Master Thesis: Econometrics & Management Science [1]

Specialisation: Quantitative Finance

Erasmus University Rotterdam

Erasmus School of Economics

Univariate Stochastic Volatilty Models with a General and Non-Causal Leverage Structure: A Comparison Study

Author:

Preidun Andkhuiy (433685)

(Co-reader) Supervisor:

Prof. dr. R.J. Lange, &

Prof. dr. D.J.C. van Dijk

August 6, 2021

Abstract

This paper proposes a novel stochastic volatility model (SV) with non-causal leverage structure. This correlation structure takes into account the influence of the future, present, and past volatility shock on the current return shock. We compare it's forecasting accuracy and ability to capture the leverage effect against the recently published SV model of Catania (2020). We do this by making use of three difficult, yet elegant filtering methods in conjunction with maximum likelihood: the particle filter, the Kalman filter and the Bellman filter of Lange (2020). We perform a simulation study to confirm that the estimators are consistent. An empirical study on the S&P 500 and AEX index returns reveals that the proposed SV model significantly outperforms Catania's SV model including a benchmark GARCH(1,1) model. Inclusion of a mean term to both SV models leads to both a better fit of the data and improvement in out-of-sample volatility predictions. Finally, the mean-corrected non-causal SV model is better able to capture the leverage effect.

[1] The content of this thesis is the sole responsibility of the author and does not reflect the view of the supervisor, second assessor, Erasmus School of Economics or Erasmus University.

Contents

1 Introduction

In finance and financial econometrics, the phenomenon "leverage effect" is well known. It is defined as the general tendency of asset returns to be negatively correlated to their corresponding changes in volatility, (e.g. Black 1976 and Christie 1982). According to Ait-Sahalia et al. (2013), the economic intuition behind this phenomenon is when asset prices (returns) decline, companies become more leveraged as their debt to equity ratio increases. It therefore makes economically sense that their stock become more risky, i.e. more volatile. Furthermore, the term "leverage propagation" is introduced in Catania (2020), which is defined as the propagation of the leverage effect over time. As illustrated in Catania (2020), it is of vital importance to correctly model the leverage effect and propagation effect when it comes to financial applications such as volatility predictions and henceforth in other areas such as option pricing.

In order to capture the leverage effect and the leverage propagation with time-series econometric models, we focus our attention on a stream of literature about parameter-driven models. Such models typically contain parameters/latent variables that vary over time as dynamic processes with idiosyncratic innovations, see also Koopman et al. (2016). More specifically, we consider a class of (extended) stochastic volatility models initially proposed by Taylor (1986). Catania (2020) considers stochastic volatility models with leverage (SVL) and debates whether the leverage effect is contemporaneous (simultaneously) or intertemporal (delayed). If the leverage is simultaneous, we obtain the SVL model of Jacquier et al. (2004), whereas if the effect is delayed we have the SVL model of Harvey and Shephard (1996). Besides the timing of the leverage effect, causality is also addressed as correlation does not imply causality (Ait-Sahalia et al. 2013). The main contribution of Catania (2020) is the extension of the basic SV model with a flexible correlation structure which encompasses both SVL models, ensuring causality and allows for leverage effects longer than one period (leverage propagation).

In this master thesis we conduct similar research as done in Catania (2020). However, based on the suggestions of Catania (2020) (for further research), we deviate from his paper and contribute to the literature by proposing a novel stochastic volatilty model with a complicated correlation structure. In this so-called non-causal correlation structure, we additionally take into account negative lags of the volatility shocks such that the influence of future volatility on the current return shocks is included. Our main aim is to feasibly estimate the proposed univariate SV model including a leading volatility lag and

compare its forecasting ability to that of Catania's model. This leads to our central research questions:

1. *"How can we feasibly estimate the proposed univariate stochastic volatility models with General/Non-causal leverage specification?"*

2. *"Can the volatility predictions be improved compared to the univariate stochastic volatility model of Catania (2020)? "*

In order to resolve the first research question, we follow the same methodology of Catania (2020) which are simulation maximum likelihood (SML) and quasi-maximum likelihood (QML). In addition, we make use of the Bellman filter proposed by Lange (2020) in conjunction with maximum likelihood. This method is computationally efficient and employs a filter-implied log-likelihood decomposition which requires only outputs of the Bellman filter and no simulations. In this paper, these three methods are used for estimating the univariate SV models with leverage.

Regarding the second question, we conduct a comparative out-of-sample forecasting study on the S&P 500 and AEX index return series. The mean squared forecast error (MSFE) and the quasi-likelihood (QLIKE) measure are used for that purpose. Doing so allows us to check whether the accuracy of the volatility predictions of the model of Catania (2020) can be significantly improved. Both the 5-min. intra-day realized variances and noisy squared returns are used for out-of-sample evaluations. As benchmark we have the standard GARCH(1,1) model but we also consider the asymmetric GARCH(1,1).

The empirical results show that the novel SV model with non-causal leverage is significantly able to outperform the SV model postulated by Catania (2020) when forecasting the one-day ahead volatilities. This also includes the outperformance of the (asymmetric) GARCH(1,1). However this claim only holds when we use the 5-min. realized variances and not the noisy squared returns. Besides that, the SV model with non-causal leverage seems to be able to capture the leverage effect properly when corrected for the mean of the return series. For Catania's model the inclusion of a mean term does not necessarily improve the ability to account for the leverage.

The remaining of this thesis is organized as follows. Section 2 provides a literature review on recent research about our topic. Sections 3 introduces the theoretic state space framework and Section 4 considers the univariate SV models with detailed explanation of the corresponding estimation procedures. Section 5 presents a simulation study where we analyze finite sample properties of our estimators. In this study we validate the estimator's consistency for both univariate SV models with leverage. The empirical application is reported in Section 6 and Section 7 concludes.

2 Literature

In order to better understand the motivation for the approaches that we apply, we discuss some of the (recent) developments that have been made in the literature of non-linear, non-Gaussian state space methods. The methods that we apply mostly stem from cited papers in Catania (2020). We discuss the CSIR method of Malik and Pitt (2011), the quasi maximum likelihood method of Harvey and Shephard (1996), and the Bellman filter of Lange (2020). As mentioned in Catania (2020), Broto and Ruiz (2004) conclude that likelihood based estimation procedures are the most efficient for estimating (extensions of) Stochastic volatility models with leverage. Therefore, we apply such methods to univariate SVL models as well.

