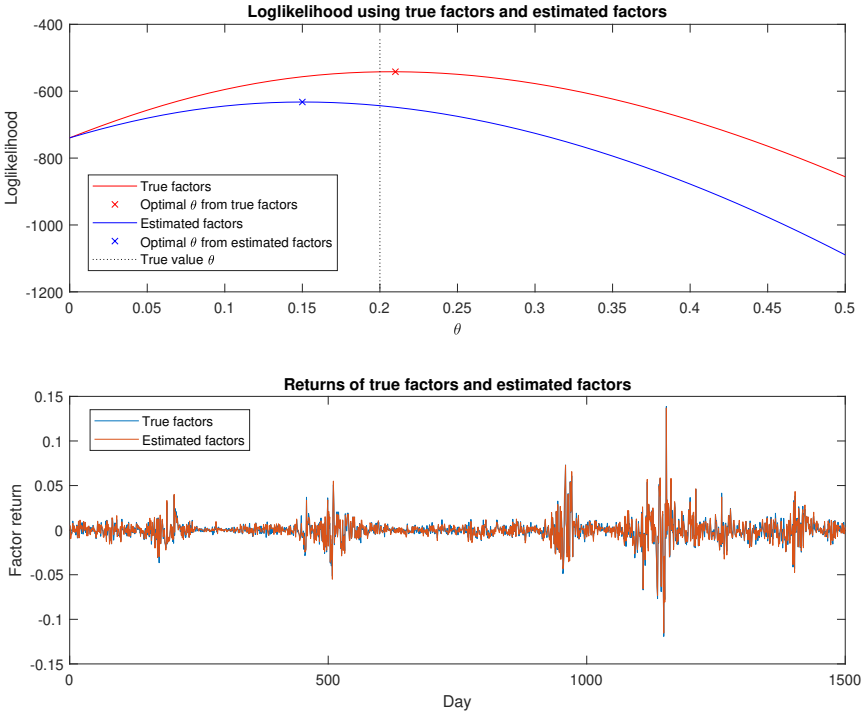

Figure 5: Top: A comparison of the log-likelihood of $(\mu_j^*, \phi_j^*, \sigma_j^*, \gamma_{j-p}^* \theta_{j-p} | \{z_{jt}\}, \{s_{jt}\}, \{h_{jt}\}, \{f_{jt}\})$ for one of the factors, where the results of the true values of f_{j-p} are compared with the results using the estimated factors $\tilde{f}_{j-p}$. Bottom: The true factor returns compared with the estimated factor returns.

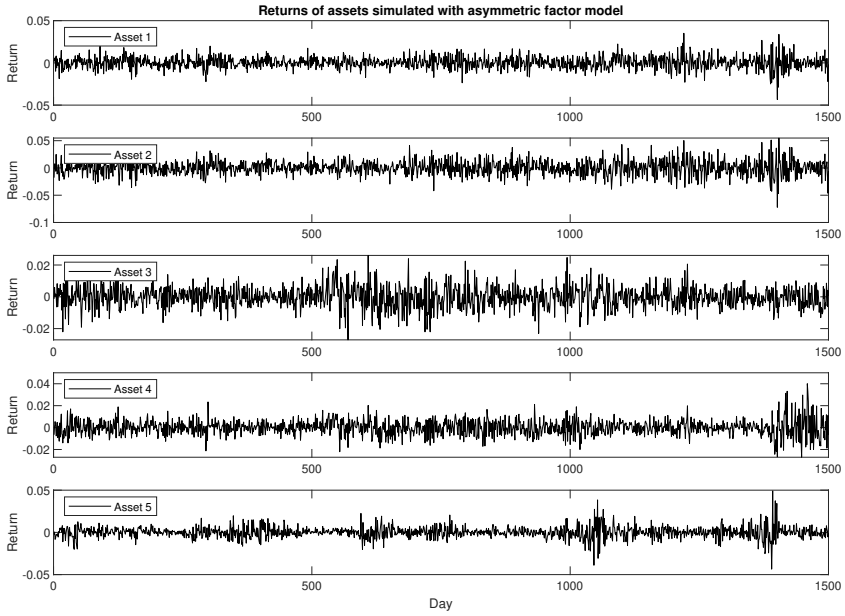

Figure 6: Returns of various assets simulated using the model explained in section 4.3

4.3. ASYMMETRIC INNOVATIONS

The main problem in Section 4.2 is that f_t is unobserved, this problem is avoided by using y_t as input and changing the logarithmic volatility equations for the innovations instead of the factors. First, a new vector d_j is defined, this vector contains the standard deviation for each asset:

$$d_{jt} = \sqrt{\sum_{i=1}^{k} B_{ji}^2 \exp(h_{p+i,t}) + \exp(h_{jt})} \tag{56}$$

In the equation for h_j, the actual returns y_j and their expected standard deviation d_j are used to apply the asymmetric term to the innovations:

$$h_{jt} = \mu_j + \phi_j(h_{j,t-1} - \mu_j) + \frac{\theta_{j-p}}{d_{j,t-1}}\left(|y_{j,t-1}| - \gamma - j - p y_{j,t-1} - d_{j,t-1}\sqrt{\frac{2}{\pi}}\right) + \sigma_j \eta_{jt} \tag{57}$$

In Figure 6, a plot similar to Figure 4 is shown. Although this model is based on observable return data rather than unobserved factor returns, the volatility correlation between assets is noticed to a lesser degree. To show this, a comparison between the trailing 50-day standard deviations is made between the model from section 4.2 and this section. This comparison can be found in Figure 7. In this figure, it can be seen that the volatility between assets doesn't move as simultaneous as the model from section 4.2.

For the reasons mentioned above, this model is not looked into further. It is still displayed here since it might be of interest to someone who wants to model the innovations asymmetrically.

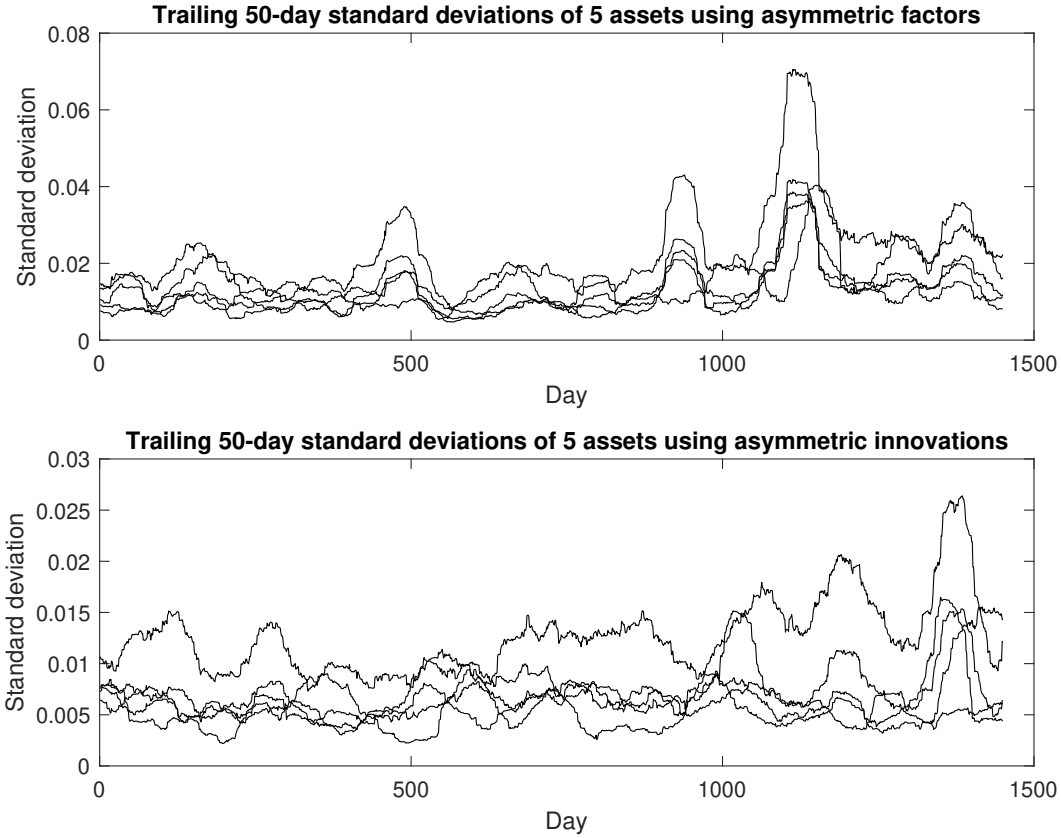

Figure 7: Comparison of the 50-day trailing standard deviation between adding the asymmetric relation to the factors and adding the asymmetric realtion to the innovations.

4.4. ASYMMETRIC FACTORS

In Section 4.2, it was found that using the sampled factors $\hat{f}_t$ did not allow for accurate estimation of the θ-parameter. To adjust for this, a new variable is introduced, the expected factor:

$$\bar{f}_t = E[f_t|B, y_t, h_t] \tag{58}$$

Note that there is a relation between $\hat{f}_t$ and $\bar{f}_t$:

$$\hat{f}_t \sim N(\bar{f}_t, F_t) \tag{59}$$

From [Chib et al., 2006]:

$$F_t = (B' V_t^{-1} B + D_t^{-1})^{-1} \tag{60}$$
$$\bar{f}_t = F_t B' V_t^{-1} y_t \tag{61}$$
$$V_t = \text{diag}(\exp(h_{1,t}), \ldots, \exp(h_{p,t})) \tag{62}$$
$$D_t = \text{diag}(\exp(h_{p+1,t}), \ldots, \exp(h_{p+k,t})) \tag{63}$$

For stability reasons, the added asymmetry term should average zero. In equation 55 the fact that $E[|\hat{f}_{j-p,t}|] = \exp(h_{j-p,t})\sqrt{\frac{2}{\pi}}$ (due to the normality of $\hat{f}_t$) could be exploited. The same solution cannot be used when working with $\bar{f}_t$ since $\bar{f}_t$ is not normally distributed. This means that $|\bar{f}_t|$ should be omitted. However, the mean of $\bar{f}_t$ (i.e. without absolute value) can easily be computed: $E[\bar{f}_t] = E[E[f_t|B, y_t, h_t]] = 0$, which means that $\bar{f}_t$ can be used. This leads to the following equation:

$$h_{jt} = \mu_j + \phi_j(h_{j,t-1} - \mu_j) - \frac{\gamma - j - p f_{j-p,t-1}}{\exp(h_{j,t-1})} + \sigma_j \eta_{jt} \tag{64}$$

This equation can be seen as a mix between the EGARCH model from [Nelson, 1991] and the Leverage Effect Model from [Harvey and Shephard, 1996]. Some quantities are defined to show the main differences between equations 55 and 64:

$$FR = \frac{f_{j-p,t-1}}{\exp\left(\frac{h_{t-1,j}}{2}\right)} \tag{65}$$
$$I_A = \frac{|f_{j-p,t-1}| - f_{j-p,t-1}}{\exp(\frac{h_{j,t-1}}{2})} - \sqrt{\frac{2}{\pi}} \tag{66}$$
$$I_B = -\frac{f_{j-p,t-1}}{\exp\left(\frac{h_{t-1,j}}{2}\right)} \tag{67}$$

Here FR is the normalized factor return, and I_X is the increase in volatility for model X (after multiplying with the corresponding parameter). Since FR is normally distributed as a $N(0,1)$ it suffices to check the behaviour for $FR \in [-3,3]$. In Figure 8 a comparison of I_A and I_B is given in this range.

In both models, the logarithmic volatility increases linearly when the factor return is negative. The big difference is when the factor returns are positive. Model A will decrease the logarithmic volatility with $\sqrt{\frac{\pi}{2}}$, On the other hand, using model B, the logarithmic volatility decreases linearly with positive factor returns.

As a result, using equation 64, one also assumes that when factor returns are positive, volatility tends to calm down. This assumption is somewhat in accordance with the behavioral tendencies of people (i.e. profit taking when returns are positive, cutting losses when returns are negative).

In Figure 9, the effect of increasing γ is shown. The easiest way to think about the parameter is that it increases the random noise, by adding the correlated part between $f_{j-p,t-1}$ and $\eta_{j,t}$ to the model.

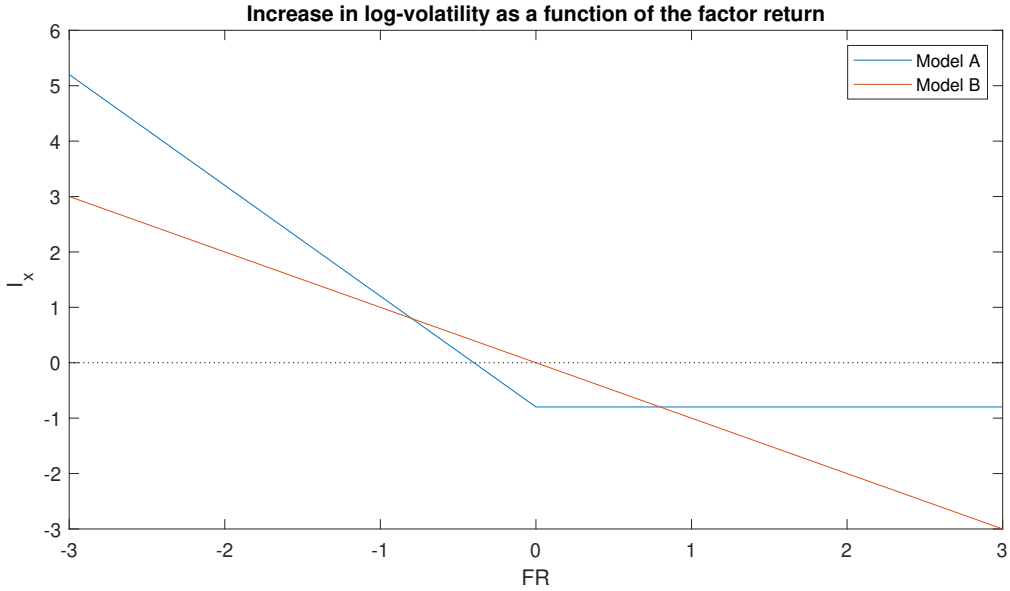

Figure 8: Comparison between model A (equation 55) and model B (equation 64). In this figure, the response I is compared to the factor return FR. The definitions of I and FR can be found in equations 65 to 67

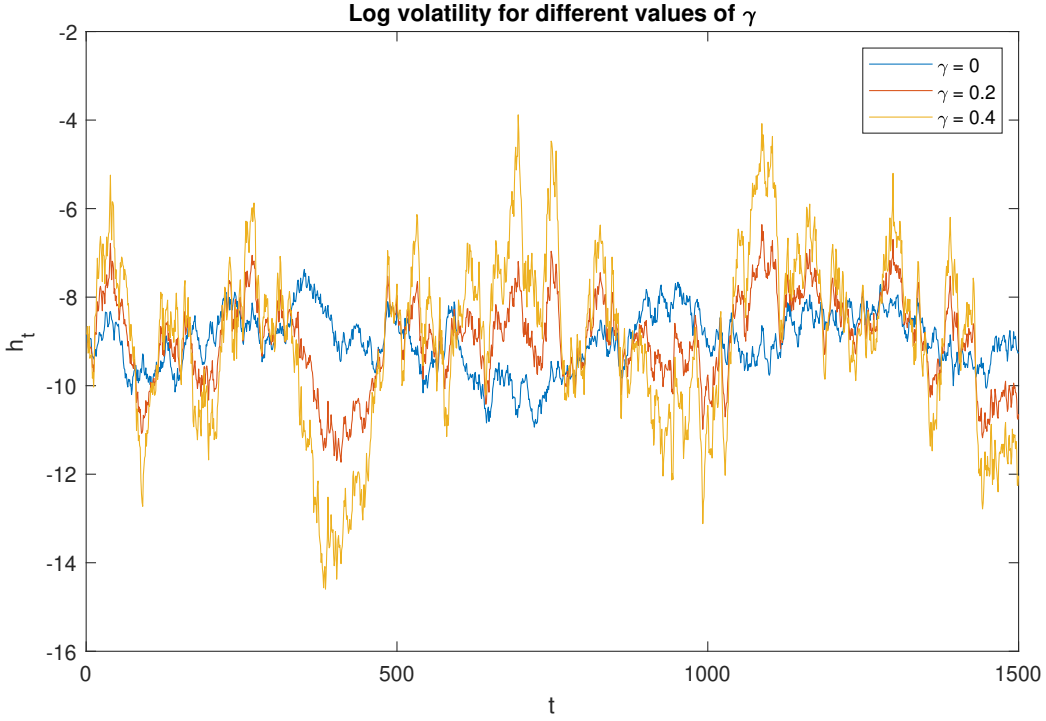

Figure 9: The path for the logarithmic volatility is shown for three different values of γ, all other variables (including the generated random numbers) are kept the same.