For simulated maximum likelihood, the most relevant paper is Malik and Pitt (2011) where a specific particle filtering method is proposed. Particle filters, initially proposed by Gordon et al. (1993), are sequential Monte Carlo algorithms that address the filtering problem in nonlinear and non-Gaussian state space model. More specifically, particle filters simply aim to approximate a continuous high-dimensional density (the joint likelihood) by using a sample of weighted draws. These weighted draw are resampled from the population of original draws from a candidate density. Gordon et al. (1993) rename the particle filter as the SIR method, which stands for sampling importance resampling. Here the candidate density is termed the importance sampler and weighted or re-sampled draws from this importance sampler are called particles. The main practical problem with applying standard particle filters lies in feasibly estimating the hyperparameter. There is a discontinuity in the log-likelihood which makes numerical optimization difficult. Malik and Pitt (2011) resolve this issue by using continuous approximations of the log-likelihood in the resampling step of algorithm SIR. Hence, the name: continuous SIR method (CSIR). The CSIR filter makes parameter estimation more feasible.

Quasi-maximum likelihood is computationally more efficient approach as compared to simulation based likelihood estimation since it uses the Kalman filter. That is why this method is commonly used in the literature of SV models with leverage and it originates in the paper of Harvey and Shephard (1996). Two key step for feasible estimation of the SV model with leverage is the logarithmic transformation of the nonlinear observation equation and doing inference conditional on the signs of the observations. These two steps render a linear state space model with conditionally uncorrelated disturbance terms which can be easily estimated with a Kalman filter algorithm. Catania (2020) generalizes these two steps to according to his proposed SV model with general correlation.

Alternatively, the recent paper of Lange (2020) proposes a promising filtering algorithm called the Bellman filter, which generalizes the kalman filter (1960). It also includes the iterated extended kalman filter (Anderson and Moore, 2012) designed for non-linear state space models. Furthermore, this method differs from SML and QML in the sense that it concentrates on estimating the posterior mode (of the complete data likelihood) instead of approximating a complex high-dimensional integral. Therefore, this method requires no simulation, is computationally efficient and offers scalability to higher dimensions for the latent state variable in state space models. Lastly, the Bellman filter allows for feasible estimation of the hyperparameters of the joint-likehood function since the log-likelihood can be computed using only the outputs of the Bellman filter.

Despite the fact we do not estimate multivariate models, it might be interesting to discuss widely used methods for estimating multivariate SV models based on importance sampling. For example, Danielsson (1998) illustrate the application of the Accelerated Gaussian Importance Sampling (AGIS) method of Danielsson and Richard (1993) for estimating a multivariate stochastic volatility model with leverage. Asai and McAleer (2007) apply the approach of Durbin and Koopman (1997) to assymetric multivariate SV models where they decompose the high dimensional likelihood function into a Gaussian component and a remainder which are evaluated with Kalman filtering methods and importance sampling, respectively. Despite these succesful application in high-dimensional SV models, these importance sampling methods still suffer from the curse of dimensionality. This phenomenon refers to the fact that the simulation variance grows exponentially as the integration dimension increases. The recent paper of Scharth and Kohn (2013) resolves this problem by introducing the particle-EIS method. It is a combination of the auxiliary particle filtering method of Pitt and Sheppard (1999) and the efficient importance sampling of Richard and Zhang (2007). Their simulation study show that their P-EIS approach strongly outperforms the individual methods when evaluating the likelihoods from univariate and bivariate SV models. To the best of my knowledge, the multivariate version of the stochastic volatility specification of Catania (2020) has never been explored before.

3 State Space Model Framework

In this section, we provide the necessary notation and foundation for the methods that we apply in this thesis. Stochastic state space models are able to describe the dynamics of observable variables by letting them depend on hidden states that follow a stochastic process. Let $\boldsymbol{y}_t = (y_{1t}, \ldots, y_{lt})'$ denote a $l \times 1$ vector of observable time-series and $\boldsymbol{\alpha}_t = (\alpha_{t1}, \ldots, \alpha_{mt})'$ the $m \times 1$ vector of hidden state variables. The general discrete-time stochastic state space model for $t = 1, \ldots, T$ is given by

$$\boldsymbol{y}_t = f(\boldsymbol{\alpha}_t, \boldsymbol{\varepsilon}_t), \tag{1}$$

$$\boldsymbol{\alpha}_t = g(\boldsymbol{\alpha}_{t-1}, \boldsymbol{\eta}_t). \tag{2}$$

Here, the observation equation (1) reflects the contemporaneous relationship of the observable variables $\boldsymbol{y}_t$ with the states $\boldsymbol{\alpha}_t$ and measurement noise vector $\boldsymbol{\varepsilon}_t$, via a function $f(\cdot)$. The state equation (2) describes the dynamics of the hidden states $\boldsymbol{\alpha}_t$ as a function $g(\cdot)$ of the previous states $\boldsymbol{\alpha}_{t-1}$ and process noise vector $\boldsymbol{\eta}_t$.

In the general state space framework it is common to have a set of assumptions. Firstly, we assume that the disturbance vectors $\boldsymbol{\varepsilon}_t$, $\boldsymbol{\eta}_t$ are i.i.d. and are independent of each other. The joint densities of the individual noise vectors and the initial state need to be specified and are formulated as

$$\boldsymbol{\varepsilon}_t \overset{iid}{\sim} p(\boldsymbol{\varepsilon}_t), \;\; \boldsymbol{\eta}_t \overset{iid}{\sim} p(\boldsymbol{\eta}_t), \;\; \boldsymbol{\alpha}_0 \sim p(\boldsymbol{\alpha}_0). \tag{3}$$

This specification implies the following generic notation for the nonlinear, non-Gaussian state space model:

$$\boldsymbol{y}_t \sim p_(\boldsymbol{y}_t \,|\boldsymbol{\alpha}_t), \;\; \boldsymbol{\alpha}_t \sim p(\boldsymbol{\alpha}_t \,|\boldsymbol{\alpha}_{t-1}), \;\; \boldsymbol{\alpha}_0 \sim p(\boldsymbol{\alpha}_0), \tag{4}$$

where the observation density can differ from the state transition density. Lastly, the generic joint densities $p(\cdot)$ are assumed to contain some additional fixed parameters $\boldsymbol{\theta}$ which need to be estimated.