MODEL AND ESTIMATION

5.1. MODEL SPECIFICATION

In this section, the model, which is heavily discussed in section 4, will be fully specified.

$$y_t = B f_t + u_t \tag{68}$$

Where $y_t = \{y_{1t}, y_{2t}, \ldots, y_{pt}\}$ are the p observations at time t, with $t \in \{1, \ldots, n\}$. B is a factor loading matrix of size p by k, with the restrictions $b_{ii} = 1$ for $i \leq k$ and $b_{ij} = 0$ for $j > i$. Visit equation 4 in Section 2.2 for an example of such a matrix. $f_t = \{f_{1t}, f_{2t}, \ldots, f_{kt}\}$ are the unobserved factors. Last $u_t = \{u_{1t}, u_{2t}, \ldots, u_{pt}\}$ are the innovations. For the innovations:

$$u_{jt} \sim N(0, \exp(h_{jt})) \tag{69}$$

And for the factors:

$$f_{jt} \sim N(0, \exp(h_{p+j,t})) \tag{70}$$

Both u_{jt} as well as f_{jt} are functions of the logarithmic volatility h_{jt}. This logarithmic volatility follows a mean-reverting process. For $j \leq p$, h_j follows the three-parameter (μ, ϕ, σ) process:

$$h_{jt} = \mu_j + \phi_j(h_{j,t-1} - \mu_j) + \sigma_j \eta_{jt} \tag{71}$$

For $j > p$, an additional term is added. This term increases the logarithmic volatility when a negative return occurs, meanwhile decreasing it when a positive return is observed. h now follows the four-parameter $(\mu, \phi, \sigma, \gamma)$ process:

$$h_{jt} = \mu_j + \phi_j(h_{j,t-1} - \mu_j) - \frac{\gamma_{j-p} f_{j-p,t-1}}{\exp\left(\frac{h_{j,t-1}}{2}\right)} + \sigma_j \eta_{jt} \tag{72}$$

The above model will be referred to as the AMSV model (Asymmetric Multivariate Stochastic Volatility model). Furthermore, MSV is used for the classical model without asymmetry in the factor returns. The results of the AMSV model will be compared to this classical model. It will also be compared to the model with a jump component (as described in [Chib et al., 2006]), this model will be abbreviated as MSVJ. Please find an overview of all the models compared in this paper in Table 6.

Model	Full Name	Explanation
MSV	Multivariate Stochastic Volatility Model	Standard model
MSVJ	Multivariate Stochastic Volatility Jump Model	MSV with a jump component
AMSV	Asymmetric Multivariate Stochastic Volatility Model	MSV with asymmetric factors

Table 6: An overview of the different models and their abbreviations used in this thesis.

It is important to note that although the MSV and AMSV model are similar in the number of parameters. The MSVJ model is significantly larger. Moreover, it also uses the most latent variables.

To give a better image: The MSV model uses $pk - \frac{k^2+k}{2}$ free elements in the factor loading matrix. It also uses $3(p+k)$ log-volatility parameters and $n(p+k)$ latent variables (h) that should be saved between iterations. For $p = 10, k = 3, n = 1500$, that is a total of 63 parameters and 19500 latent variables. Compared with MSV, the AMSV model adds k parameters (the γ-parameters) and no additional latent variables should be saved between iterations. This makes the total 66 parameters and 19500 latent variables. For MSVJ (again compared with MSV), $2p$ variables are added (δ and κ) as well as $2n(p+k)$ latent variables (jump intensity matrix K and jump indicator matrix q). Making the total for MSVJ 83 parameters and 58500 latent variables. This all is summarized in Table 7.

Model	Parameters	Latent Variables	Parameters*	Latent Variables*
MSV	$pk - \frac{k^2+k}{2} + 3(p+k)$	$n(p+k)$	63	19500
MSVJ	$pk - \frac{k^2+k}{2} + 5p + 3k$	$3n(p+k)$	83	58500
AMSV	$pk - \frac{k^2+k}{2} + 3p + 4k$	$n(p+k)$	66	19500

Table 7: An overview of the sizes of the different models. For parameters* and Latent Variables*, the size is given when $p = 10$, $k = 3$ and $n = 1500$

5.2. MCMC-ALGORITHM

In this section, the sampling process using the MCMC method is explained. As discussed in Section 2.3, one of the results from [Chib et al., 2006] is that it is more efficient to sample B (or β, the free elements of B) and f_t in one block. In the second step, the independence of the logarithmic volatilities from each other is exploited and the parameters $\Theta_j = (\mu_j, \phi_j, \sigma_j, \gamma_j)$[2] conditioned on (y, B, f_t) are sampled. Finally, in the same block, $\{h_{j.}\}$ conditioned on $(y, B, f_t, \{\Theta_j\})$ is sampled. As explained in 2.3, this is done for all t simultaneously, as was shown by [Chib and Greenberg, 1995] to be faster to converge than sampling all h_{jt} independently.

The AMSV model is compared with both the MSV and the MSVJ model from [Chib et al., 2006]. For the MSVJ model, a few additional sampling steps to sample the jump frequency and jump size parameters are necessary. In their paper, you can find the required extra sampling steps. To sample using the MSV model, the scheme below can still be used, however, when sampling Θ, the γ parameters can be omitted.

SAMPLING THE FACTOR LOADING MATRIX

First, the likelihood function for β is constructed:

$$\pi(\beta|y, \{h_{j.}\}) \propto \pi(\beta) \prod_{i=1}^{n} p(y_t|B, h_t) \tag{73}$$

Where $\pi(\beta)$ is the prior for β. Following [Chib and Greenberg, 1995], the Metropolis-Hastings algorithm (MH-algorithm) as explained in Section 2.4 is used. A candidate for β gets proposed from a multivariate t-distribution: $T(\beta|m, \Sigma, \nu)$ where m is the mode of $l = \log(\prod_{i=1}^{n} p(y_t|B, h_t))$, and Σ is minus the inverse of the hessian of l at the mode. ν is chosen arbitrarily at 15, as in [Chib and Greenberg, 1995]. To find the mode m, first, the log-likelihood of β should be optimized:

$$\begin{aligned} l &= \log(\prod_{i=1}^{n} p(y_t|B, h_t)) \\ &= c - \frac{1}{2}\sum_{t=1}^{n} \log|\Omega_t| - \frac{1}{2}\sum_{t=1}^{n} y_t' \Omega_t^{-1} y_t \end{aligned} \tag{74}$$

With c a constant. The matrix Ω_t is a function of the log-volatilities h_{jt} and the factor loading matrix B.

$$V_t = \text{diag}(\exp(h_{1,t}), \ldots, \exp(h_{p,t})) \tag{75}$$

$$D_t = \text{diag}(\exp(h_{p+1,t}), \ldots, \exp(h_{p+k,t})) \tag{76}$$

$$\Omega_t = V_t + B D_t B' \tag{77}$$

Although this can be solved numerically, by providing an analytic first derivative vector, the optimum is found significantly faster:

$$\begin{aligned} l'_{ij} &= \frac{\delta l}{\delta B_{ij}} \\ &= \sum_{t=1}^{n} Tr\left(y_t y_t' \Omega_t^{-1} B D_t \frac{\delta B}{\delta B_{ij}}' \Omega_t^{-1} - \Omega_t^{-1} B D_t \frac{\delta B}{\delta B_{ij}}' \right) \end{aligned} \tag{78}$$

Note that the formulation of this derivative is different than the formulation in [Chib et al., 2006]. By formulating it this way, it is easier to calculate the second derivative. In the end, the Newton Raphson method chosen

[2] Since only the factors are modelled asymmetrically, $\theta_j = 0$ for $j \leq p$

to optimize l does not require this second derivative. Since some maximizing algorithms are faster to converge when an analytical second derivative is provided as well, the interested reader can find this second derivative in Appendix A. Within the derivations of the first and especially the second derivative, many identities from the matrix cookbook [Petersen and Pedersen, 2012] were used, mostly from chapter 2.

Using the objective function (equation 74), and the derivative (equation 78), the mode m and inverse of the hessian at the mode Σ can be found using iterative schemes. A new candidate value β^* is proposed from $T(m,\Sigma,\nu)$, the candidate generating density.

Using the MH-algorithm, this proposal value β^* gets accepted with probability:

$$\alpha(\beta,\beta^*|y,\{h_{j.}\}) = \min\left(1, \frac{\pi(\beta^*)\prod_{t=1}^n N(y_t|0,V_t+B^*D_t(B^*)')T(\beta|m,\Sigma,\upsilon)}{\pi(\beta)\prod_{t=1}^n N(y_t|0,V_t+BD_tB')T(\beta^*|m,\Sigma,\upsilon)}\right) \tag{79}$$

Where B^* is the factor loading matrix with free elements corresponding to β^*. If β^* is not accepted, the current value for β is used again.

Sampling the factors

The first block is completed by sampling $f_t|y,B,\{h_{j.}\}$. Note that f_t was not used when sampling β, so the sampler is still in the same block. Since all the time steps are independent of each other, the sampling of f can be broken up into the product of the distributions $f_t|y_t,B,h_t$. This distribution is Gaussian with variance: $F_t = (B'(V_t^*)^{-1}B + D_t^{-1})^{-1}$ and mean $\bar{f}_t = F_tB'(V_t^*)^{-1}y_t$, hence:

$$f_t \sim N(\bar{f}_t, F_t) \tag{80}$$

As discussed in Section 4, in the next block, in addition to a sample of f_t, a best estimate $\bar{f}_t$ is required as well. This best estimate is just the mean of the normal distribution in equation 80.

During this thesis, is is discovered that much better results are obtained when e_t^f is defined:

$$e_t^f = \frac{f_t}{\exp\left(\frac{h_t}{2}\right)} \tag{81}$$

It is clear that $e_t^f \sim N(0,1)$. In this block the best estimate of e_t is saved. This estimate will be denoted $\bar{e}_t^f$, and is obtained by plugging in $\bar{f}_t$ in equation 81.

If this step is skipped and in the following block the sampling is conditioned on $\bar{f}_t$ instead of $\bar{e}_t$, the estimation of Θ and h becomes much harder. To sample these quantities, an EKF (see section 2.6) is needed with the following equation for $f(x)$ within this EKF:

$$f(x,\bar{f}_{j-p,t-1}) = \mu_j + \phi_j(x-\mu) - \gamma_{j-p}\frac{\gamma\bar{f}_{j,t-1}}{\exp\left(\frac{x}{2}\right)} \tag{82}$$

And consequently, its derivative:

$$\frac{\delta f}{\delta x} = \phi_j + \frac{\gamma_{j-p}\bar{f}_{j-p,t-1}}{2\exp\left(\frac{x}{2}\right)} \tag{83}$$

As discussed, in Section 2.6, some deficiencies can be expected when working with EKFs. The most noteworthy one, is that σ_j gets underestimated. Another problem is encountered when sampling h. Although some methods are available to get an estimate for h ([Shumway and Stoffer, 2000], for example provides a Kalmann Smoother Algorithm), no known method is available to sample h, where h keeps all its characteristics.

The problem with using a smoother to estimate h, is that the time-series h gets smoothed out. This returns a series which has less variance than the original time series. As a result, in the next cycle σ gets underestimated even more. So although the EKF is very good at estimating the $(\mu_j,\phi_j,\gamma_{j-p})$ parameter set, the underestimation of σ_j is a problem. At first, this problem only seemed a minor inconvenience, since it did not worsen the estimate of h. However, in Section 7.3 it turned out to that this underestimation did actually result in problems when comparing the different methods against each other. More about this, including the sampling results of this method can be found in Appendix B.

SAMPLING LOGARITHMIC VOLATILITY PARAMETERS

Given $(y, B, f_t, \bar{e}_t^f)$, now Θ_j for $j \in \{1, \ldots, p+k\}$ is sampled. Using the fact that the innovations (equation 69) and the factors (equation 70) are all independent, the model can be treated as individual state space models. First, a vector α_t is created, which corresponds to the predicted movement of y_t, given the factor loading matrix B and the factors f_t, i.e. $\alpha_t = Bf_t$. Next, the realized logarithmic volatilities are calculated:

$$z_{jt} = \ln(y_{jt} - \alpha_{jt} + c)^2 \qquad \text{for } j \leq p \tag{84}$$

$$z_{jt} = \ln(f_{j-p,t} + c)^2 \qquad \text{for } j > p \tag{85}$$

For the factors, the log of the square of the estimated factors is taken, while for the innovations z is calculated by taking the log of the square of the returns minus the projected returns from the factors. This subtraction should equal the innovations. $c = 10^{-8}$ is added to make sure the natural logarithm is always defined. In Section 2.5, it was shown that: $z_{jt} = h_{jt} + \log(\chi_1^2)$. So z_{jt} goes like h_{jt} plus a log chi-squared random variable with one degree of freedom. h and its parameters will be estimated using the Kalman Filter, for which it is necessary to use Gaussian distributions. Therefore, the logarithmic chi-squared random variable is approximated by a mixture of seven Gaussian distributions. A comparison between the logarithmic chi-squared random variable and the estimation using seven Gaussian distributions is made in [Kim et al., 1998]. The probabilities, means, and variances of these seven Gaussian distributions can be found in Table 1 in Section 2.5.

Using this approximation:

$$z_{jt}|s_{jt}, h_{jt} \sim N(h_{jt} + m_{s_{jt}}, v_{s_{jt}}^2) \tag{86}$$

Since the series are independent for each j, each series is treated separately. For each j, first $s_{j.}$ gets sampled from:

$$p(s_{j.}|z_{j.}, h_{j.}) = \prod_{t=1}^{n} p(s_{jt}|z_{jt}, h_{jt}) \tag{87}$$

There is also no independence between the different time steps, so s_{jt} can be sampled individually:

$$p(s_{jt}|z_{jt}, h_{jt}) \propto \pi(s_{jt})\phi(z_{jt}|h_{jt} + m_{s_{jt}}, v_{s_{jt}}^2) \tag{88}$$

Where $\phi(.)$ denotes the normal probability distribution function. This equation only has seven points of mass, which makes the sampling straightforward.

Next Θ_j is sampled from the density

$$\begin{aligned} p(\Theta_j|z_{j.}, s_{j.}) &\propto \pi(\Theta_j)p(z_{j.}|s_{j.}, \Theta_j) && \text{for } j \leq p \\ p(\Theta_j|z_{j.}, s_{j.}, \bar{e}_{j-p.}^f) &\propto \pi(\Theta_j)p(z_{j.}|s_{j.}, \bar{e}_{j-p.}^f, \Theta_j) && \text{for } j > p \end{aligned} \tag{89}$$

Where:

$$\begin{aligned} p(z_{j.}|s_{j.}, \Theta_j) &= p(z_{j1}|s_{j1}, \Theta_j) \prod_{t=2}^{n} p(z_{jt}|\mathscr{F}_{t-1}^*, s_{jt}, \Theta_j) && \text{for } j \leq p \\ p(z_{j.}|s_{j.}, \bar{e}_{j-p.}^f, \Theta_j) &= p(z_{j1}|s_{j1}, \Theta_j) \prod_{t=2}^{n} p(z_{jt}|\mathscr{F}_{t-1}^*, s_{jt}, \bar{e}_{j-p,t-1}^f, \Theta_j) && \text{for } j > p \end{aligned} \tag{90}$$

Both $p(z_{jt}|\mathscr{F}_{t-1}^*, s_{jt}, \Theta_j)$ as well as $p(z_{jt}|\mathscr{F}_{t-1}^*, s_{jt}, \bar{e}_{j-p,t-1}^f, \Theta_j)$ are normal densities with parameters the result of the Kalman Filter recursion (explained in detail in 2.5).

The sampling of each Θ_j is similar to the sampling of β. Using equation 90, the mode m and the hessian at the mode Σ are found. A candidate for Θ_j is generated from $T(\Theta|m, \Sigma, \nu)$, again setting $\nu = 15$. Then using the MH algorithm, the candidate is accepted or rejected using the acceptance ratio.

Last, $h_{j.}$ is sampled. This time series is sampled from $[h_{j.}|z_{j.}, s_{j.}, \Theta_j]$ for $j \leq p$, and for the factors ($j > p$), h_j is sampled from: $[h_{j.}|z_{j.}, s_{j.}, \bar{e}_{j-p.}^f \Theta_j]$. The sampling is done by running the Kalman Filter on equation 90 again, and using the Simulation Smoother algorithm from [De Jong and Shephard, 1995] afterwards. This might be the weakest part of the algorithm since for certain combinations of ϕ_j, σ_j the algorithm has a small probability to define negative variances within the alogrithm of [De Jong and Shephard, 1995]. This problem is not noted by the authors of [Chib et al., 2006], since the range of values of ϕ_j, σ_j used by the authors does not run into these problems. Especially for a combination of low ϕ and low σ, this problem is observed.

MODEL COMPARISON

In this section, a method to compare the different models with each other is explained. To accomplish this, the likelihood of the data given each model will be compared and a Bayes-factor will be calculated and discussed, as is done is [Chib, 1995], and [Chib and Jeliazkov, 2001], for example. Similar to [Chib et al., 2006], first the method of [Chib, 1995] gets used, exploiting the Bayes-definition:

$$p(y|\mathscr{M}) = \frac{p(y|\mathscr{M},\psi^*)p(\psi^*|\mathscr{M})}{p(\psi^*|\mathscr{M},y)} \tag{91}$$

Where $p(y|\mathscr{M})$ is the likelihood function under the given model, $p(y|\mathscr{M},\psi^*)$ is the likelihood function under the given model and given parameters ψ^* and $p(\psi^*|\mathscr{M})$, and $p(\psi^*|\mathscr{M},y)$ are respectively the prior and posterior of ψ under the model $\mathscr{M}$.

To compare two models, the quantity of interest is: $B = \frac{p(y|\mathscr{M}_1)}{p(y|\mathscr{M}_2)}$, or equivalently $b = \log(B)$. Using equation 91:

$$\begin{aligned} b &= \log(p(y|\mathscr{M}_1)) - \log(p(y|\mathscr{M}_2)) \\ &= \log(p(y|\mathscr{M}_1,\psi_1^*)) + \log(p(\psi_1^*|\mathscr{M}_1)) - \log(p(\psi_1^*|\mathscr{M}_1,y)) \\ &\quad - \log(p(y|\mathscr{M}_2,\psi_2^*)) - \log(p(\psi_2^*|\mathscr{M}_2)) + \log(p(\psi_2^*|\mathscr{M}_2,y)) \end{aligned} \tag{92}$$

The prior density is chosen within the model and therefore already available, the two obvious next steps are determining the likelihood of the data given the model and the parameters, and determining the posterior density of the parameters, this will be done in the following two sections.