When we allow for a linear Gaussian state equation we obtain the framework mentioned in Lange (2020):

$$\boldsymbol{y}_t \sim p_(\boldsymbol{y}_t \,|\boldsymbol{\alpha}_t), \quad \boldsymbol{\alpha}_t = \boldsymbol{c} + \boldsymbol{T}\boldsymbol{\alpha}_{t-1} + \boldsymbol{\eta}_t, \quad \boldsymbol{\eta}_t \overset{iid}{\sim} \mathcal{N}_m(\boldsymbol{0}, \boldsymbol{Q}), \quad \boldsymbol{\alpha}_0 \sim p(\boldsymbol{\alpha}_0). \tag{5}$$

In this framework, the observation equation can be nonlinear and allows for non-Gaussian error terms. The classic linear state space model can easily be recovered from by letting $f(\cdot)$ be linear in $\boldsymbol{\alpha}_t$ and that observation noise vector $\boldsymbol{\varepsilon}_t$ is Gaussian. The fixed parameters

$\boldsymbol{\theta}$ corresponding to the linear state-space model can then be estimated with likelihood based estimation via the Kalman filtering recursions (1960), see e.g. Koopman and Durbin (2012).

4 Univariate Models

Two univariate extensions of the stochastic volatility models will be considered in the upcoming subsections. The first one is the stochastic volatility with general leverage specification proposed by Catania (2020). The second is a different stochastic volatility model and contains a novel complicated non-causal leverage structure. Afterwards, three estimation procedures will be derived for each univariate model. We follow the notation of Catania (2020) throughout the rest of the paper.

4.1 Stochastic Volatilty with General Leverage

Catania (2020) assumes that the financial log-returns y_t at time t are generated according to the following data generating process with mean term $\mu = 0$:

$$\begin{aligned} y_t &= \mu + \beta \exp\left\{\frac{h_t}{2}\right\} \varepsilon_t, \quad \varepsilon_t \overset{iid}{\sim} \mathcal{N}(0,1), \\ h_t &= \phi h_{t-1} + \kappa\, \eta_t, \\ \eta_t &= \sum_{j=0}^{m} \rho_j \varepsilon_{t-j} + \sigma_b\, b_t, \quad b_t \overset{iid}{\sim} \mathcal{N}(0,1). \end{aligned} \tag{6}$$

where $\beta > 0$, $\kappa > 0$, and $|\phi| < 1$ are the level, variance and persistence parameters of the log-volatility process ($h_t = \log \sigma_t^2$), respectively. Furthermore, the return shocks ε_t are assumed to be independent of the the standard normal noise term b_t. For identification of the model, it is required that $\sigma_b = \sqrt{1 - \sum_{j=0}^{m} \rho_j^2} > 0$, where $\rho_j \in (-1, 1)$ denotes the unconditional correlations between the return shock and the volatility shocks for $j = 0, \ldots, m$. This model encompasses a few models. When $m = 0$ and $\rho_0 = 0$, we obtain the basic stochastic volatility model of Taylor (1986). When $m = 1$ and $\rho_0 = 0$, we see it's equivalent to the intertemporal stochastic volatility model of Harvey and shephard (1996). Lastly, when $m = 0$ but $\rho_0 \neq 0$, we retrieve the contemporaneous model of Jacquier et al. (2004). Catania (2020) explains more details about the statistical properties of his model, but this is left as a reference to the reader.

As we can see from the general leverage specification (6), the volatility shocks depend on multiple lags of the return shocks, which makes it an ARMA(1,m) kind of specification. This correlation structure of the noise terms enables us to explicitly compute the (implied) leverage effect and leverage propagation. However, we theoretically show this for $\rho_0 = 0$ only, since relaxing this restriction leads to theoretically inconvenient expressions for the leverage effect, Yu (2005). The definitions of Catania (2020) are given below:

Definition 1 (Leverage effect) If we denote y_t as the financial returns following a stochastic process, then the leverage effect exists if the following inequality holds:

$$\text{var}[y_{t+1} \,|\, y_t < 0] > \text{var}[y_{t+1} \,|\, y_t \geq 0].$$

The leverage propagation refers to the leverage effect over longer horizons and can be determined with $\text{var}(y_t \,|\, y_{t-k} < 0)$ for $k > 0$. The analytical expression for the propagation, derived by Catania (2020), is given by:

$$\text{var}(y_t \,|\, y_{t-k} < 0) = \text{var}(\sigma_t^2 \,|\, y_{t-k} < 0) \overset{\rho_0 = 0}{=} \begin{cases} 2\mu_\sigma \Phi(-\kappa \sum_{j=0}^{k} \phi^{k-j} \rho_j), & \text{if } 0 < k \leq m, \\ 2\mu_\sigma \Phi(-\kappa \phi^k \sum_{j=0}^{m} \phi^{m-j} \rho_j), & \text{if } k > m, \end{cases}$$

where $\mu_\sigma = E(\sigma_t^2) = \beta^2 \exp\{\sigma_h^2/2\}$, and $\Phi(\cdot)$ is the Gaussian cumulative distribution function. The unconditional variance σ_h^2 is given in section 2.1 of Catania (2020).

In order to apply state space methods, Catania (2020) rewrites model (6) as a nonlinear, Gaussian state space model with uncorrelated noise terms. The resulting SV model with leverage that we employ is given as:

$$\begin{aligned} y_t - \mu \;=\; & \beta \exp\left\{\frac{h_t}{2}\right\} \Bigg[\sqrt{1 - \frac{\rho_0^2}{1 - \sum_{j=1}^{m} \rho_j^2}}\, u_t + \frac{\rho_0}{\kappa \left(1 - \sum_{j=1}^{m} \rho_j^2\right)} \\ & \times \left[h_t - \phi h_{t-1} - \frac{\kappa}{\beta} \sum_{j=1}^{m} \rho_j \left(y_{t-j} - \mu\right) \exp\left\{\frac{h_{t-j}}{2}\right\} \right] \Bigg], \end{aligned}$$

$$h_t \;=\; \phi h_{t-1} - \frac{\kappa}{\beta} \sum_{j=1}^{m} \rho_j \left(y_{t-j} - \mu\right) \exp\left\{\frac{h_{t-j}}{2}\right\} + \kappa \sqrt{1 - \sum_{j=1}^{m} \rho_j^2}\, \varpi_t, \tag{7}$$

where u_t and ϖ_t are uncorrelated noise terms who follow a standard normal distribution. If we condition on previous states $\boldsymbol{\alpha}_t = (h_t, \ldots, h_{t-m+1})'$ and observations $\boldsymbol{y}_t = (y_t, \ldots, y_{t-m+1})'$, we get that $(y_t - \mu) \,|\, \boldsymbol{\alpha}_t, \boldsymbol{y}_{t-1} \sim \mathcal{N}(\mu_{y,t},\, \sigma_{y,t}^2)$ and $h_t \,|\boldsymbol{\alpha}_{t-1}, \boldsymbol{y}_{t-1} \sim \mathcal{N}(\mu_{h,t}\, \sigma_{h,t}^2)$. The formulas for the conditional mean and variances are shown below:

$$
\begin{aligned}
\mu_{y,t} &= \beta \exp\left\{\frac{h_t}{2}\right\} \frac{\rho_0}{\kappa\left(1-\sum_{j=1}^{m}\rho_j^2\right)} \\
&\qquad \times \left(h_t - \phi h_{t-1} - \frac{\kappa}{\beta}\sum_{j=1}^{m}\rho_j\left(y_{t-j}-\mu\right)\exp\left\{\frac{h_{t-j}}{2}\right\} \right) \\
\sigma_{y,t}^2 &= \beta^2 \exp\left\{h_t\right\}\left(1-\frac{\rho_0^2}{1-\sum_{j=1}^{m}\rho_j^2}\right) \\
\mu_{h,t} &= \phi h_{t-1} - \frac{\kappa}{\beta}\sum_{j=1}^{m}\rho_j\left(y_{t-j}-\mu\right)\exp\left\{\frac{h_{t-j}}{2}\right\} \\
\sigma_{h,t}^2 &= \kappa^2\left(1-\sum_{j=1}^{m}\rho_j^2\right). \qquad (8)
\end{aligned}
$$

4.2 New Stochastic Volatility Model

We now present a new stochastic volatility model with a complicated non-causal structure in the observation equation. Here, we assume that the log-financial returns y_t are generated from the following model:

$$
\begin{aligned}
y_t &= \mu + \exp\left\{\frac{h_t}{2}\right\}\tilde{\varepsilon}_t, \quad \tilde{\varepsilon}_t \sim \mathcal{N}(0,1), \\
h_t &= c + \varphi h_{t-1} + \sigma_\eta\,\eta_t, \quad \eta_t \overset{iid}{\sim} \mathcal{N}(0,1), \qquad (9)
\end{aligned}
$$

where c, $|\phi|<1$ and σ_η^2 are the unknown level, persistence and noise variance parameters of the model. Furthermore, the observation equation contains return shocks $\tilde{\varepsilon}_t$ that are not i.i.d. and unknown mean parameter μ. We deviate from the paper of Catania (2020) by taking the complicated correlation structure:

$$
\tilde{\varepsilon}_t = \left[\sum_{i=-1}^{+1}\rho_i\,\eta_{t+i}\right] + \sigma_\varepsilon\,\varepsilon_t, \quad \varepsilon_t \overset{iid}{\sim} \mathcal{N}(0,1). \qquad (10)
$$

The novelty of the SV model lies in above non-casaul equation where the current return shock $\tilde{\varepsilon}_t$ depends on the volatility shocks of yesterday (η_{t-1}), today (η_t) and tomorrow (η_{t+1}). These volatility shocks influence the current return shock via ρ_{-1}, ρ_0 and ρ_1, respectively. Also, we can easily verify that $\tilde{\varepsilon}_t \sim \mathcal{N}(0,1)$ remains true. Similarly to the model of Catania (2020), we require that $\sigma_\varepsilon = \sqrt{1-\sum_{i=-1}^{+1}\rho_i^2}$ for identification.

If $Corr(\tilde{\varepsilon}_t, \eta_{t-i}) = \rho_i = 0$ for all i, we obtain the basic SV model without leverage since $\tilde{\varepsilon}_t = \varepsilon_t$ in that case. The standard SVL model of harvey and Shephard (1996) is recovered when only $\rho_{-1} \neq 0$.

As before, we can rewrite the model given in equations (9) as a Gaussian state space model with a linear state equation:

$$y_t - \mu = f(\boldsymbol{\alpha}_t, \boldsymbol{\varepsilon}_t), \quad \boldsymbol{\alpha}_t = \boldsymbol{c} + \boldsymbol{T}\boldsymbol{\alpha}_{t-1} + \boldsymbol{\eta}_t, \quad \boldsymbol{\eta}_t \overset{iid}{\sim} \mathcal{N}_4(\boldsymbol{0}, \boldsymbol{Q}), \tag{11}$$

where

$$f(\boldsymbol{\alpha}_t, \boldsymbol{\varepsilon}_t) = \exp\left\{\frac{h_t}{2}\right\}\left[\left(\sum_{i=-1}^{+1} \rho_i \, \eta_{t+i}\right) + \sigma_\varepsilon \varepsilon_t\right],$$

$$\boldsymbol{\alpha}_t = \begin{pmatrix} h_t \\ \eta_{t+1} \\ \eta_t \\ \eta_{t-1} \end{pmatrix}, \; \boldsymbol{c} = \begin{pmatrix} c \\ 0 \\ 0 \\ 0 \end{pmatrix}, \; \boldsymbol{T} = \begin{pmatrix} \varphi & \sigma_\eta & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix}, \; \boldsymbol{\eta}_t = \begin{pmatrix} 0 \\ \eta_{t+1} \\ 0 \\ 0 \end{pmatrix}.$$