6.1. POSTERIOR ORDINATE ESTIMATION

First, the posterior density is estimated. The likelihood of the model can be broken down in the following way:

$$\begin{aligned} p(\psi^*|\mathscr{M},y) &= p(\beta^*,\theta^*|\mathscr{M},y) \\ &= p(\beta^*|\mathscr{M},y)p(\Theta^*|\beta^*,\mathscr{M},y) \end{aligned} \tag{93}$$

This means that first the posterior density of the factor loading matrix can be analyzed. Thereafter, fixing the factor loading matrix at it's posterior mean, an analysis for Θ can be made.

Starting with the first term, it is noted that the density of the free elements of the factor loading matrix is close to normal. Hence, to estimate this term, the mean vector and covariance matrix obtained within the MCMC run are used. The posterior density is estimated using a normal density with these parameters.

Next, $p(\theta^*|\beta^*,M,y)$ is estimated. This is done by fixing β^* and all but one of the Θ^* at their posterior mean. Starting with $j = 1$, a "reduced run" is performed. In this reduced run $\{\Theta_j,\{h_j.\},f_t\}$ are sampled, while $\{B,\{\Theta_i\}\}$ for $i \neq j$ are kept fixed. The ordinate is then estimated using the kernel smoothing approach (discussed in [Silverman, 2018]). The process is repeated in sequence for $j \in \{1,\ldots,p+k\}$.

6.2. LIKELIHOOD ESTIMATION USING THE PARTICLE FILTER

METHOD

The goal of this section is to estimate the log-likelihood ordinate: $\log(f(y_1,\ldots,y_n|\mathscr{M},\psi^*))$.

$$\log(f(y_1,\ldots,y_n|\mathscr{M},\psi^*)) = \sum_{t=1}^{n} \log(f(y_t|\mathscr{M},\mathscr{F}_{t-1},\psi^*)) \tag{94}$$

The one step ahead density of y_t in turn can be simplified to:

$$f(y_t|\mathscr{M},\mathscr{F}_{t-1},\psi^*) = \int N(y_t|0,\Omega_t(h_t,B^*))p(h_t|\mathscr{M},\mathscr{F}_{t-1},\psi^*)dh_t \tag{95}$$

This equation contains a normal distribution which is already available, and the the one step ahead density of h_t, which is not yet available. This can be rewritten as:

$$p(h_t|\mathscr{M},\mathscr{F}_{t-1},\psi^*) = \int p(h_t|\mathscr{M},h_{t-1},\psi^*)p(h_{t-1}|M,\mathscr{F}_{t-1},\psi^*)dh_{t-1} \tag{96}$$

Now the density of h_t, given $(\mathscr{M}, h_{t-1}, \psi^*)$ is just:

$$p(h_t|\mathscr{M}, h_{t-1}, \psi^*) = \prod_{j=1}^{p} N(h_{jt}|\mu_j^* + \phi_j^*(h_{j,t-1} - \mu_j^*), \sigma_j^{*2}) \tag{97}$$

$$\times \prod_{j=p+1}^{p+k} N(h_{jt}|\mu_j^* + \phi_j^*(h_{j,t-1} - \mu_j^*) + \gamma_{j-p}^* \frac{f_{j-p,t-1}}{\exp(\frac{h_{j,t-1}}{2})}, \sigma_j^{*2}) \tag{98}$$

$p(h_{t-1}|M, \mathscr{F}_{t-1}, \psi^*)$ in equation 96 is the filtered density of h_{t-1}. To estimate the one step ahead density of y_t and h_t, the particle filter is used, as described by [Pitt and Shephard, 1999] and [Doucet et al., 2001]. This method uses particles from the previous filtered distribution to estimate the current distribution. That is, a sample of M values $\{h_{t-1}^{(1)}, \ldots, h_{t-1}^{(M)}\}$ from the distribution $p(h_{t-1}|\mathscr{M}, \mathscr{F}_{t-1}, \psi^*))$ is used as a starting point. Using these samples, the density of h_t is estimated using:

$$p(h_t|\mathscr{M}, \mathscr{F}_{t-1}, \psi^*) \approx \frac{1}{M} \sum_{i=1}^{M} p(h_t|\mathscr{M}, h_{t-1}^{(i)}, \psi^*) \tag{99}$$

The Using this approximation, equation 95 can be rewritten as:

$$f(y_t|\mathscr{M}, \mathscr{F}_{t-1}, \psi^*) \approx \int N(y_t|0, \Omega_t(h_t, B^*)) \frac{1}{M} \sum_{i=1}^{M} p(h_t|\mathscr{M}, h_{t-1}^{(i)}, \psi^*) dh_t \tag{100}$$

Where, of course, a new M values to simulate the next time step is sampled. A detailed description of the particle filter used can be found in the section below.

Particle Filter Algorithm

Step 1 The particle filter algorithm starts with $\{h_{t-1}^{(1)}, \ldots, h_{t-1}^{(M)}\}$, which are obtained from the previous time step (at $t = 1$, $h_0^g = \mu$ is used for $g \in \{1, \ldots, M\}$). Now the expected value of h given the previous time step is calculated: $\hat{h}_t^{*(g)} = E[h_t^{(g)}|h_{t-1}^{(g)}]$ for $g \in \{1, \ldots, M\}$. For this expectation, f_{t-1}^g is required, this can be extracted using the method from 4, setting $F_{t-1}^{(g)} = (B'(V_{t-1}^{(g)})^{-1}B + (D_{t-1}^{(g)})^{-1})^{-1}$ and $f_{t-1}^{(g)} = F_{t-1}^{(g)} B'(V_{t-1}^{(g)})^{-1} y_{t-1}$. With $V_t^{(g)}$ and $D_t^{(g)}$ from equation 75 and 76 (using $h_t^{(g)}$ instead of h_t).

Now an estimate for $N(y_t|0, \Omega_t)$ is made:

$$w_g = N(y_t|0, \Omega_t(\hat{h}_t^{*(g)}, B^*))$$

This factor is also used as an indication of how likely the data is given the current combination of $\hat{h}_t^{*(g)}$ and B^*. The integers $1, \ldots, M$ are sampled R times with probability

$$\bar{w}_g = \frac{w_g}{\sum_{i=1}^{M} w_i} \tag{101}$$

These sampled indices are labeled $k_1, \ldots, k_R$ and map to $\hat{h}_t^{*(k_1)}, \ldots, \hat{h}_t^{*(k_R)}$

Step 2 Now for each sampled index k_g, the original value of $h_{t-1}^{(k_g)}$ (the original values from the previous time step) are used to sample $h_t^{*(g)}$:

$$h_{j,t}^{*(g)} = \mu_j^* + \phi_j^*(h_{j,t-1}^{(k_g)} - \mu_j^*) + \sigma_j^* \eta_{j,t}^{(g)} \qquad \text{for } j \leq p \tag{102}$$

$$h_{j,t}^{*(g)} = \mu_j^* + \phi_j^*(h_{j,t-1}^{(k_g)} - \mu_j^*) - \gamma_{j-p}^* \frac{f_{j-p,t-1}^{(k_g)}}{\exp(\frac{h_{j,t-1}^{(k_g)}}{2})} + \sigma_j^* \eta_{j,t}^{(g)} \qquad \text{for } j > p \tag{103}$$

This produces the sample: $\{h_t^{*(1)}, \ldots, h_t^{*(R)}\}$

Step 3 Finally, the values $\{h_t^{*(1)}, \ldots, h_t^{*(R)}\}$ are re-sampled M times using the weight from equation 105 as sample-probability:

$$w_g^* = \frac{N\left(y_t|0, \Omega_t\left(h_t^{*(g)}, B^*\right)\right)}{N\left(y_t|0, \Omega_t\left(\hat{h}_t^{*(k_g)}, B^*\right)\right)} \tag{104}$$

$$\bar{w}_g^* = \frac{w_g^*}{\sum_{i=1}^{M} w_i^*} \tag{105}$$

for $g = 1, \ldots, R$. This produces the filtered sample $\{h_t^{(1)}, \ldots, h_t^{(M)}\}$ from $(h_t | M, F_t, \psi^*)$. The desired sample, which is used as a starting point for the next time step.

Step 4 The estimate for the log-likelihood can now be obtained by the method of [Pitt, 2002]:

$$\hat{f}(y_t | \mathscr{M}, \mathscr{F}_{t-1}, \psi^*) \approx \left(\frac{1}{M} \sum_{g=1}^{M} w_g \right) \left(\frac{1}{R} \sum_{g=1}^{R} w_g^* \right)$$

After this cycle has repeated for $t = 1, \ldots, N$, the logarithmic likelihood (equation 94) can be estimated using:

$$\log(f(y_1, \ldots, y_n | \mathscr{M}, \psi^*)) \approx \sum_{t=1}^{n} \log \left(\left(\frac{1}{M} \sum_{g=1}^{M} w_g^t \right) \left(\frac{1}{R} \sum_{g=1}^{R} w_g^{t*} \right) \right)$$

Where w_g^t and w_g^{t*} are notation for w_g and w_g^* computed at time step t respectively.

SIMULATION RESULTS

7.1. MODEL SET-UP AND PRIORS

In this section, the accuracy of the models and the ability of the method from Section 6 to provide a criterion to distinguish between data generated with different models is studied. After explaining the details of the parameters of the model and the prior distributions in this section, in Section 7.2 the accuracy of the estimation of the parameters for all three models is tested. In Section 7.3, the log-likelihood of all three models will be compared to each other, when underlying data sets are generated by one of the models. The results of this section will be used in section 8, where real-world data is sampled using all three methods.

For the simulation of the artificial data sets, the following set of parameters is taken. The free elements of B are $N(0.75, 0.5)$ distributed, $(\mu_j, \phi_j, \sigma_j, \gamma_j)$ are set at $(-9; 0.98; 0.14; 0.2)$, where γ_j is omitted in the MSV and MSVJ models. The reason the Θ parameters are fixed instead of simulated is for bundling purposes. A full run of the sampler takes quite a long time. By simulating the innovations the same way, they can be bundled. This way more data on the distributions of the parameters are available, instead of little data being available for many different parameters. For the same reason, δ_j is fixed at 0.06 and κ_j is fixed at 0.02 for all j.

For the Model Accuracy Section (Section 7.2), B is simulated once, and then this factor loading matrix is used for all the simulations. For the Model Comparison Section (Section 7.3), $B$is re-simulated each time a new synthetic data set is generated.

In Section 7.2, 6.000 samples are accumulated for 40 runs. In each run, the first 1.000 simulations are discarded (the burn-in period). The last 5.000 iterations are used for inference.

In section 7.3, 6.000 samples are accumulated for 10 runs, again discarding the first 1.000 samples of each run. After the sampling is done, the Reduced Sampling from Section 6.2 starts for 5.000 iterations each reduced run. Finally, the Particle Filter is run (also from section 6.2), with $M = 5.000$, and $R = 25.000$.

The number of samples, particles, and reduced run samples are based on [Chib et al., 2006], where a convergence analysis is performed. These results are scaled down by a factor of a maximum of two since the model in this paper is at least a factor of two smaller.

During the MCMC method, the following priors are taken. For the free elements of B, the used prior is $N(0.9, 1)$, a slightly wider prior with a slightly different mean as the distribution function of the free elements. For the logarithmic volatility parameters, the prior for μ_j is set at $\mu_j \sim N(-9, 1)$. ϕ_j is distributed by simulating a transformed version: ϕ_j^* from a beta distribution and then taking $\phi_j = 2\phi_j^* - 1$. This is done in such a way that the mean of $\phi_j = 0.94$ and its standard deviation 0.08. σ_j is simulated from an Inverse Gamma distribution, such that its mean is 0.2 and its standard deviation 0.1. Finally, the prior for γ_j is a $N(0; 0.25)$. Note that these priors are all quite close to the chosen values, with an exception for γ, since part of the goal of the accuracy section is to see if γ can actually be detected from the data, it is desirable to give an unbiased prior about its sign. For the jump parameters: $\log(\delta_j) \sim N(-3, 0.4)$ is used as a prior for δ and $\kappa_j \sim \text{Beta}(2, 100)$ as a prior for κ. These priors indicate about 4 or 5 jumps a year (probability of a jump is 1.96% per day) with an intensity of about 5%.

The choice of all priors, are based on analysis done in [Chib et al., 2006] and [Albert and Hu, 2019], except for γ.

7.2. MODEL ACCURACY

AMSV

In this section, the accuracy of the different sampling methods is analyzed. Starting with AMSV, where-after the MSV and MSVJ model are analyzed and compared with the AMSV model.

For all 40 runs mentioned in section 7.1, the posterior mean for the factor loading matrix (B_i^* for run i) is collected. Using these means, a grand posterior average, B^* is calculated, together with the grand posterior standard deviation.

In Figure 10 the true values of the free elements of the factor loading matrix B are compared with the grand posterior mean obtained from the sampler. It can be seen that the true value and the estimate are close for all factor loadings. In Figure 11 the actual errors are compared with the standard errors. As can be seen, most of the factor loadings are within one standard error of the true value, with a few of the errors being slightly larger,

but always within two standard errors. This means that using the AMSV model, the factor loading matrix can be accurately estimated. Later in this section, these results are compared with the MSV and MSVJ models respectively.

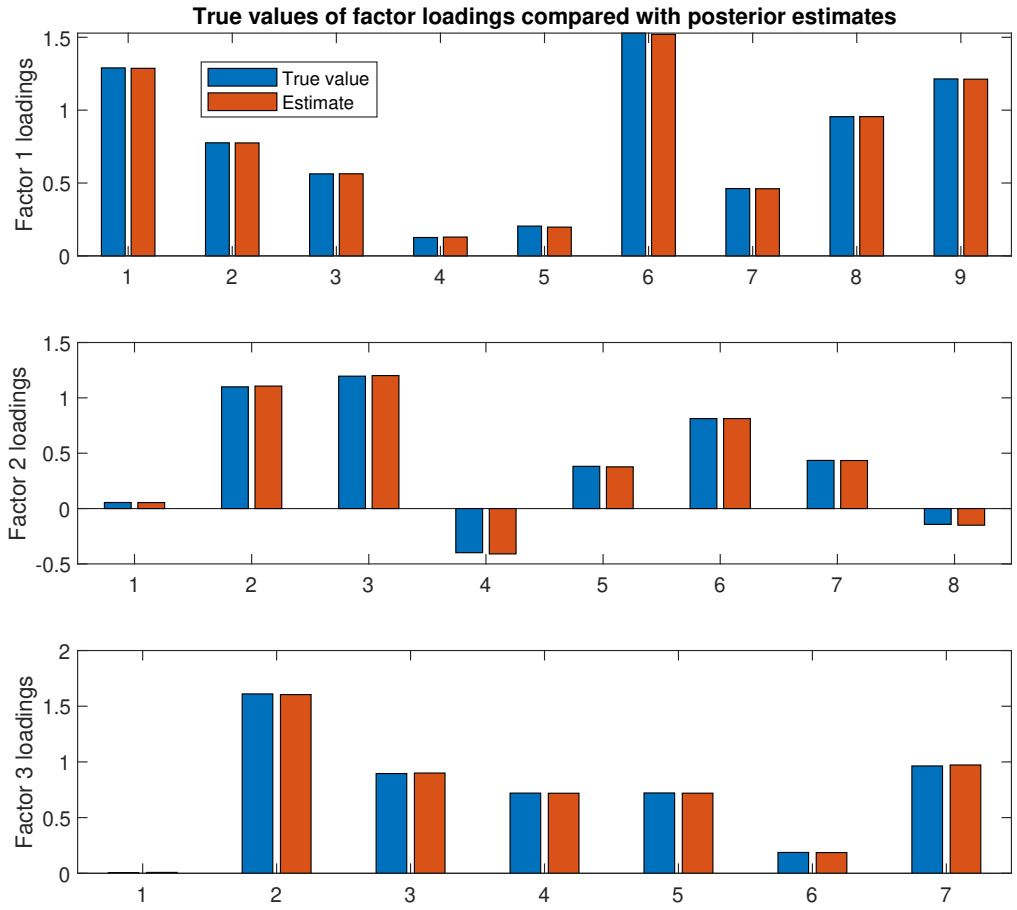

Figure 10: The parameters of the factor loading matrix. The values used for the generation of the data are shown under "True Value". The values under "Estimate" are the average of posteriors of 40 simulations using 5.000 samples (after 1.000 discarded burn-in samples), generated by the AMSV model.

Parameter	True Value	Posterior mean	Posterior s.d.
μ_i	-9	-9.00	0.18
μ_f	-9	-8.97	0.20
ϕ_i	0.98	0.963	0.040
ϕ_f	0.98	0.978	0.005
σ_i	0.14	0.161	0.023
σ_f	0.14	0.182	0.020
θ_f	0.2	0.186	0.021

Table 8: The parameters of the logarithmic volatility. The values used for the generation of the data are shown under "True Value". The values under "Posterior mean" and "Posterior s.d." are respectively the average and the standard deviation of posteriors of 40 simulations using 5.000 samples (after 1.000 discarded burn-in samples), generated by the AMSV model.