The (singular) symmetric and positive definite covariance matrix of the process noise $\boldsymbol{\eta}_t$ is given by

$$\boldsymbol{Q} \;=\; \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}. \tag{12}$$

For practical purposes, we can instead use pseudo-inverses of the covariance matrix $\boldsymbol{Q}$. Conditional on the current states (and observations), the observations are distributed as $(y_t - \mu) \,|\, \boldsymbol{\alpha}_t, \, \boldsymbol{y}_{1:t-1} \sim N(\mu_{y,t}, \, \sigma^2_{y,t})$, where $\boldsymbol{y}_{1:t-1}$ denotes the filtration set of observations $\{y_1, \, \ldots \, , y_{t-1}\}$. The conditional mean and variance of y_t are

$$\mu_{y,t} \;=\; \exp\left\{\frac{h_t}{2}\right\}\left(\sum_{i=-1}^{+1} \rho_i \, \eta_{t+i}\right),$$

$$\sigma^2_{y,t} \;=\; \exp\{h_t\}\left(1 - \sum_{i=-1}^{+1} \rho_i^2\right).$$

4.3 SML Estimation

Application of simulated maximum likelihood on the stochastic volatility model with general leverage seems remarkably difficult. Therefore, we follow the methodology of Catania (2020) by using the CSIR algorithm of Malik and Pitt (2011) to evaluate the joint (log-) likelihood function which is not available in closed form. We estimate the collection of unknown parameters in $\boldsymbol{\theta}$ by maximizing the numerically evaluated log-likelihood. For the general SV model of Catania (2020), we have $\boldsymbol{\theta} = (\beta, \phi, \kappa, \mu, \boldsymbol{\rho})$ with $\boldsymbol{\rho} = (\rho_1, \ldots, \rho_m)$.

Before we discuss and provide the CSIR algorithm of Malik and Pitt (2011), we provide some preliminaries on particle filtering. The core problem lies in the fact that the marginal likelihood $p(y_t \,|\, \mathcal{F}_{t-1})$ is not available in closed form since the states $\boldsymbol{\alpha}_t$ enter the observation density in a non-linear way. We can write the marginal likelihood as:

$$p(y_t \,|\, \boldsymbol{\theta};\, \mathcal{F}_{t-1}) = \int p(y_t \,|\boldsymbol{\alpha}_t; \boldsymbol{\theta})\, p(\boldsymbol{\alpha}_t \,|\, \boldsymbol{\theta};\, \mathcal{F}_{t-1}) d\boldsymbol{\alpha}_t. \tag{13}$$

The above integral can be approximated with the Monte Carlo mean:

$$\hat{p}(y_t \,|\, \boldsymbol{\theta};\, \mathcal{F}_{t-1}) = 1/N \sum_{k=1}^{N} p(y_t \,|\tilde{\boldsymbol{\alpha}}_t^k; \boldsymbol{\theta}), \tag{14}$$

where $\tilde{\boldsymbol{\alpha}}_t^1, \ldots, \tilde{\boldsymbol{\alpha}}_t^N$ are simulation draws from the the predictive density $p(\boldsymbol{\alpha}_t \,|\, \boldsymbol{\theta};\, \mathcal{F}_{t-1})$.

The particle filtering method: SIR proposed by Gordon et al. (1993) is given in algorithm 1 on the next page. The algorithm recursively solves the issue of obtaining simulation draws, $\boldsymbol{\alpha}_t^k$, for each time t by making use of the Bayesian updating rule:

$$p(\boldsymbol{\alpha}_t \,|\, \boldsymbol{\theta};\, \mathcal{F}_t) \propto p(y_t \,|\boldsymbol{\alpha}_t; \boldsymbol{\theta}) \int p(\boldsymbol{\alpha}_t \,|\, \boldsymbol{\alpha}_{t-1}; \boldsymbol{\theta})\, p(\boldsymbol{\alpha}_{t-1} \,|\, \boldsymbol{\theta};\, \mathcal{F}_{t-1})\, d\boldsymbol{\alpha}_{t-1}. \tag{15}$$

Essentially, the intuition of this recursive algorithm is as follows. Suppose, we have a set of "particles", $(\boldsymbol{\alpha}_{t-1}^1, \ldots, \boldsymbol{\alpha}_{t-1}^N)$ drawn from the filtered state density $p(\boldsymbol{\alpha}_{t-1} \,|\, \boldsymbol{\theta};\, \mathcal{F}_{t-1})$, with discrete probability weights $\pi_{t-1}^1, \ldots, \pi_{t-1}^N$. We propagate the old set of particles to a new one $(\tilde{\boldsymbol{\alpha}}_t^1, \ldots, \tilde{\boldsymbol{\alpha}}_t^N)$ by sampling from the state transition density $p(\boldsymbol{\alpha}_t \,|\, \boldsymbol{\alpha}_{t-1}^k; \boldsymbol{\theta})$. In order to filter the propagated particles, we resample from them according to a multinomial sampling scheme with associating normalized probabilities $\pi_t^1, \ldots, \pi_t^N$. We then obtain again the set of filtered particles $(\boldsymbol{\alpha}_t^1, \ldots, \boldsymbol{\alpha}_t^N)$ but now at time t. The SIR method assumes that we can compute the obervation density $p(y_t \,|\boldsymbol{\alpha}_t; \boldsymbol{\theta})$ and sample from the transition state density $p(\boldsymbol{\alpha}_t \,|\, \boldsymbol{\alpha}_{t-1}; \boldsymbol{\theta})$ for $t = 0, \ldots, T$.

Algorithm 1: SIR

1. *For* $t = 0$*, we sample a set of particles* $\{\boldsymbol{\alpha}_0^k\}_{k=1}^N$ *from the stationary distribution* $p(\boldsymbol{\alpha}_0)$.

   For $t = 1, \ldots, T$*, do:*

2. *For* $k = 1, \ldots, N$*: sample* $\tilde{\boldsymbol{\alpha}}_t^k \sim p(\boldsymbol{\alpha}_t \,|\, \boldsymbol{\alpha}_{t-1}^k; \boldsymbol{\theta})$

3. *For* $k = 1, \ldots, N$*: compute the normalized weights:*

$$\pi_t^k = \frac{\omega_t^k}{\sum_{i=1}^N \omega_t^i}, \quad \textit{where} \quad \omega_t^k = p(y_t \,|\tilde{\boldsymbol{\alpha}}_t^k; \boldsymbol{\theta}).$$

4. *For* $k = 1, \ldots, N$*: sample from the mixture:* $\boldsymbol{\alpha}_t^k \sim \sum_{i=1}^N \pi_t^i \, \delta(\boldsymbol{\alpha}_t - \tilde{\boldsymbol{\alpha}}_t^i)$*, where* $\delta(\cdot)$ *denotes the Diract-delta function centered at* $\mathbf{0}$.

With the output of Algorithm 1, the log-likelihood can be approximated with

$$\hat{\ell}_{N,T}(\boldsymbol{\theta}) = \sum_{t=1}^{T} \log \hat{p}(y_t \,|\, \boldsymbol{\theta};\, \mathcal{F}_{t-1}) = \sum_{t=1}^{T} \log \left(\frac{1}{N} \sum_{k=1}^{N} \omega_t^k \right). \tag{16}$$

However, Malik and Pitt (2011) argue that (step 4) resampling from the mixture empirical distribution function $\hat{F}_N(\boldsymbol{\alpha}_t) = \sum_{i=1}^N \pi_t^i \, \delta(\boldsymbol{\alpha}_t - \tilde{\boldsymbol{\alpha}}_t^i)$ leads to discontinuities in $\hat{\ell}_{N,T}(\boldsymbol{\theta})$. Such discontinuities may be detrimental for estimating the parameter $\boldsymbol{\theta}$ with maximum likelihood. They remedy this by replacing the mixture function of step 4, with a continuous approximate function $\tilde{F}_N(\boldsymbol{\alpha}_t)$, which leads to CSIR algorithm. As in Catania (2020) we apply the resampling step only for the first element of the state vector $\boldsymbol{\alpha}_t = (h_t, \ldots, h_{t-m+1})'$ and leave the remaining m elements unchanged while filtering. For details on practical implementation of continuous resampling, we refer to appendix A.

Let the continuous function be

$$\tilde{F}_N(h_t) = \sum_{k=0}^{N} \lambda_t^k \, G_K \left(\frac{h_t - h_t^{(k)}}{h_t^{(k+1)} - h_t^{(k)}} \right), \tag{17}$$

where we set

$$\lambda_t^k = \begin{cases} \pi_t^1/2 \,, & \text{if } k = 0, \\ \pi_t^{k+1} + \pi_t^k)/2 \,, & \text{if } k = 1, \ldots, N-1, \\ \pi_t^N/2 \,, & \text{if } k = N, \end{cases} \quad G_k(z) = \begin{cases} \mathbb{1}(z > 0) \,, & \text{if } k = 0, \\ F_U(z) \,, & \text{if } k = 1, \ldots, N-1, \\ \mathbb{1}(z > 0) \,, & \text{if } k = N. \end{cases}$$

Here, $\mathbb{1}(z > 0)$ denotes the indicator function taking on 1 if $z > 0$ and $F_U(z)$ is the uniform cumulative function. Furthermore, we need that $h_t^{(0)} = -\infty$ and $h_t^{(N+1)} = \infty$, where $h_t^{(k)}$ denotes the k-th order statistic. As the number of particles N goes to infinity, we know that particles $(\tilde{h}_t)$ drawn from $\tilde{F}_N(h_t)$ will be equivalent to draws from the true predictive density $p(h_t \,|\mathcal{F}_{t-1})$, with probability 1. The output of the CSIR algorithm can then be used to evaluate the log-likelihood function given in equation (26), which now becomes continuous in $\boldsymbol{\theta}$. The SML estimates are then obtained as

$$\hat{\boldsymbol{\theta}}_{\text{SML}} = \arg\max_{\theta} \hat{\ell}_{N,T}(\boldsymbol{\theta}). \tag{18}$$

4.4 QML Estimation

In contrast to the SML method, the quasi-maximum likelihood approach does not require simulation draws to estimate $\boldsymbol{\theta}$ for the SV model of Catania (2020). We cannot apply QML to the non-causal SV model, because taking the logarithm of the squared return shocks $\tilde{\varepsilon}_T^2$ in equation (9) leads to complications. However, this could be a neat extention for future research. Since it is already explicitly discussed in Catania (2020), we will not go into much details when discussing the QML method.

Essentially, the crux of the QML approach is to write model (6) as a state space model with linear equations where the observation and volatility shocks are uncorrelated. If we square and take logarithms, the observation equation becomes

$$y_t^* = \omega + h_t + \xi_t, \;\; h_{t+1} = \phi\, h_t + \kappa\, g_t, \tag{19}$$

where $y_t^* = \log(y_t - \mu)^2$, $\omega = 2\log\beta + \zeta$, $\xi_t = E(\log\varepsilon_t^2) - \zeta$ with $\zeta = E(\log\varepsilon_t^2) \approx -1.270$. Here, the above state space model is linear, but not Gasssian and the return shocks $\xi_t \sim iid(0, \sigma_\xi^2)$ are correlated with the volatility shocks $g_t = \eta_{t+1}$. In order to make QML estimation via the Kalman filter feasible, the strategy of Harvey and Shephard (1996) is adopted to construct a linear state space model with uncorrelated error terms. In this strategy the information about the correlation between the noise terms are recovered by using the auxiliary variables $\boldsymbol{s}_t = (s_t, \ldots, s_{t-m+1})$ where $s_t = \text{sign}(y_t)$. In addition, by introducing the latent i.i.d stochastic variables q_t and z_t, Catania (2020) derives that the log-volatility h_t can be written as

$$h_{t+1} = c_t + \sum_{j=1}^{m} \phi_{j,t}\, h_{t-j+1} + \kappa\upsilon \sum_{j=1}^{m} \rho_j q_{t-j+1} + \kappa\upsilon\, z_t, \tag{20}$$

where

$$c_t = \kappa \left[\mu^* \sum_{j=1}^{m} \rho_j s_{t-j+1} + (\gamma^*/\sigma_\xi^2) \sum_{j=1}^{m} \rho_j s_{t-j+1} \left(y^*_{t-j+1} - \omega \right) \right], \;\; \upsilon^2 = 1 - \left(\mu^{*2} + \frac{\gamma^{*2}}{\sigma_\xi^2} \right) \sum_{j=1}^{m} \rho_j^2,$$

$$\phi_{1,t} = \phi - \frac{\kappa\, \gamma^*\, s_t\, \rho_1}{\sigma_\xi^2}, \qquad \phi_{j,t} = - \frac{\kappa\, \gamma^*\, \rho_j\, s_{t-j+1}}{\sigma_\xi^2}, \qquad \mu^* = E(\,|\varepsilon_t|\,), \qquad \gamma^* = E(\,|\varepsilon_t| \cdot \log \varepsilon_t^2\,).$$