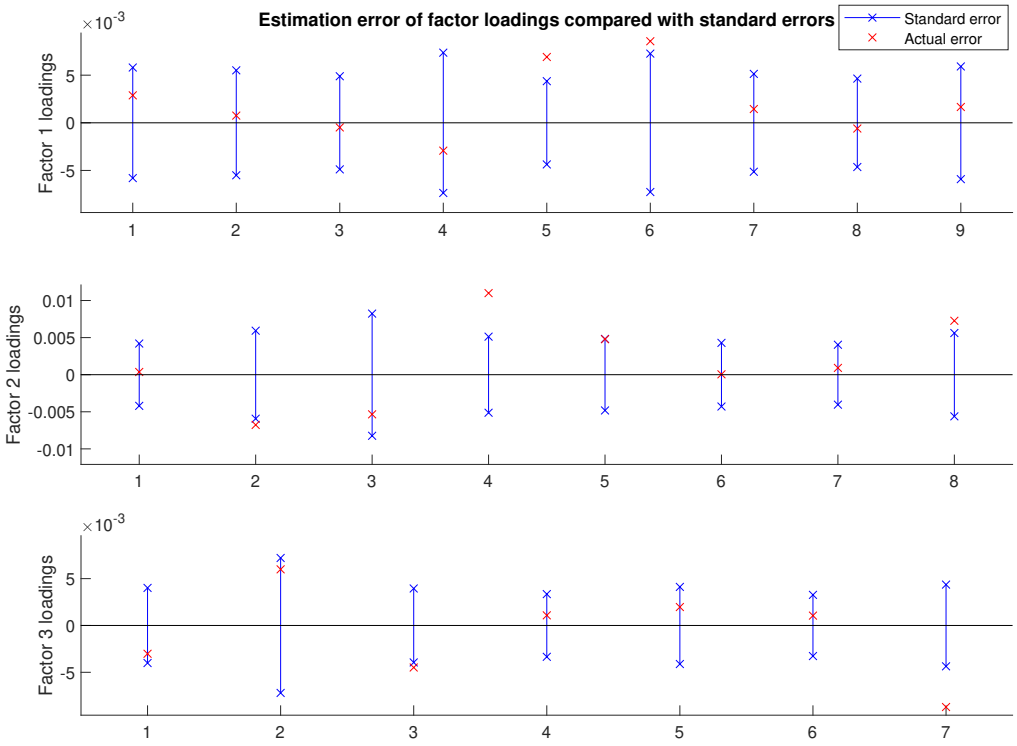

Figure 11: The standard errors and the actual errors of the values of the factor loading matrix. These values are the result of 40 simulations using 5.000 samples (after 1.000 discarded burn-in samples), generated by the AMSV model.

Similarly, the posterior means for Θ are collected. This time, as discussed in Section 7.1, all Θ_j from all runs for $j \leq p$ get bundled together to calculate Θ_i^*: the grand posterior average of logarithmic volatility parameters for the innovations. The same is done for the factors to create: Θ_f^*. The results are summarized and compared to the true values in Table 8

Note that in Table 8 the standard deviation is displayed (not the standard error). Comparing the grand posterior averages with the true values, we find that (μ_i, μ_f) are accurately estimated within 2 standard errors. $(\phi_i, \phi_f, \gamma_f)$ are estimated within 5 standard errors, while (σ_i, σ_f) are estimated less accurately (respectively 6 and 14 standard errors off). It is also observed in [Chib et al., 2006] that the estimation of σ and ϕ is generally less accurate. This will also be seen during the analysis of the MSV and the MSVJ models later this section.

A possible explanation for the larger overestimation of σ_f is that the best estimate for the factors are used in the estimation scheme (and not the sampled factors), this removes some of the randomness captured by the term that uses the previous random element generated by the factors. In turn, this randomness is absorbed by σ_f. Moreover, σ_f is already slightly overestimated, as can be seen in this section (σ_i) and the two sections to follow. However, this trade-off is necessary to have an accurate estimate for γ_f, as was discussed in Section 4, and especially in Figure 5.

In Figure 12, the distribution for one of the factor loadings and three of the logarithmic volatility parameters for factor one are shown. In these Figures, a vertical red line marks the true value of the parameter. In Figure 13, the path of the sampler for these parameters can be found. The true value of the parameters is added to the figures as well, using a horizontal red line. The most important take away from figure 13, is that the sampler has found the equilibrium posterior distribution (i.e. it is not trending anymore).

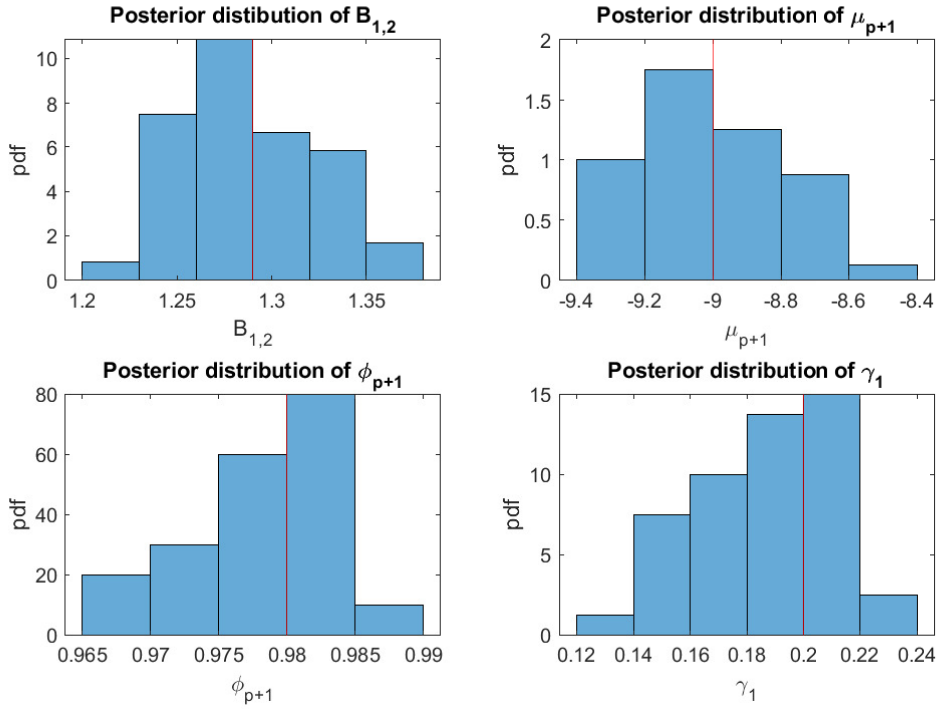

Figure 12: Histogram of the posterior values for one of the factor loadings and the parameters for the first factor. For this histogram, 40 data sets using 5.000 samples are used. The data is generated by the AMSV model.

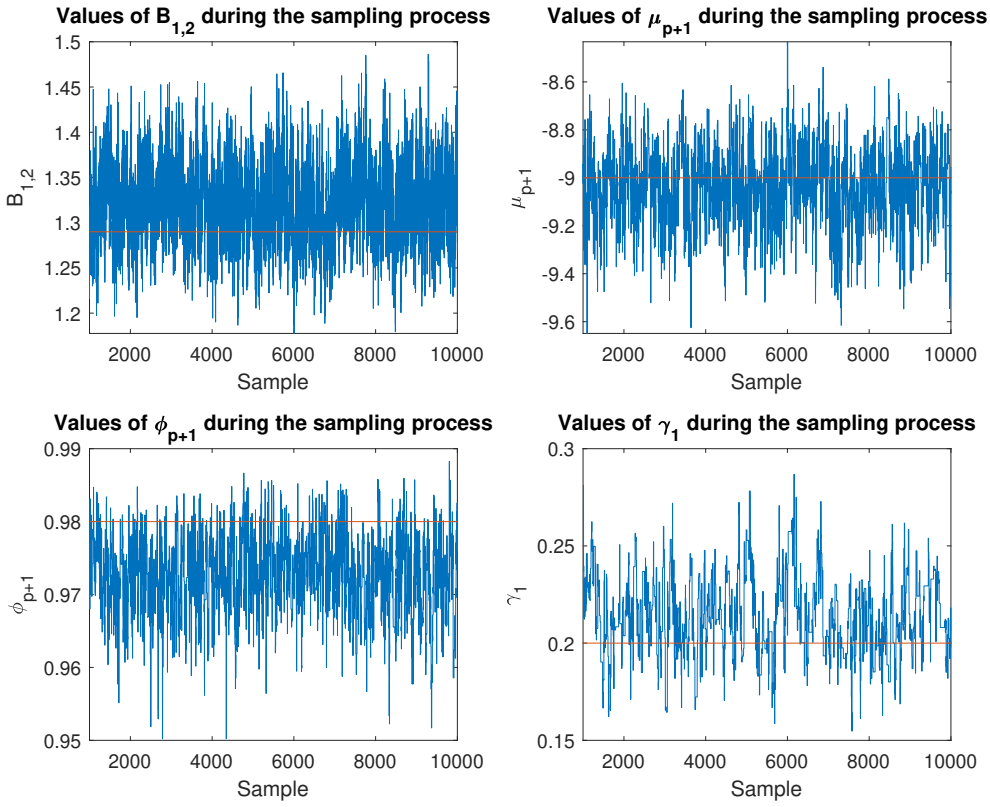

Figure 13: Path of the sampling of the parameters in Figure 12 of an arbitrary run using the AMSV model.

Finally, in Figure 14 the distribution for the three γ parameters in one of the runs is shown. The distributions shown here can be compared with the distribution for γ in Section 8, where of course only one run can be observed for each selection of assets.

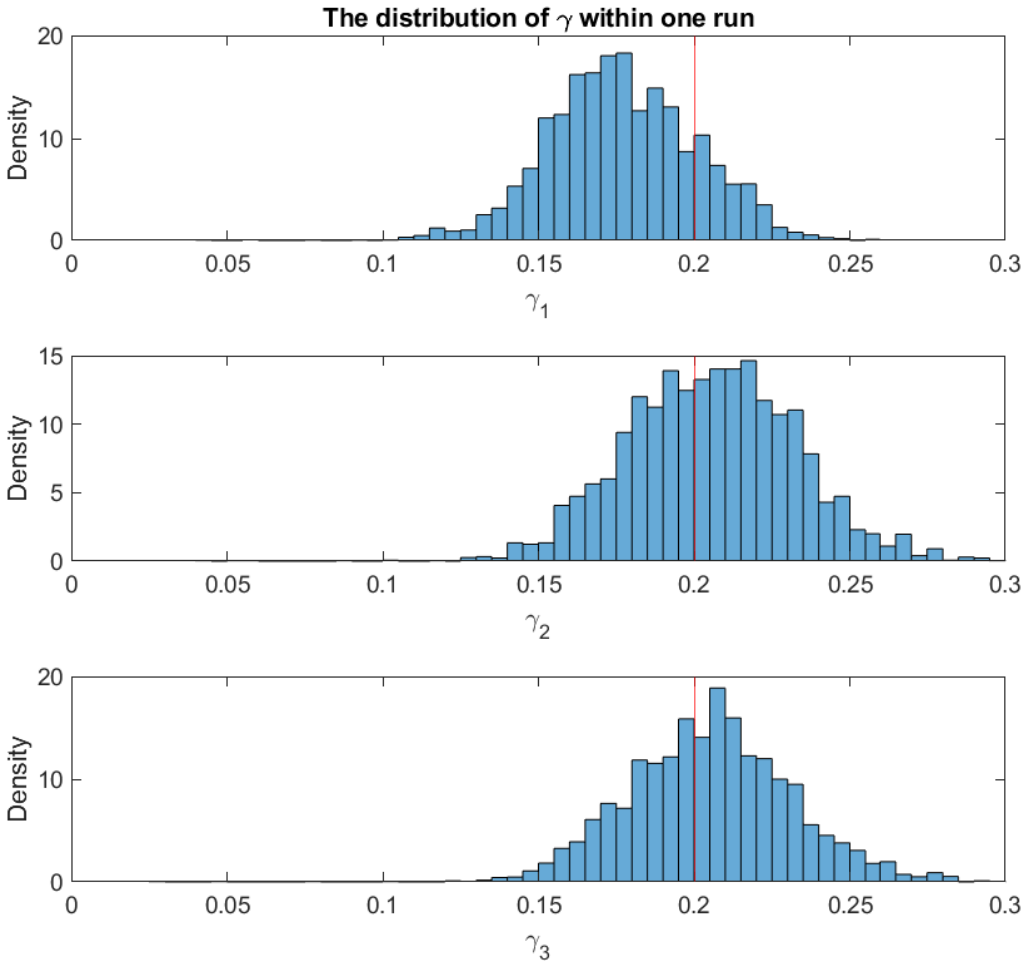

Figure 14: The distribution of γ_1, γ_2, and γ_3 within one arbitrary run.

MSV

Similar as in the AMSV part, B^*, Θ_i^*, and Θ_f^* are calculated, together with their standard deviations. The results for B are displayed in Figure 15 and the results for Θ are shown in Table 9, an analysis of the errors in B is made in Figure 16.

Comparing Figure 15 with 10, and Figure 16 with 11, similar results are observed. Both models estimate the factor loading matrix accurately, and similar behaviour is observed when comparing the estimation errors with the standard errors.

For the logarithmic volatility parameters in Table 9, the results are also comparable to Table 8. Again, (μ_i, μ_f) are within two standard errors of the true value, (ϕ_i, ϕ_f) are within 5 standard errors and the (σ_i, σ_f) pair are about 6 standard errors from their true value. It is also noteworthy that ϕ is slightly underestimated using both the AMSV model and the MSV model, while both models estimate σ a bit too high. Each time the error is in the direction of the prior mean. [Chib et al., 2006] mentions that the factor loadings for B and μ are accurately estimated (accuracy around 0.98) while ϕ and σ record somewhat larger deviations (accuracy around 0.90). The same is observed in this thesis.

It can be concluded that adding the γ-parameter in the model, only influences the σ_f-parameters, which get overestimated a bit more.

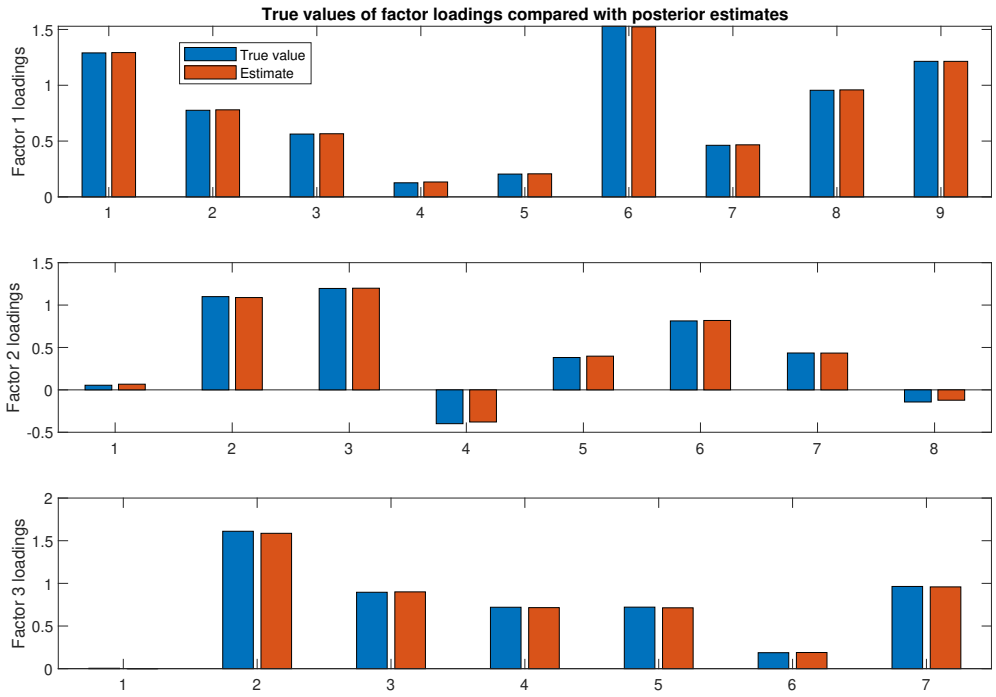

Figure 15: The parameters of the factor loading matrix. The values used for the generation of the data are shown under "True Value". The values under "Estimate" are the average of posteriors of 40 simulations using 5.000 samples (after 1.000 discarded burn-in samples), generated by the MSV model.

Parameter	True Value	Posterior mean	Posterior s.d.
μ_i	-9	-8.97	0.18
μ_f	-9	-9.00	0.18
ϕ_i	0.98	0.965	0.035
ϕ_f	0.98	0.972	0.011
σ_i	0.14	0.161	0.024
σ_f	0.14	0.160	0.024

Table 9: The parameters of the logarithmic volatility. The values used for the generation of the data are shown under "True Value". The values under "Posterior mean" and "Posterior s.d." are respectively the average and the standard deviation of posteriors of 40 simulations using 5.000 samples (after 1.000 discarded burn-in samples), generated by the MSV model.

Finally, in Figure 17 the distribution of one of the factor loadings and the logarithmic volatility parameters of one of the factors are shown, with the sampling path displayed in Figure 18. In this figure, it can again be seen that the sampler has reached the equilibrium posterior distribution.

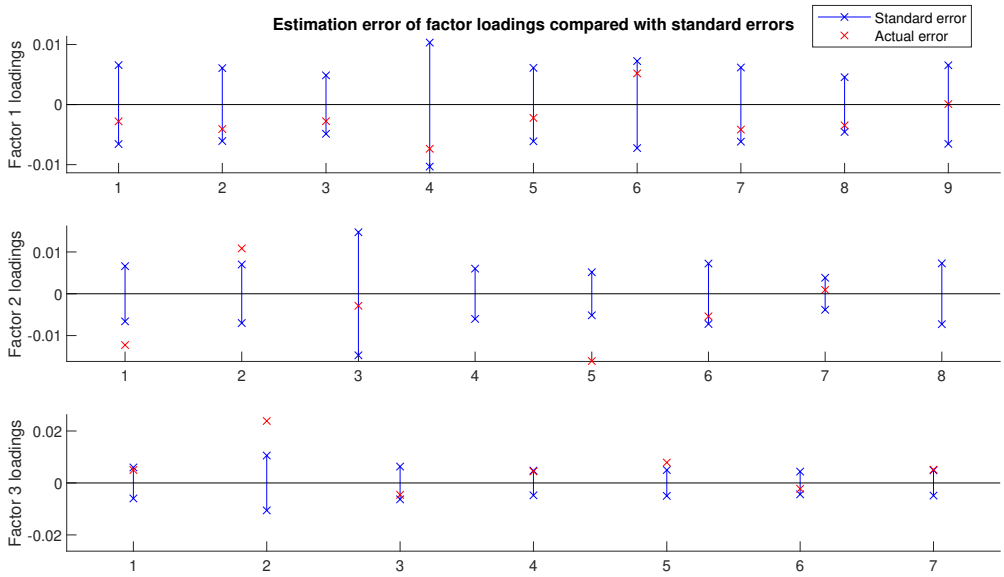

Figure 16: The standard errors and the actual errors of the values of the factor loading matrix. These values are the result of 40 simulations using 5.000 samples (after 1.000 discarded burn-in samples), generated by the MSV model.