By using the formulas given above, we can write (19) compactly and linearly as

$$y_t^* = \omega + \boldsymbol{Z}\boldsymbol{\alpha}_t + \xi_t, \quad \xi \sim iid\,(0, \sigma_\xi^2), \tag{21}$$

$$\boldsymbol{\alpha}_{t+1} = \boldsymbol{b}_t + \boldsymbol{T}_t \boldsymbol{\alpha}_t + \boldsymbol{H}\, \boldsymbol{w}_t, \quad \boldsymbol{w}_t \sim iid \left(\boldsymbol{0}, \begin{bmatrix} \sigma_z^2 & 0 \\ 0 & \sigma_q^2 \end{bmatrix} \right), \tag{22}$$

where $\boldsymbol{\alpha}_t = (h_t, \ldots, h_{t-m+1}, q_t, \ldots, q_{t-m+1})'$, $\boldsymbol{w}_t = (z_t, q_{t+1})'$ are the new state vector, noise terms and the $1 \times 2m$ vector $\boldsymbol{Z} = (1, 0, \ldots, 0)$, respectively. The variances of the state shocks $\boldsymbol{w}_t$ are

$$\sigma_q^2 = \left(1 - \mu^{*2} - \frac{\gamma^{*2}}{\sigma_\xi^2} \right) \frac{1}{\upsilon^2}, \qquad \sigma_z^2 = 1 - \sigma_q^2 \sum_{j=1}^{m} \rho_j^2.$$

The time-varying fixed parameter matrices are defined as

$$\boldsymbol{b}_t = \begin{bmatrix} c_t \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \; \boldsymbol{T}_t = \begin{bmatrix} \phi_{1,t} & \phi_{2,t} & \ldots & \phi_{m,t} & \kappa\upsilon\rho_1 & \ldots & \kappa\upsilon\rho_{m-1} & \kappa\upsilon\rho_m \\ 1 & 0 & \ldots & 0 & 0 & \ldots & 0 & 0 \\ 0 & 1 & \ldots & 0 & 0 & \ldots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \ldots & 0 & 0 & \ldots & 0 & 0 \\ 0 & 0 & \ldots & 0 & 1 & \ldots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \ldots & 0 & 0 & \ldots & 1 & 0 \end{bmatrix}, \; \boldsymbol{H} = \begin{bmatrix} \kappa\upsilon & 0 \\ 0 & 0 \\ 0 & 0 \\ \vdots & \vdots \\ 0 & 1 \\ 0 & 0 \\ \vdots & \vdots \\ 0 & 0 \end{bmatrix},$$

where $\boldsymbol{b}_t$ is $2m \times 1$, $\boldsymbol{T}_t$: $2m \times 2m$ and $\boldsymbol{H}$: $2m \times 2$. It is important to note that Catania (2020) assumes that $\rho_0 = 0$ as otherwise we would obtain a non-linear system for $\log(y_t - \mu)^2$ rendering the efficient Kalman filter obsolete. Despite the restriction, we can assume ξ_t and $\boldsymbol{w}_t$ to be normally distributed. Given that we know μ^*, γ^* and σ_ξ^2, the QML estimator is obtained by maximizing the prediction error decomposition via the Kalman filter:

$$\hat{\boldsymbol{\theta}}_{\text{QML}} = \arg\max_{\theta} \sum_{t=1}^{T} \left[\log var(y_t^* \,|\mathcal{F}_{t-1}) + \frac{(y_t^* - E(y_t^* \,|\mathcal{F}_{t-1}))^2}{var(y_t^* \,|\mathcal{F}_{t-1})} \right]. \tag{23}$$

4.5 Bellman Filter

The Bellman filter introduced by Lange (2020) is a computationally efficient filtering algorithm based on the dynamic programming technique of Bellman (1956). The Bellman filter generalizes the standard Kalman filter and iterated extender Kalman filter for non-linear state space models. Since the algorithm of Lange (2020) specifically requires that the state equation is linear and Gaussian, we apply this method only to the SV model with non-causal leverage.

The method aims to find a set of filtered states $\{\tilde{\boldsymbol{\alpha}}_{1:1}, \ldots, \tilde{\boldsymbol{\alpha}}_{T\,|\,T}\}$ by maximizing the complete data log likelihood $\ell(\boldsymbol{\alpha}_{1:T}, \boldsymbol{y}_{1:T}) = \log p(\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_t, \boldsymbol{y}_1, \ldots, \boldsymbol{y}_T)$. The key insight for making the Bellman filter computationally efficient is that we can recursively write the joint log-likelihood as

$$\ell(\boldsymbol{\alpha}_{1:t}, \boldsymbol{y}_{1:t}) = \quad \ell(\boldsymbol{y}_t\,|\boldsymbol{\alpha}_t) + \ell(\boldsymbol{\alpha}_t\,|\,\boldsymbol{\alpha}_{t-1}) + \ell(\boldsymbol{\alpha}_{1:t-1}, \boldsymbol{y}_{1:t-1}),$$

for $2 \leq t \leq T$. If we take the value function:

$$V_t(\boldsymbol{\alpha}_t) = \max_{\boldsymbol{\alpha}_{1:t-1}} \ell(\boldsymbol{\alpha}_{1:t}, \boldsymbol{y}_{1:t}),$$

then we can write the recursive Bellman equation with its solution definition as

$$\begin{aligned} V_t(\boldsymbol{\alpha}_t) &= \ell(\boldsymbol{y}_t\,|\boldsymbol{\alpha}_t) + \max_{\boldsymbol{\alpha}_{t-1}} \left\{\, \ell(\boldsymbol{\alpha}_t\,|\,\boldsymbol{\alpha}_{t-1}) + V_{t-1}(\boldsymbol{\alpha}_{t-1}) \,\right\}, \\ \boldsymbol{\alpha}_{t|t} &= \max_{\boldsymbol{\alpha}_{t-1}} V_t(\boldsymbol{\alpha}_t), \end{aligned} \tag{24}$$

where $\boldsymbol{\alpha}_{t|t} = \tilde{\boldsymbol{\alpha}}_{t|t}$ for all $t = 2, \ldots, T$.