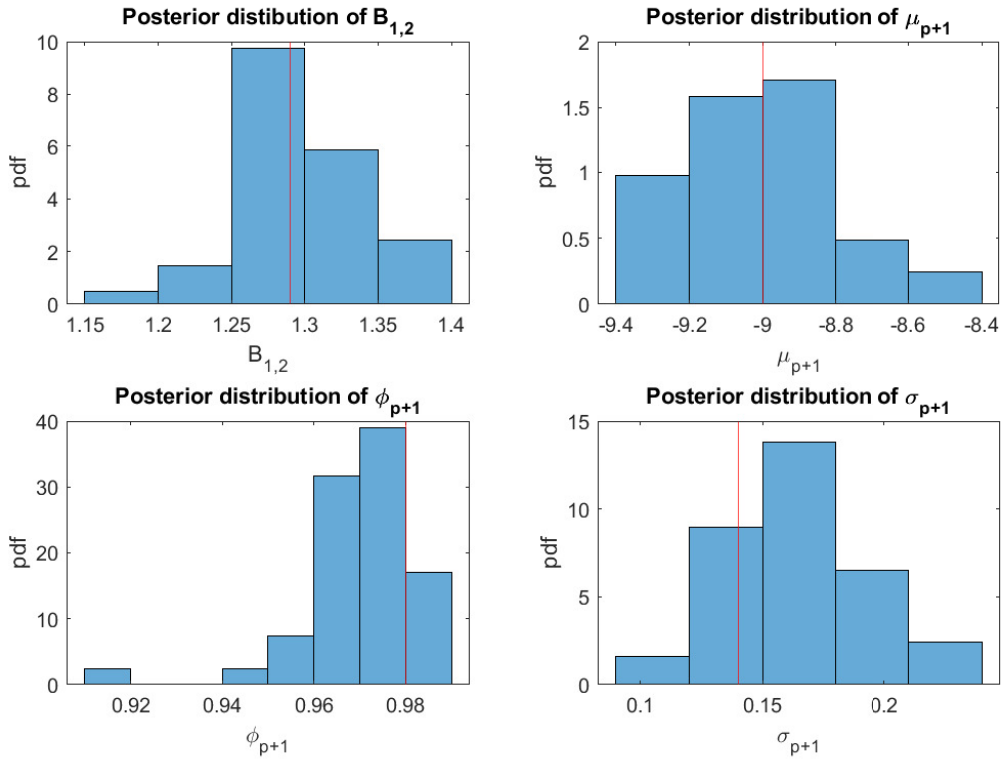

Figure 17: Histogram of the posterior values for one of the factor loadings and the parameters for the first factor. For this histogram, 40 data sets using 9.000 samples are used. The data is generated by the MSV model.

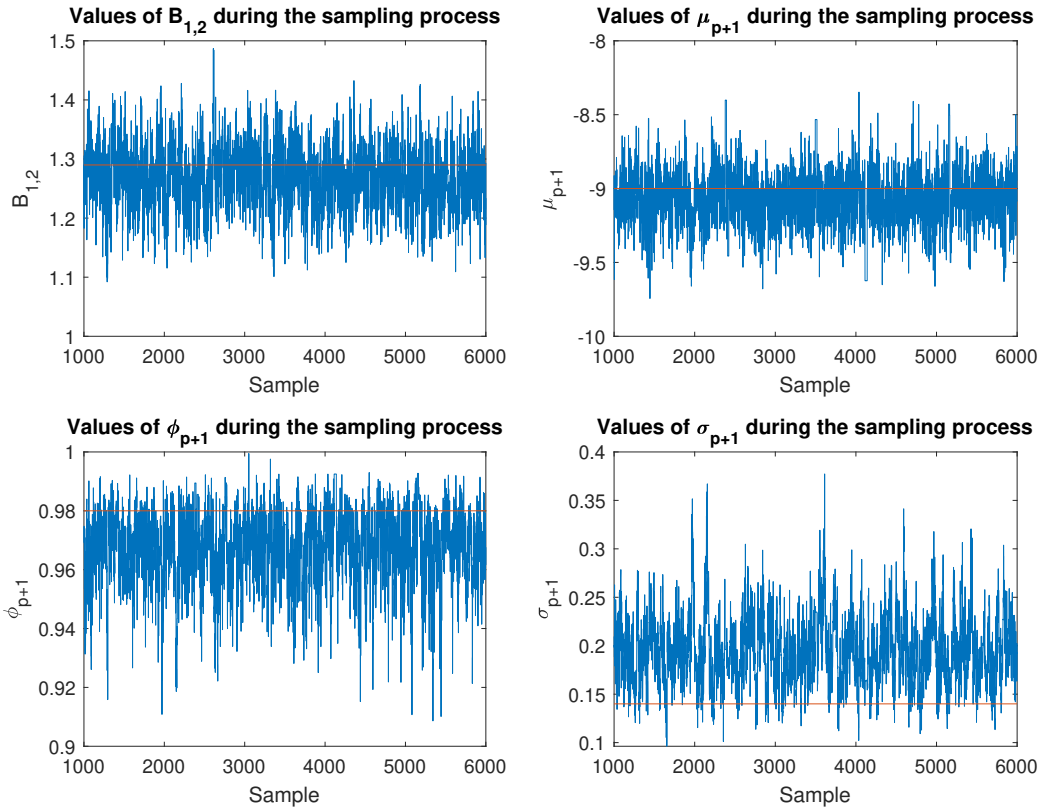

Figure 18: Path of the sampling of the parameters in Figure 17 of an arbitrary run using the MSV model.

MSVJ

The results for the factor loading matrix and the logarithmic volatility parameters are shown without further comments in appendix C. The behaviour of the sampling of the B and Θ parameters is similar to the behaviour in the AMSV and MSV sections. The only noteworthy difference is that B is estimated a bit worse.

Instead, in this section, a quick analysis of the distribution of the jump parameters κ and δ is made.

The grand posterior average for κ during the samples is 0.023 with a standard deviation of 0.013. Sufficiently close to the true value 0.02. For δ the grand posterior average is 0.0597 with a standard deviation of 0.0088, again very close to its true value 0.06. In Figure 19, the beta-like shape can be seen in the samples for κ

In [Chib et al., 2006], it is discussed that the prior has quite a large influence on the final estimate of the parameters for the jump parameters (especially for κ). For this reason, a prior which has a large probability mass at the true value used to set up the model is added. In Figure 19 the used prior is added to the distribution as well. In both cases, the samples are shifted away from the prior towards the true value of the parameters. Confirming that the choice of prior with high probability mass at the true values is sufficient to have good estimates for the parameters.

In Figure 20, the sampling path of an arbitrary run is shown. For κ and δ, the MSVJ sampler has also reached the equilibrium state after 1.000 initial iterations (also confirmed for the factor loading and logarithmic volatility parameters in Appendix C).

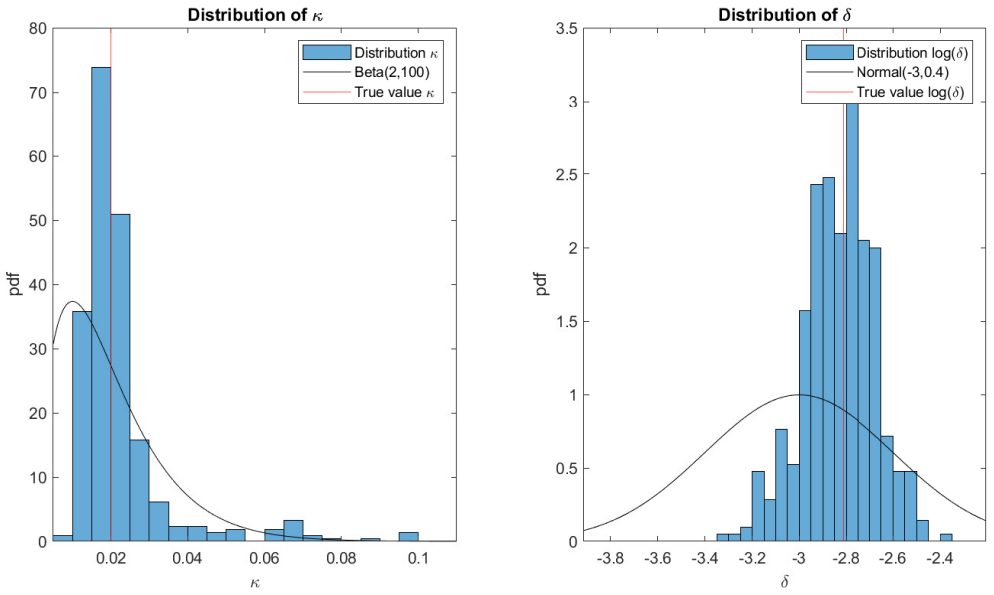

Figure 19: Histogram of the posterior values for δ and κ compared with the prior distribution of these parameters. For this histogram, 40 data sets using 9.000 samples are used.

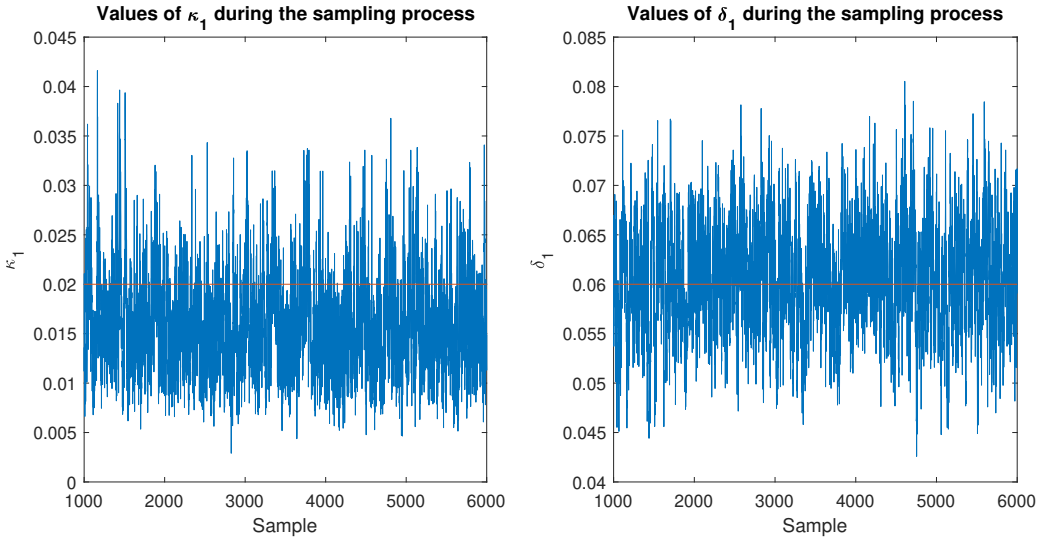

Figure 20: Path of the sampling of the parameters in Figure 19 of an arbitrary run.

7.3. Model comparison

Data from AMSV

In this section, the following sequence of steps is gone through:

- Generate a data set using one of the models.
- Sample the data sets with samplers from multiple models.
- Compute the Bayes factor b.

The models sampled with are again MSV, AMSV, and MSVJ. First, the results of all the samplers when using data generated with the AMSV model are analyzed. Later, the same is done with data coming from the MSV and MSVJ models.

Ten different synthetic data sets are generated using the AMSV model. The results after sampling, reduced sampling and running the particle filter are summarized in Table 10 and 11.

Run	$\log(p(y\|\mathcal{M}_{\text{AMSV}}))$	$\log(p(y\|\mathcal{M}_{\text{MSVJ}}))$	b
1	40953	40976	157.3
2	40441	40147	294.0
3	40924	40762	161.3
4	41431	41235	194.9
5	40687	40458	228.4
6	41377	41186	190.7
7	40891	40733	158.0
8	40618	40459	163.1
9	41287	41132	154.3
10	41155	40979	176.1

Table 10: The bayes factor between AMSV and MSVJ, when the underlying data is generated by the AMSV model.

Run	$\log(p(y\|\mathcal{M}_{\text{AMSV}}))$	$\log(p(y\|\mathcal{M}_{\text{MSV}}))$	b
1	40953	40910	42.6
2	40441	40355	85.4
3	40924	40855	68.7
4	41431	41337	93.0
5	40687	40614	72.6
6	41377	41316	61.1
7	40891	40840	50.8
8	40618	40518	100.4
9	41287	41232	55.2
10	41155	41072	82.9

Table 11: The bayes factor between AMSV and MSV, when the underlying data is generated by the AMSV model.

In Table 10 and 11 it can be seen that the method explained in Section 6 is clearly decisive when comparing AMSV with MSVJ. When comparing AMSV with MSV the score is less decisive, but it is still in favour of the AMSV model every time. The reason that b is smaller between AMSV and MSV is that these two models are closer to each other, only differing in three parameters (for the given problem size). Meanwhile, AMSV and MSVJ differ in 23 parameters for the same problem size.

In Table 12, a detailed comparison for three runs between AMSV and MSV is made. In this breakdown, it can be seen why the value for b is positive: generally, the AMSV model has a lower prior value and a higher ordinate value (which acts as a penalty to b). However, the higher value for $\log(p(y|\mathcal{M},\psi^*))$ compensates for this and tips the score in favour to the AMSV model.

Run	Model	$\log(p(y\|\mathcal{M},\psi^*))$	$\log(p(\psi^*\|\mathcal{M}))$	$\log(p(\psi^*\|\mathcal{M},y))$	$\log(p(y\|\mathcal{M}))$
1	AMSV	41096.5	-25.4	118.0	40953.0
1	MSV	41038.7	-18.7	109.7	40795.7
1	Δ	57.7	-6.7	8.34	31.2
2	AMSV	40587.6	-25.7	120.6	40441.3
2	MSV	40485.4	-18.9	110.6	40355.9
2	Δ	102.2	-6.8	10.0	85.4
3	AMSV	41068.6	-26.5	118.1	40924.0
3	MSV	40983.9	-20.9	107.7	40855.3
3	Δ	84.6	-10.8	10.4	68.7

Table 12: A breakdown of the bayes factor of three runs between AMSV and MSV. The underlying data is generated by the AMSV model.

This part of the section is concluded by showing what happens when trying to fit the jump parameters to the AMSV model. As discussed before, the parameters tend to go to their priors. In these simulations a total of 200 values of δ and κ are collected. Theoretically they should be zero (since the data is generated without jumps), however in Figure 21 it can be seen that these values are still highly influenced by the prior. The mean value found for κ and δ are 0.054 and 0.0093 with a standard deviation of 0.023 and 0.0056 respectively.

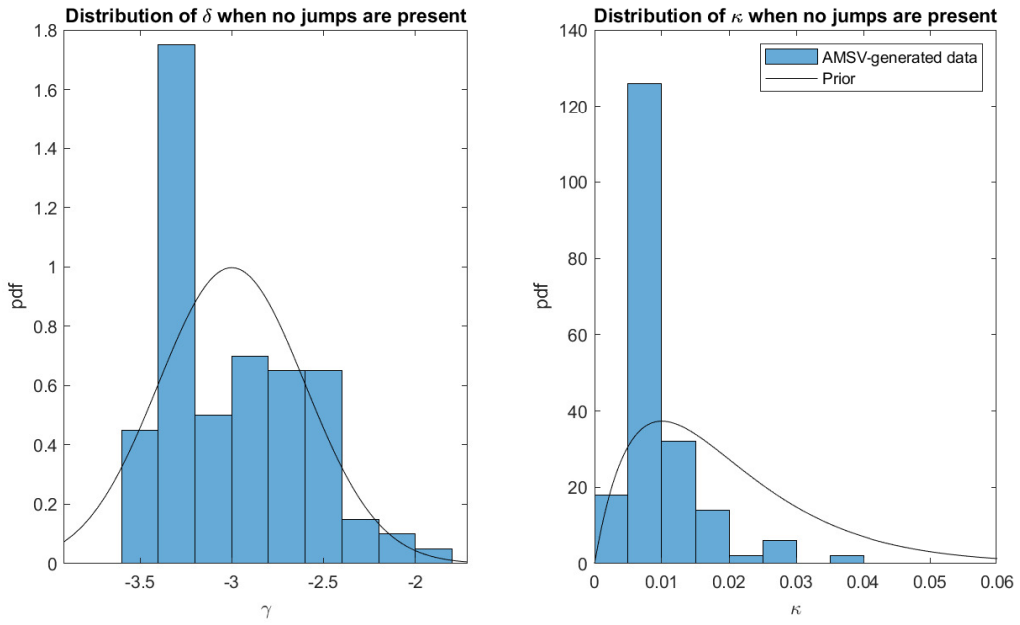

Figure 21: Histogram of values of δ and κ found when no jumps are in the model used to simulate underlying data. The prior used to sample the data is added as well.

Data from MSV

In the previous section, it is shown that the method from Section 6 scores in favour of the AMSV model against the MSV model, when the underlying data is indeed generated by an AMSV model. In this section, the reverse is looked into. Ten different synthetic data sets are generated using the MSV model. The results after sampling, reduced sampling, and running the particle filter are again compared between MSV and AMSV, summarized in Table 13.

Run	$\log(p(y\|\mathcal{M}_{\text{MSV}}))$	$\log(p(y\|\mathcal{M}_{\text{AMSV}}))$	b
1	40977	40967	10.5
2	40427	40407	19.7
3	42188	42175	12.7
4	41175	41134	41.7
5	41130	41114	16.5
6	41362	41348	13.5
7	41029	41000	28.5
8	40165	40142	22.7
9	41563	41535	28.2
10	40971	40953	17.3

Table 13: The bayes factor between MSV and AMSV, when the underlying data is generated by the MSV model.

As can be seen in this table, when there is no asymmetry in the data, the method from Section 6 scores in favour of the MSV model. The values of b are even smaller this time. The reason that the likelihood of the models is closer this time, is that the γ parameters are estimated at or close to zero. This leads to similar results from the particle filter. The only distinguishing factors now are the priors and the posterior ordinate penalty, in Table 12

it can be seen that the difference between these values are typically in the range of 10. Combining the results of Tables 11 and 13, a reference can be created. This reference will be used when applying the different models to real equity returns in Section 8.

In Figure 22, the distribution from the 30 values of γ found in the above simulation can be found. In the left half of this figure, it can be seen that compared to the prior, γ is centered around 0. The right half figure is provided to have a better look at the actual distribution of γ. This shows that the model doesn't give a significant γ value when no asymmetry is present in the data.

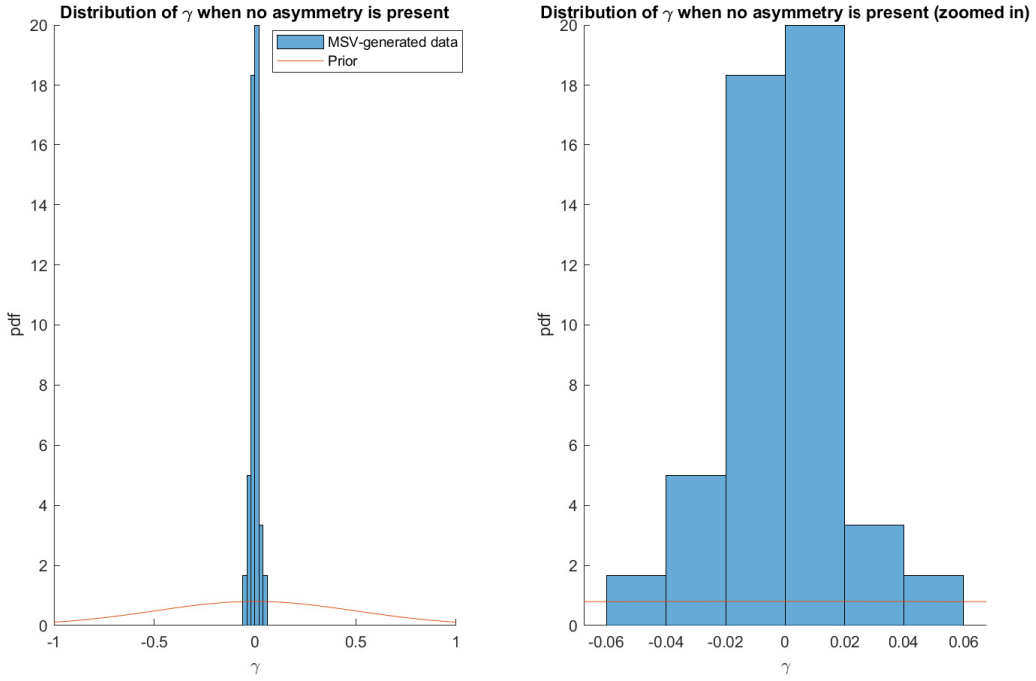

Figure 22: Histogram of values of γ found when no asymmetry is in the model used to simulate underlying data. The prior is added to the figure as well. The mean of the MSV-generated data is -0.001 with a standard deviation of 0.021.

Data from MSVJ

Finally, ten different synthetic data sets are generated using the MSVJ model. The results after sampling, reduced sampling and running the particle filter are summarized in Table 14.

Run	$\log(p(y\|\mathcal{M}_{\text{MSVJ}}))$	$\log(p(y\|\mathcal{M}_{\text{AMSV}}))$	b
1	39078	38620	457.3
2	38830	38585	245.4
3	39318	38731	587.0
4	39468	38983	484.6
5	39928	39611	317.3
6	39541	38939	602.1
7	40554	39739	815.0
8	39724	39003	721.2
9	39121	38378	742.7
10	40365	39741	615.6

Table 14: The bayes factor between MSVJ and AMSV, when the underlying data is generated by the MSVJ model.

Just as in Table 10, using the Bayes method one can correctly identify whether asymmetry or jumps are in the data.

The values for γ found when sampling a data set that has jumps, but no asymmetry is analyzed as well. The distribution of these values of γ can be seen in Figure 23. Similar to the MSV-generated data, no significant asymmetry parameter is found.

Another noteworthy observation that can be made when using the AMSV method to sample MSVJ-generated data, is that the logarithmic volatility parameters for the innovations do not get estimated correctly anymore. Generally, higher values for μ_i and σ_i, and lower values for ϕ_i are found. A probable explanation for this is that the jumps get estimated using the logarithmic volatility, which generates a higher mean value for the logarithmic volatility (higher μ). This also requires the volatility to move more quickly to cover for the jumps (higher σ). Finally, the volatility is less dependent on the previous value of the volatility since the jumps are independent at each time step (lower ϕ). This observation is also true for the MSV model.

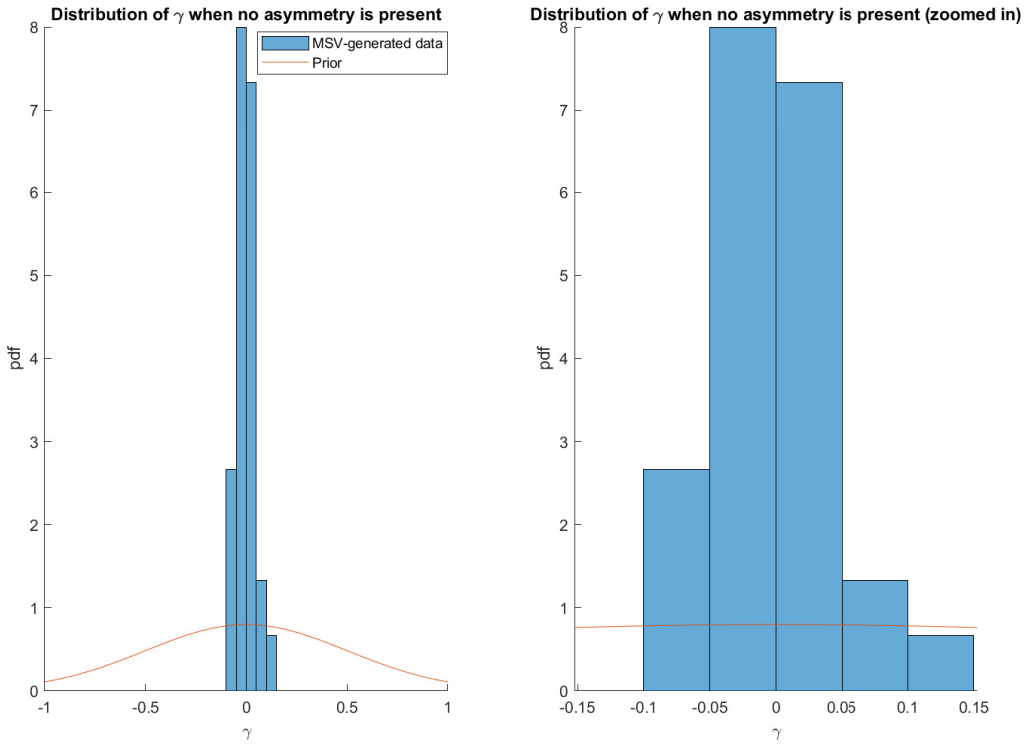

Figure 23: Histogram of values of γ found when jumps are but asymmetry is not in the model used to simulate underlying data. The prior is added to the figure as well. The mean for γ of the MSVJ-generated data is -0.003 with a standard deviation of 0.038.

APPLICATION TO REAL EQUITY RETURNS

8.1. THE DATA SET

In this section, the AMSV model will be tested on real equity returns. Afterward, a comparison will be made between the AMSV, MSV, and MSVJ models. First, the data set gets introduced.

Similar to section 3, a set of assets from the SP500 are selected. These assets are selected to have some higher market caps, some medium market caps, and some lower market caps. The chosen assets for this analysis are:

- Apple Inc. (AAPL)
- Southwest Airlines (LUV)
- Northrop Grumman (NOC)
- Microsoft (MSFT)
- Mastercard (MA)
- Walmart (WMT)
- Progressive Corporation (PGR)
- Air Products (APD)
- First Republic Bank (FRC)
- Dr Pepper Snapple Group (KDP)

This will also be the order in which they will be selected, this will be important when analyzing the factor loading matrix. The stock prices will be analyzed starting from 2015-01-01 until 2021-01-01, this period contains 1510 trading days. In Figure 24, the prices of the selected stocks are displayed for this period.

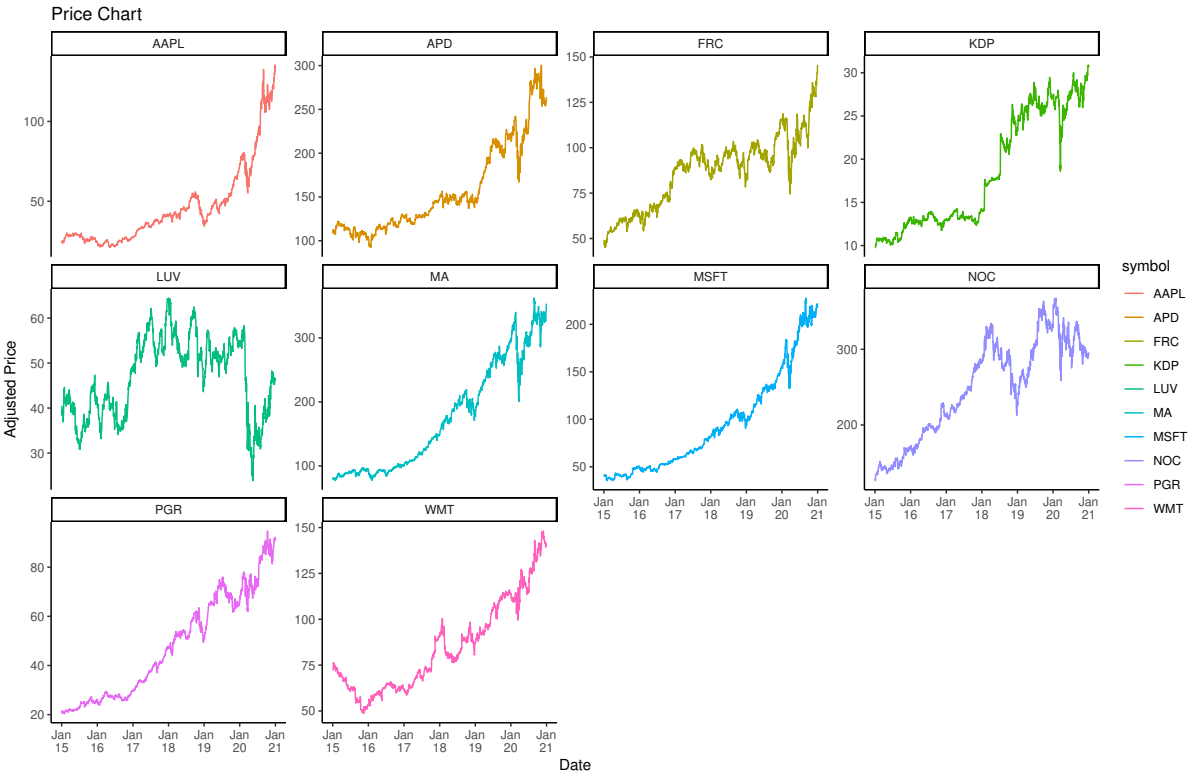

Figure 24: Prices of ten selected assets from 2015-01-01 until 2021-01-01.

Let a_t^j be the asset price of asset j at time t, then the returns analyzed are:

$$y_t^{*j} = \log(\frac{a_t^j}{a_{t-1}^j}) \tag{106}$$

$$y_t^j = y_t^{*j} - \frac{1}{N}\sum_{i=1}^{N} y_i^{*j} \tag{107}$$

So the analysis is made on the log-returns of the asset, with the mean removed.

8.2. PRIORS

Using the MCMC method, the following priors are taken:

For the free elements of B, $N(1,9)$ is taken as a prior. For the logarithmic volatility parameters, the prior of μ_j is set at $N(-9,25)$, ϕ_j is distributed by simulating a transformed version: ϕ_j^* from a beta distribution and then taking $\phi_j = 2\phi_j^* - 1$. This is done in such a way that the mean of $\phi_j = 0.84$ and its standard deviation is 0.11. σ_j is simulated from an Inverse Gamma distribution, such that its mean is 0.25 and its standard deviation 0.4. The prior for γ_j is still $N(0;0.25)$. When using MSVJ, for the jump parameters, $\log(\delta_j) \sim N(-3,1)$ as a prior for δ and Beta(2,100) as a prior for κ are taken. These priors indicate about 4 or 5 jumps a year (probability of a jump is 1.96% per day) with an intensity of about 5%.

The priors in this section have the same (or a similar) mean as in section 7. The distributions, however, are chosen to be much wider. The reason for these wider distributions is that not as much is known about this data set, as was the case for the synthetic data.

8.3. RESULTS AMSV MODEL

First, when trying to estimate the above data set with a model with $k = 3$ factors, only one significant factor is found. The other two factors are estimated to be more than 10 times less significant. Moreover, this leads to problems when estimating the optimal value for B. When some factors are insignificant compared to other factors, their influence is less noticeable. In turn, the likelihood function for B is barely influenced by the factor loadings for those factors, resulting in the optimizer having difficulties finding an optimal value for B. This is also the case using the MSV and the MSVJ models.

Instead, only one factor is used in this section. The results of the sampling using the AMSV model can be found in Table 15.

Asset	Factor 1 Loading	μ	ϕ	σ
AAPL	1 (fixed)	-8.94 (0.06)	0.639 (0.127)	0.295 (0.054)
LUV	0.753 (0.045)	-8.50 (0.13)	0.890 (0.026)	0.475 (0.089)
NOC	0.733 (0.033)	-9.35 (0.14)	0.882 (0.030)	0.455 (0.120)
MSFT	1.042 (0.032)	-9.52 (0.06)	0.703 (0.105)	0.268 (0.052)
MA	1.00 (0.033)	-9.98 (0.17)	0.868 (0.028)	0.638 (0.122)
WMT	0.496 (0.028)	-9.29 (0.06)	0.798 (0.070)	0.281 (0.055)
PGR	0.733 (0.031)	-9.37 (0.10)	0.869 (0.050)	0.298 (0.076)
APD	0.862 (0.035)	-9.43 (0.07)	0.834 (0.053)	0.271 (0.049)
FRC	0.791 (0.039)	-8.93 (0.11)	0.900 (0.025)	0.321 (0.075)
KDP	0.538 (0.028)	-9.16 (0.06)	0.747 (0.082)	0.324 (0.051)
Factor	μ	ϕ	σ	γ
1	-9.45 (0.20)	0.947 (0.012)	0.350 (0.041)	0.215 (0.036)

Table 15: The factor loadings and logarithmic volatility parameters for the simulation of the selected assets using the AMSV model. The standard deviations are also provided, between brackets.

The most important result in Table 15 is γ, which is significantly larger than 0. This means that asymmetric behaviour is indeed observed. In Figure 25, the sampling path and the posterior density of γ are shown. In these figures, it can be seen that the sampler found a stable equilibrium and that the minimum value of γ during the sampling process is about 0.1, which is a clear sign that the asymmetry is observed. Note that this

distribution looks similar to the distribution from Figure 14 from Section 7.2, where the synthetic data set with $\gamma = 0.2$ was sampled.

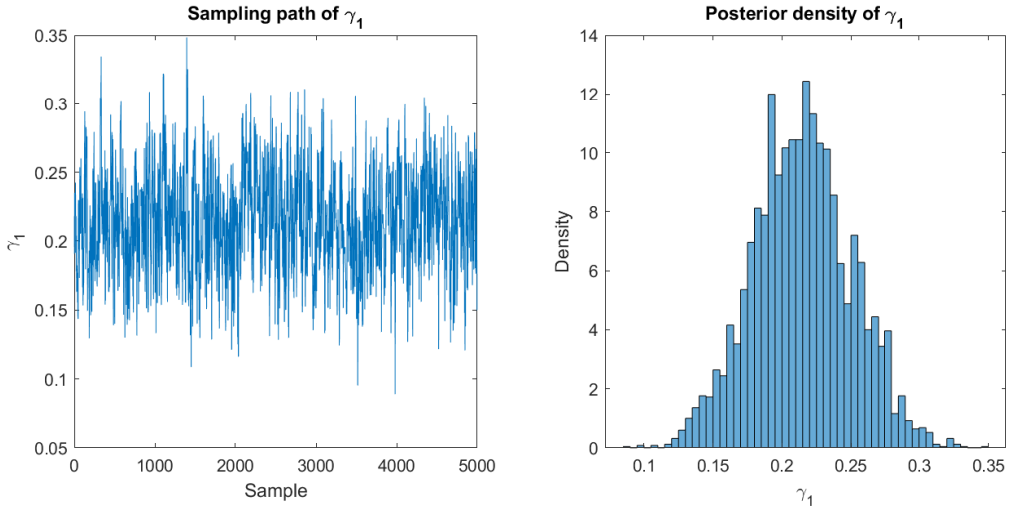

Figure 25: The sampling path and the posterior density of γ_1.

8.4. Comparison with MSV

Similar as in the previous section, first the simulation results using one factor are shown in Table 16, this time using the MSV model.

Asset	Factor 1 Loading	μ	ϕ	σ
AAPL	1 (fixed)	-8.93 (0.05)	0.631 (0.129)	0.297 (0.054)
LUV	0.777 (0.045)	-8.50 (0.13)	0.890 (0.026)	0.472 (0.089)
NOC	0.683 (0.032)	-9.36 (0.14)	0.882 (0.030)	0.466 (0.120)
MSFT	1.059 (0.032)	-9.52 (0.06)	0.700 (0.103)	0.267 (0.051)
MA	1.018 (0.032)	-9.98 (0.17)	0.866 (0.029)	0.641 (0.120)
WMT	0.496 (0.028)	-9.27 (0.06)	0.795 (0.073)	0.280 (0.053)
PGR	0.689 (0.030)	-9.36 (0.09)	0.868 (0.050)	0.293 (0.074)
APD	0.825 (0.035)	-9.43 (0.07)	0.831 (0.056)	0.270 (0.046)
FRC	0.745 (0.039)	-8.92 (0.10)	0.900 (0.026)	0.319 (0.074)
KDP	0.493 (0.028)	-9.16 (0.06)	0.750 (0.083)	0.325 (0.051)
Factor		μ	ϕ	σ
1		-9.50 (0.24)	0.955 (0.012)	0.367 (0.044)

Table 16: The factor loadings and logarithmic volatility parameters for the simulation of the selected assets using the MSV model. The standard deviations are also provided, between brackets.

Comparing Table 16 with Table 15, it is first noticed that the factor loadings are similar to the AMSV simulation results in Table 15. Moreover, all values for (μ, ϕ, σ) are close to each other for the two models as well (both the values and the standard deviations).

It can be concluded, that the addition of γ does not alter the factor loading matrix nor the other logarithmic volatility parameters. In Figure 26, the logarithmic volatility of the factor is compared between the MSV and the AMSV model. This is done by drawing $\{f_s, s_s | B^*, \Theta^*, h_s\}$ and subsequently $\{h_s | B^*, \Theta^*, f_s, s_s\}$ 1000 times, saving the factor values for h_s, initializing with the final value for h_s found when sampling the data using the AMSV and MSV model respectively. In Figure 26, the average of all the draws is shown. It can be seen that the factor volatility for AMSV and MSV are very close to each other. Showing that only the internal method of calculating the volatility is altered and not the volatility itself.

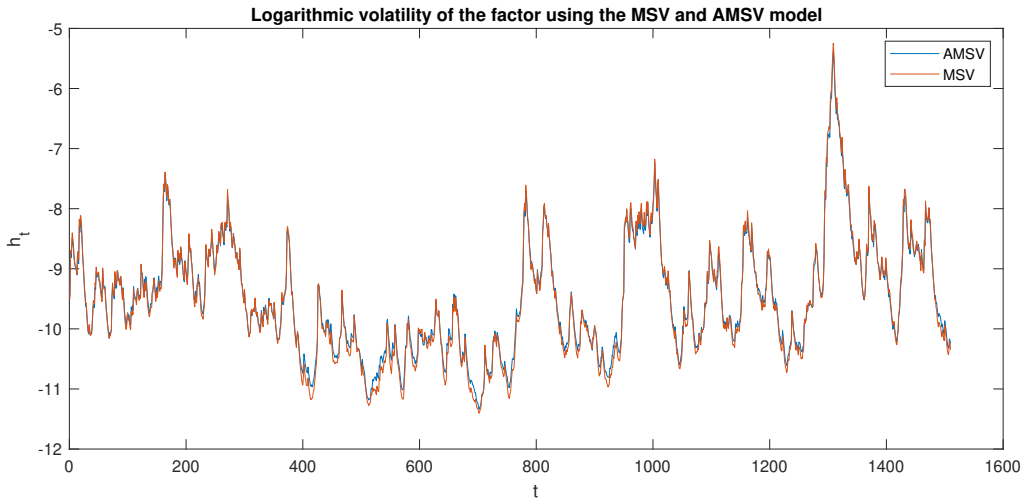

Figure 26: A comparison between the factor log volatility using the MSV and the AMSV model

The Bayes factor (from Section 6) between the two models is $b = 23.0$ in favour of the AMSV model. In Section 7.3, it was observed that when asymmetry is present, the Bayes criterion is in favour of the AMSV model, while when asymmetry is not present, it is in favour of the MSV model. In Figure 27 two empirical probability distribution functions for the Bayes factor are shown in one figure. For one of the probability distribution functions, the underlying data is generated by the MSV model, while for the other the underlying data is generated by the AMSV model[3]. In the figure, the value for the real data set in this section is added as well. This comparison isn't totally fair, since in the reference data three factors were used. However, it will still give a good indication.

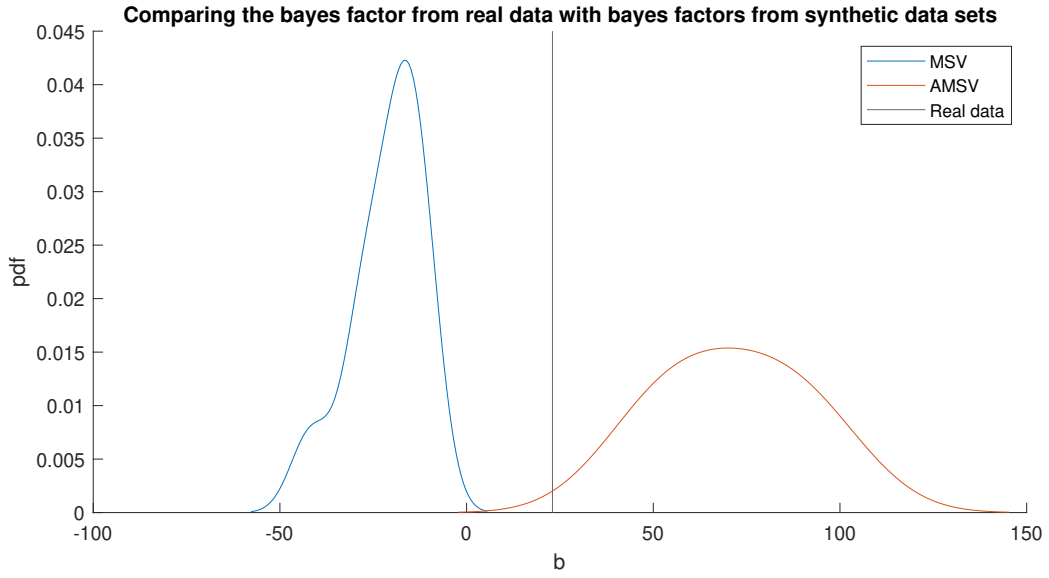

Figure 27: The empirical probability distribution functions for the bayes factor between the MSV and the AMSV model. The underlying data generated via either MSV or AMSV, as indicated by the legend. In this figure, a negative value is in favour of MSV, while a positive value is in favour of AMSV. The value for the real data set is added to the figure as well. The empirical probability distribution functions are estimated by the KS algorithm.

[3]These are the results from Section 7.3.

8.5. Comparison with MSVJ

Again, as in the previous sections, first the simulation results using one factor are shown in Table 17, this time using the MSVJ model.

Asset	Factor 1 Loading	μ	ϕ	σ	κ	δ
AAPL	1 (fixed)	-9.23 (0.12)	0.923 (0.029)	0.272 (0.074)	0.027 (0.010)	0.0409 (0.0043)
LUV	0.814 (0.045)	-8.62 (0.25)	0.973 (0.012)	0.205 (0.043)	0.038 (0.011)	0.0464 (0.0047)
NOC	0.705 (0.035)	-9.50 (0.20)	0.946 (0.030)	0.311 (0.095)	0.032 (0.012)	0.0303 (0.0037)
MSFT	1.083 (0.036)	-9.98 (0.15)	0.900 (0.030)	0.453 (0.118)	0.016 (0.006)	0.0457 (0.0054)
MA	1.074 (0.037)	-9.98 (0.15)	0.891 (0.038)	0.510 (0.144)	0.013 (0.009)	0.0340 (0.0083)
WMT	0.475 (0.029)	-9.56 (0.11)	0.927 (0.022)	0.274 (0.056)	0.019 (0.006)	0.0517 (0.0048)
PGR	0.717 (0.032)	-9.61 (0.14)	0.940 (0.025)	0.267 (0.066)	0.022 (0.007)	0.0375 (0.0042)
APD	0.827 (0.035)	-9.76 (0.12)	0.912 (0.028)	0.352 (0.094)	0.018 (0.008)	0.0379 (0.0050)
FRC	0.786 (0.042)	-9.03 (0.12)	0.924 (0.023)	0.302 (0.059)	0.015 (0.008)	0.0433 (0.0064)
KDP	0.426 (0.031)	-9.48 (0.12)	0.915 (0.027)	0.331 (0.082)	0.018 (0.007)	0.0696 (0.0073)

Factor	μ	ϕ	σ
1	-9.67 (0.25)	0.958 (0.011)	0.367 (0.042)

Table 17: The factor loadings and logarithmic volatility parameters for the simulation of the selected assets using the MSVJ model.

The Bayes factor (from Section 6) between the two models is $b = 262.9$ in favour of the MSVJ model. In Figure 28 two empirical distribution functions similar to Figure 27 are shown. Again using the results from Section 7.3.

Moreover, in the MSVJ part of Section 7.3 it was noticed that when sampling data that contains jumps with an AMSV model (or an MSV model), the μ and σ parameters of the innovations get estimated too high, while the ϕ parameters of the innovations are estimated too low. Comparing Table 17 with Table 15, this is exactly what is found. Indicating that the jumps in the data distort the volatility parameters for the innovations in the AMSV and MSV models.

These two insights, lead to the conclusion that the jump model outperforms the asymmetric factor model for this data set.

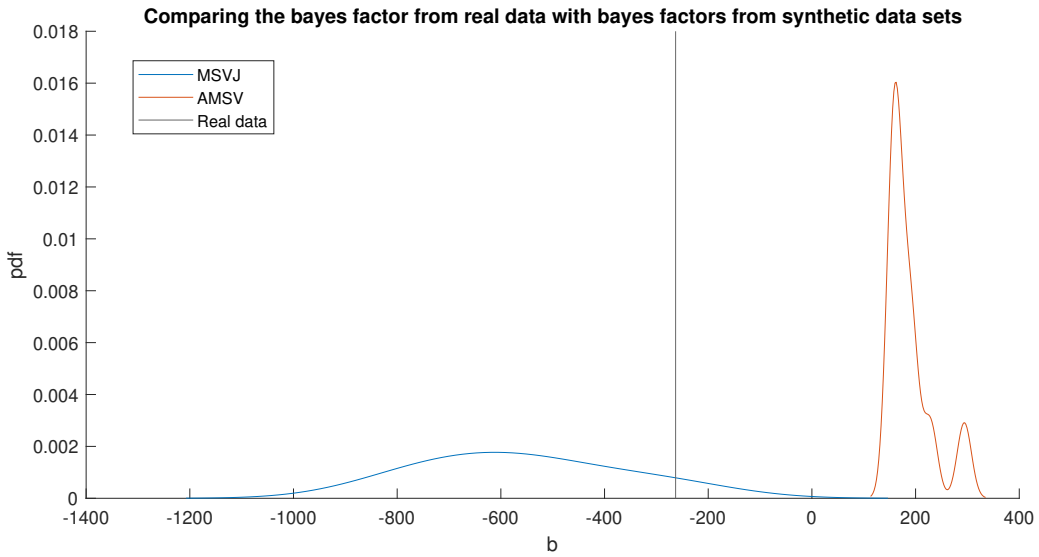

Figure 28: The empirical probability distribution functions for the bayes factor between the MSV and the AMSV model. The underlying data generated via either MSV or AMSV, as indicated by the legend. In this figure, a negative value is in favour of MSV, while a positive value is in favour of AMSV. The value for the real data set is added to the figure as well. The empirical probability distribution functions are estimated by the KS algorithm.

CONCLUSION AND DISCUSSION

In this thesis, a multivariate model is extended to contain an asymmetric relation between asset returns and the volatility of these assets. In section 7.2, it was shown that the sampler provided is able to successfully estimate the parameters of the extension. In section 7.3 it was shown that (i) the model is distinguishable from the classical factor model and the factor model with jump extensions, and (ii) that the asymmetric extension correctly gets estimated at zero when this asymmetry is non-present in the underlying data. Finally, in Section 8 it is shown that the asymmetrical component is observed in real equity time series. Moreover, the method provided in Section 6 prefers the factor model with asymmetry extension over the classical factor model. However, the model with jump extensions still performs best.

The results from this thesis, implicate that the observed asymmetry in the data in Section 3 can be modelled by applying an asymmetric term to the factors. It also shows that implementing the asymmetry gives an improvement over the classical model, by implementing relatively few extra parameters.

However, it was perhaps a bit far-fetched and optimistic to expect the relatively simple asymmetric extension to compete with the more complex jump extension. Although it is known that the model with jump extension requires a well-thought-out prior, it still is the preferred algorithm. It could be the case that a combination between the two models would perform even better, however, that would require additional research, since the jump extension alters the logarithmic volatility update equations quite a bit, as was seen in Sections 7.3 and 8.

The most limiting factor in this thesis is the number of samples for the Bayes factor simulated in Section 7.3. While 40 runs in Section 7.2 seem enough, the 10 values for the Bayes factor between each combination of two models are not enough to have a full overview of the distribution of this parameter. The reason not more of these simulations are performed, is that they are very time-consuming, especially the reduced runs for the MSVJ algorithm take a long time (between 1 and 2 days). This is partly due to the choice of programming software. To have a better overview, 40 values of b are recommended.

In addition to the factor model with jumps and asymmetric factors, other further research recommendations include the addition of more heavy-tailed distributions to the innovations and/or factors using the asymmetric extensions, and a better way to estimate the logarithmic volatility given the parameters for its equation, which was identified as the weakest part in the Monte Carlo Markov chain algorithm in Section 5.2.

It would also be interesting to see if the asymmetric model from this thesis outperforms a model that captures the asymmetry in the innovations. Especially since fewer extra parameters are necessary to capture the asymmetry in the factors. Using the methods in this thesis, this could be easily tested and does not necessarily require new research.

BIBLIOGRAPHY

[Albert and Hu, 2019] Albert, J. and Hu, J. (2019). *Probability and Bayesian Modeling*. Chapman and Hall/CRC.

[Asai and McAleer, 2006] Asai, M. and McAleer, M. (2006). Asymmetric multivariate stochastic volatility. *Econometric Reviews*, 25(2-3):453–473.

[Asai and McAleer, 2011] Asai, M. and McAleer, M. (2011). Alternative asymmetric stochastic volatility models. *Econometric Reviews*, 30(5):548–564.

[Asai et al., 2006] Asai, M., McAleer, M., and Yu, J. (2006). Multivariate stochastic volatility: A review. *Econometric Reviews*, 25(2-3):145–175.

[Bollerslev, 1986] Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of econometrics*, 31(3):307–327.

[Chan et al., 2005] Chan, D., Kohn, R., and Kirby, C. (2005). Multivariate stochastic volatility models with correlated errors. *Econometric Reviews*, 25:245–274.

[Chen et al., 2019] Chen, H., Zhang, J., Tao, Y., and Tan, F. (2019). Asymmetric garch type models for asymmetric volatility characteristics analysis and wind power forecasting. *Protection and Control of Modern Power Systems*, 4(1):1–11.

[Chib, 1995] Chib, S. (1995). Marginal likelihood from the gibbs output. *Journal of the american statistical association*, 90(432):1313–1321.

[Chib and Greenberg, 1995] Chib, S. and Greenberg, E. (1995). Understanding the metropolis-hastings algorithm. *The American Statistician*, 49(4):327–335.

[Chib and Jeliazkov, 2001] Chib, S. and Jeliazkov, I. (2001). Marginal likelihood from the metropolis–hastings output. *Journal of the American statistical association*, 96(453):270–281.

[Chib et al., 2002] Chib, S., Nardari, F., and Shephard, N. (2002). Markov chain monte carlo methods for stochastic volatility models. *Journal of Econometrics*, 108(2):281–316.

[Chib et al., 2006] Chib, S., Nardari, F., and Shephard, N. (2006). Analysis of high dimensional multivariate stochastic volatility models. *Journal of Econometrics*, 134(2):341–371.

[Christie, 1982] Christie, A. A. (1982). The stochastic behavior of common stock variances: Value, leverage and interest rate effects. *Journal of Financial Economics*, 10(4):407–432.

[Danielsson, 1998] Danielsson, J. (1998). Multivariate stochastic volatility models: Estimation and a comparison with vgarch models. *Journal of Empirical Finance*, 5(2):155–173.

[De Jong and Shephard, 1995] De Jong, P. and Shephard, N. (1995). The simulation smoother for time series models. *Biometrika*, 82(2):339–350.

[Ding et al., 1993] Ding, Z., Granger, C. W., and Engle, R. F. (1993). A long memory property of stock market returns and a new model. *Journal of Empirical Finance*, 1(1):83–106.

[Dodge, 2008] Dodge, Y. (2008). *The Concise Encyclopedia of Statistics*, pages 283–287. Springer, New York, NY.

[Doucet et al., 2001] Doucet, A., De Freitas, N., Gordon, N. J., et al. (2001). *Sequential Monte Carlo methods in practice*, volume 1. Springer.

[Ghysels et al., 1996] Ghysels, E., Harvey, A. C., and Renault, E. (1996). Handbook of statistics. In *Statistical Methods in Finance*, volume 14, pages 119–191. Elsevier.

[Glosten et al., 1993] Glosten, L. R., Jagannathan, R., and Runkle, D. E. (1993). On the relation between the expected value and the volatility of the nominal excess return on stocks. *The Journal of Finance*, 48(5):1779–1801.

[Harvey and Shephard, 1996] Harvey, A. C. and Shephard, N. (1996). Estimation of an asymmetric stochastic volatility model for asset returns. *Journal of Business & Economic Statistics*, 14(4):429–434.

[Huang et al., 2008] Huang, G. P., Mourikis, A. I., and Roumeliotis, S. I. (2008). Analysis and improvement of the consistency of extended kalman filter based slam. In *2008 IEEE International Conference on Robotics and Automation*, pages 473–479. IEEE.

[Kim et al., 1998] Kim, S., Shephard, N., and Chib, S. (1998). Stochastic volatility: likelihood inference and comparison with arch models. *The review of economic studies*, 65(3):361–393.

[Nelson, 1991] Nelson, D. B. (1991). Conditional heteroskedasticity in asset returns: A new approach. *Econometrica: Journal of the econometric society*, pages 347–370.

[Petersen and Pedersen, 2012] Petersen, K. B. and Pedersen, M. S. (2012). *The Matrix Cookbook*. `http://matrixcookbook.com`.

[Pitt, 2002] Pitt, M. K. (2002). Smooth particle filters for likelihood evaluation and maximisation. The Warwick Economics Research Paper Series (TWERPS) 651, University of Warwick, Department of Economics.

[Pitt and Shephard, 1999] Pitt, M. K. and Shephard, N. (1999). Filtering via simulation: Auxiliary particle filters. *Journal of the American statistical association*, 94(446):590–599.

[Shumway and Stoffer, 2000] Shumway, R. H. and Stoffer, D. S. (2000). *Time series analysis and its applications*, volume 3. Springer.

[Silverman, 2018] Silverman, B. W. (2018). *Density estimation for statistics and data analysis*. Routledge.

[Tsiotas, 2012] Tsiotas, G. (2012). On generalised asymmetric stochastic volatility models. *Computational Statistics And Data Analysis*, 56(1):151–172.

[Wan, 2006] Wan, E. (2006). Sigma-point filters: An overview with applications to integrated navigation and vision assisted control. In *2006 IEEE Nonlinear Statistical Signal Processing Workshop*, pages 201–202.

A

SECOND DERIVATIVE OF LOGARITHMIC LIKELIHOOD OF FACTOR LOADING MATRIX

$$
\begin{aligned}
H_{ij,mn} &= \frac{\delta^2 l}{\delta B_{ij}\delta B_{mn}} \\
&= \frac{\delta}{\delta B_{mn}}\left(\sum_{t=1}^{n} Tr\left(y_t y_t'\Omega_t^{-1}BD_t\frac{\delta B}{\delta B_{ij}}'\Omega_t^{-1} - \Omega_t^{-1}BD_t\frac{\delta B}{\delta B_{ij}}'\right)\right) \\
&= \sum_{t=1}^{n} Tr\left(\frac{\delta}{\delta B_{mn}}\left(y_t y_t'\Omega_t^{-1}BD_t\frac{\delta B}{\delta B_{ij}}'\Omega_t^{-1}\right) - \frac{\delta}{\delta B_{mn}}\left(\Omega_t^{-1}BD_t\frac{\delta B}{\delta B_{ij}}'\right)\right)
\end{aligned}
$$

Now the two terms are calculated separately, starting with the second term within the sum:

$$
\begin{aligned}
H_{ij,mn}^{1,t} &= \frac{\delta\left(\Omega_t^{-1}BD_t\frac{\delta B}{\delta B_{ij}}'\right)}{\delta B_{mn}} \\
&= \frac{\delta\Omega_t^{-1}}{\delta B_{mn}}BD_t\frac{\delta B}{\delta B_{ij}}' + \Omega_t^{-1}\frac{\delta B}{\delta B_{mn}}D_t\frac{\delta B}{\delta B_{ij}}' \\
&= -\Omega_t^{-1}\frac{\delta\Omega_t}{\delta B_{mn}}\Omega_t^{-1}BD_t\frac{\delta B}{\delta B_{ij}}' + \Omega_t^{-1}\frac{\delta B}{\delta B_{mn}}D_t\frac{\delta B}{\delta B_{ij}}' \\
&= -\Omega_t^{-1}\left(\frac{\delta B}{\delta B_{mn}}D_tB' + BD_t\frac{\delta B}{\delta B_{mn}}'\right)\Omega_t^{-1}BD_t\frac{\delta B}{\delta B_{ij}}' + \Omega_t^{-1}\frac{\delta B}{\delta B_{mn}}D_t\frac{\delta B}{\delta B_{ij}}'
\end{aligned}
$$

The first term gets split into two terms again:

$$
\frac{\delta}{\delta B_{mn}}\left(y_t y_t'\Omega_t^{-1}BD_t\frac{\delta B}{\delta B_{ij}}'\Omega_t^{-1}\right) = \tilde{y}_t\tilde{y}_t'\frac{\delta\left(\Omega_t^{-1}BD_t\frac{\delta B}{\delta B_{ij}}'\right)}{\delta B_{mn}}\Omega_t^{-1} + \tilde{y}_t\tilde{y}_t'\Omega_t^{-1}BD_t\frac{\delta B}{\delta B_{ij}}'\frac{\delta\Omega_t^{-1}}{\delta B_{mn}}
$$

So, starting with the first term:

$$
\begin{aligned}
H_{ij,mn}^{2,t} &= y_t y_t'\frac{\delta\left(\Omega_t^{-1}BD_t\frac{\delta B}{\delta B_{ij}}'\right)}{\delta B_{mn}}\Omega_t^{-1} \\
&= y_t y_t' H_{ij,mn}^{1,t}\Omega_t^{-1}
\end{aligned}
$$

And finally:

$$
\begin{aligned}
H_{ij,mn}^{3,t} &= y_t y_t'\Omega_t^{-1}BD_t\frac{\delta B}{\delta B_{ij}}'\frac{\delta\Omega_t^{-1}}{\delta B_{mn}} \\
&= -y_t y_t'\Omega_t^{-1}BD_t\frac{\delta B}{\delta B_{ij}}'\Omega_t^{-1}\left(\frac{\delta B}{\delta B_{mn}}D_tB' + BD_t\frac{\delta B}{\delta B_{mn}}'\right)\Omega_t^{-1}
\end{aligned}
$$

The full hessian matrix can be described using:

$$
H_{ij,mn} = \sum_{t=1}^{n} -H_{ij,mn}^{1,t} + H_{ij,mn}^{2,t} + H_{ij,mn}^{3,t} \tag{A.1}
$$

B

RESULTS AMSV MAKING USE OF THE EKF

Parameter	True Value	Posterior mean	Posterior s.d.
μ_i	-9	-9.00	0.18
μ_f	-9	-8.98	0.20
ϕ_i	0.98	0.966	0.035
ϕ_f	0.98	0.972	0.010
σ_i	0.14	0.160	0.023
σ_f	0.14	0.093	0.009
θ_f	0.2	0.200	0.03

Table B.1: The parameters of the logarithmic volatility. The values used for the generation of the data are shown under "True Value". The values under "Posterior mean" and "Posterior s.d." are respectively the average and the standard deviation of posteriors of 40 simulations using 5.000 samples (after 1.000 discarded burn-in samples), generated by the AMSV model.

The main difference between Tables B.1 and 8 are the results of γ_f and σ_f. Although the value for γ_f is estimated better using the EKF, the value for σ_f is heavily underestimated.

Exporting this version of the estimation scheme to Section 7.3, it was found that the method using the Bayes factor was not sufficient to determine whether AMSV or MSV was used. The main reason being $p(y|\mathscr{M}_{\text{AMSV}}, \Psi^*)$ is estimated significantly lower when σ_f is underestimated. Otherwise said, the loss in likelihood because of the underestimation of σ_f canceled out the gain in likelihood by the extra parameter γ_f. Because of this, no conclusion could be made in Section 8. Therefore the conditioning on $\bar{e}_t^f$ is preferred over the conditioning on $\bar{f}_t$.

C

RESULTS FACTOR LOADINGS AND LOGARITHMIC VOLATILITY MSVJ

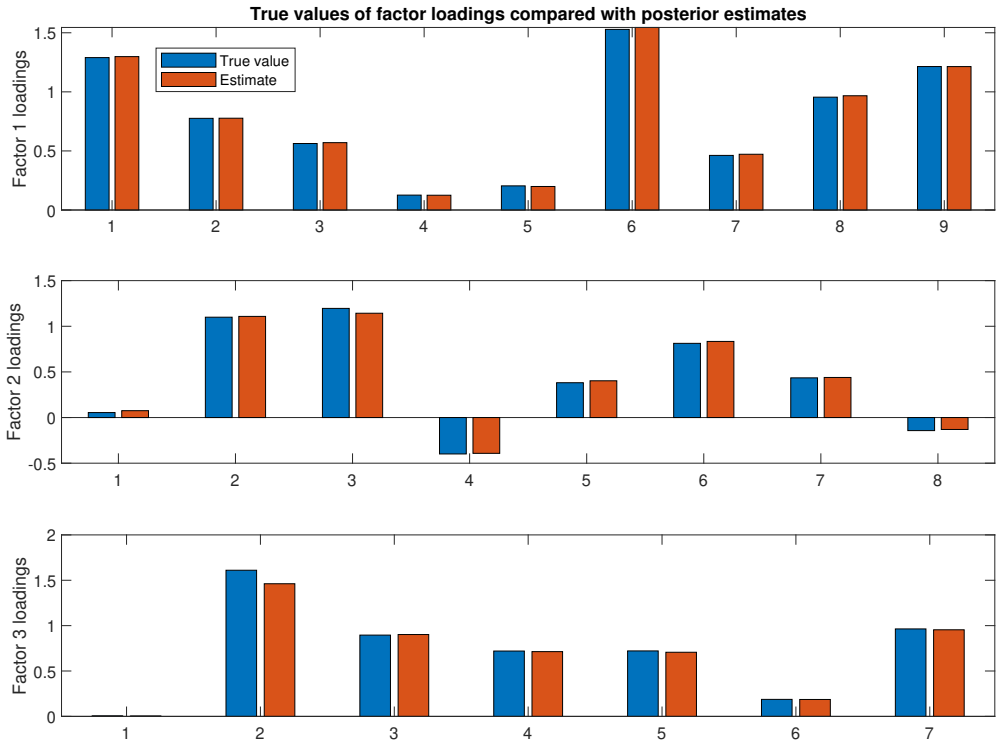

Figure C.1: The parameters of the factor loading matrix. The values used for the generation of the data are shown under "True Value". The values under "Estimate" are the average of posteriors of 40 simulations using 5.000 samples (after 1.000 discarded burn-in samples). The data is generated by the MSVJ model.

Parameter	True Value	Posterior mean	Posterior s.d.
μ_i	-9	-8.93	0.17
μ_f	-9	-9.06	0.16
ϕ_i	0.98	0.947	0.062
ϕ_f	0.98	0.968	0.021
σ_i	0.14	0.170	0.027
σ_f	0.14	0.158	0.022

Table C.1: The parameters of the logarithmic volatility. The values used for the generation of the data are shown under "True Value". The values under "Posterior mean" and "Posterior s.d." are respectively the average and the standard deviation of posteriors of 40 simulations using 5.000 samples (after 1.000 discarded burn-in samples), generated by the MSVJ model.

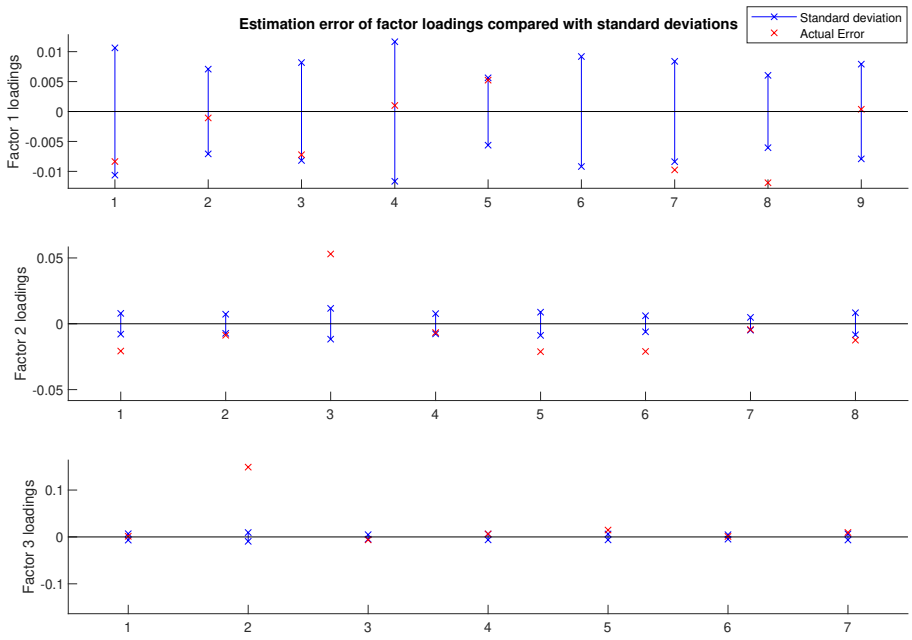

Figure C.2: The standard deviations and the actual errors of the values of the factor loading matrix. The standard deviations and the errors are the result of 40 simulations using 5.000 samples (after 1.000 discarded burn-in samples), generated by the MSVJ model.

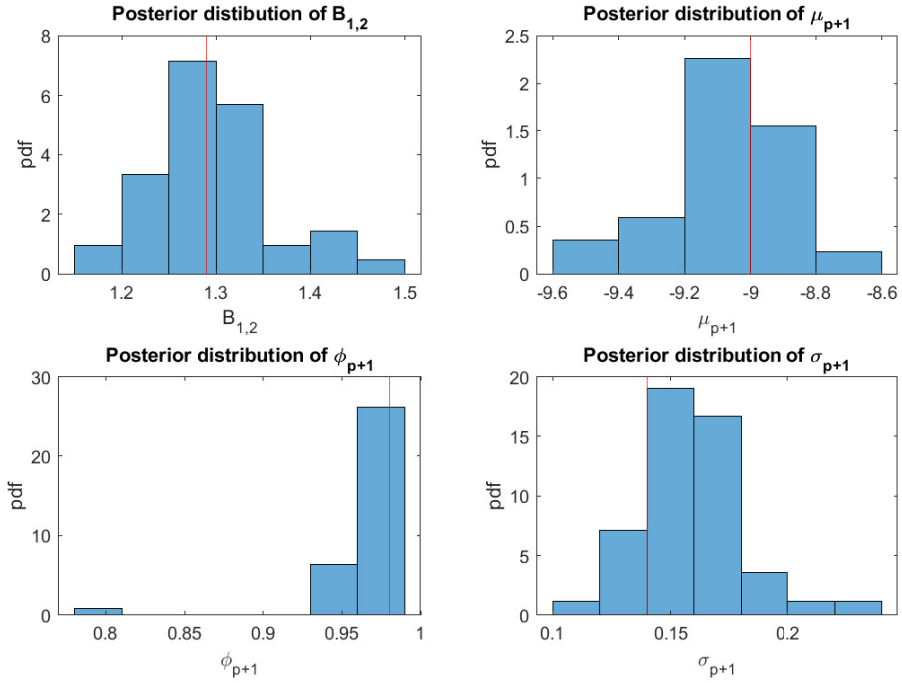

Figure C.3: Histogram of the posterior values for one of the factor loadings and the parameters for the first factor. For this histogram, 40 data sets using 5.000 samples are used. The data sets are generated by the AMSV model.

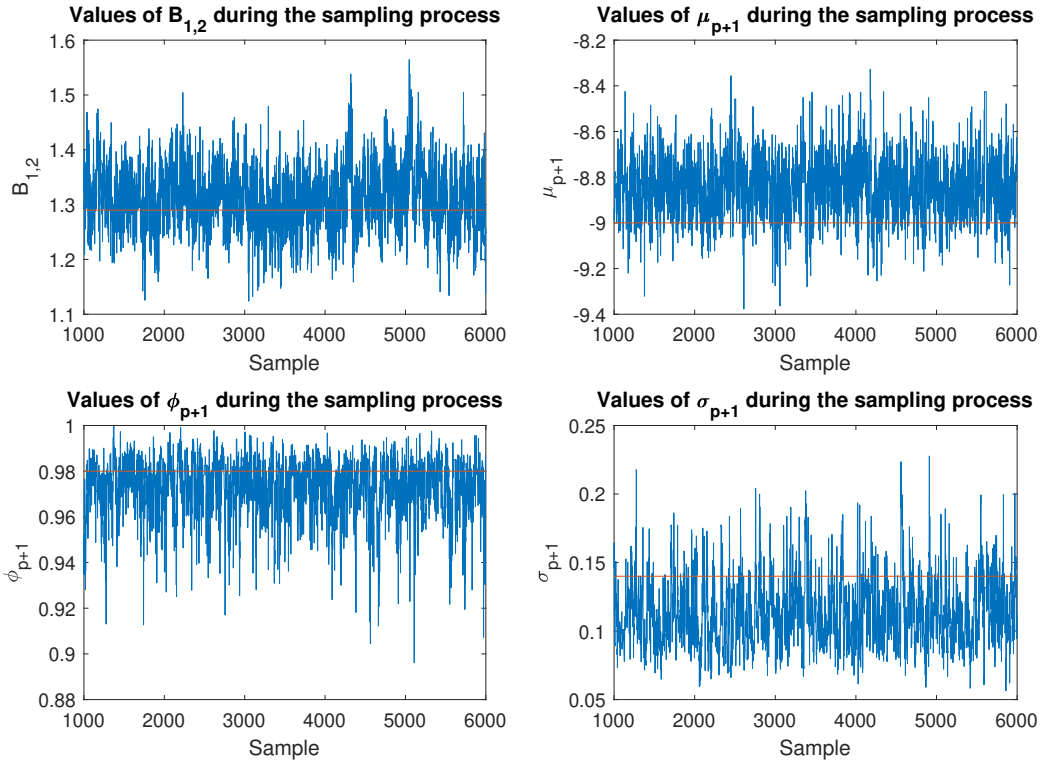

Figure C.4: Path of the sampling of the parameters in Figure C.3 of an arbitrary run using the MSVJ model.

D

Code

The code, written in Matlab R2019b is available upon request.