Lange (2020) proposes to solve the Bellman equations for each time point by using a multivariate quadratic value function. However, the Bellman filter algorithm we use, works only in the state space framework (5), where the state is linear and Gaussian. The observation equation may involve nonlinearity. Therefore, this filtering method perfectly suits the SV model with non-causal leverage given in (9). The prediction and updating steps are shown in algorithm 2 on the next page. For derivations of the necessary score and the information matrix, we refer to appendix B.

In order to obtain feasible estimates of the unknown parameters $\boldsymbol{\theta}$, Lange (2020) decomposes the marginal likelihood into a "fit" term and a "penalty" term, which is the difference between the filtered and predicted states. The components are easily computed by the output of Algorithm 2. Finally, the Bellman Filter estimates are obtained as:

$$\hat{\boldsymbol{\theta}}_{\mathrm{BF}} = \arg\max_{\theta} \sum_{t=1}^{T} \left[\ell(\boldsymbol{y}_t\,|\boldsymbol{\alpha}_t) + \frac{1}{2}\log\det(\boldsymbol{I}_{t\,|\,t}^{-1}\boldsymbol{I}_{t\,|\,t-1}) - \frac{1}{2}\,(\boldsymbol{\alpha}_{t\,|\,t} - \boldsymbol{\alpha}_{t\,|\,t-1})'\boldsymbol{I}_{t\,|\,t-1}(\boldsymbol{\alpha}_{t\,|\,t} - \boldsymbol{\alpha}_{t\,|\,t-1}) \right]. \tag{25}$$

Algorithm 2: Bellman Filter

For $t = 0$*:*

1. *Initialization step:*
   Set $\boldsymbol{\alpha}_{0|0} = (\mathbf{1} - \boldsymbol{T})^{-1}\boldsymbol{c}$ *and* $vec(\boldsymbol{I}_{0|0}) = (\mathbf{1} - \boldsymbol{T} \otimes \boldsymbol{T})^{-1} vec(\boldsymbol{Q})$.

For $t = 1, \ldots, T$*:*

2. *Prediction step:*
   $\boldsymbol{\alpha}_{t|t-1} = \boldsymbol{c} + \boldsymbol{T}\boldsymbol{\alpha}_{t-1|t-1}$.
   $\boldsymbol{I}_{t|t-1} = (\boldsymbol{T}\boldsymbol{I}_{t-1|t-1}^{-1}\boldsymbol{T}' + \boldsymbol{Q})^{-1}$.

3. *Optimization step:*
   Set $\boldsymbol{\alpha}_{t|t}^{(0)} = \boldsymbol{\alpha}_{t|t-1}$. *Choose one of the three steps for* $i = 1, \ldots, i_{max}$ *:*

   Newton :

$$\boldsymbol{\alpha}_{t|t}^{(i)} = \boldsymbol{\alpha}_{t|t}^{(i-1)} + \left\{ \boldsymbol{I}_{t|t} - \frac{d^2\ell(\boldsymbol{y}_t|\,\boldsymbol{\alpha})}{d\boldsymbol{\alpha} d\boldsymbol{\alpha}'} \right\}^{-1} \left\{ \frac{d\ell(\boldsymbol{y}_t|\,\boldsymbol{\alpha})}{d\boldsymbol{\alpha}} - \boldsymbol{I}_{t|t-1}(\boldsymbol{\alpha} - \boldsymbol{\alpha}_{t|t-1}) \right\} \Bigg|_{\boldsymbol{\alpha} = \boldsymbol{\alpha}_{t|t}^{(i-1)}}$$

   Fisher :

$$\boldsymbol{\alpha}_{t|t}^{(i)} = \boldsymbol{\alpha}_{t|t}^{(i-1)} + \left\{ \boldsymbol{I}_{t|t} + \mathbb{E}\left[- \frac{d^2\ell(\boldsymbol{y}_t|\,\boldsymbol{\alpha})}{d\boldsymbol{\alpha} d\boldsymbol{\alpha}'} \,|\, \boldsymbol{\alpha} \right] \right\}^{-1} \left\{ \frac{d\ell(\boldsymbol{y}_t|\,\boldsymbol{\alpha})}{d\boldsymbol{\alpha}} - \boldsymbol{I}_{t|t-1}(\boldsymbol{\alpha} - \boldsymbol{\alpha}_{t|t-1}) \right\} \Bigg|_{\boldsymbol{\alpha} = \boldsymbol{\alpha}_{t|t}^{(i-1)}}$$

   BHHH :

$$\boldsymbol{\alpha}_{t|t}^{(i+1)} = \boldsymbol{\alpha}_{t|t}^{(i)} + \left\{ \boldsymbol{I}_{t|t} + \frac{d\ell(\boldsymbol{y}_t|\,\boldsymbol{\alpha})}{d\boldsymbol{\alpha}} \frac{d\ell(\boldsymbol{y}_t|\,\boldsymbol{\alpha})}{d\boldsymbol{\alpha}'} \right\}^{-1} \left\{ \frac{d\ell(\boldsymbol{y}_t|\,\boldsymbol{\alpha})}{d\boldsymbol{\alpha}} - \boldsymbol{I}_{t|t-1}(\boldsymbol{\alpha} - \boldsymbol{\alpha}_{t|t-1}) \right\} \Bigg|_{\boldsymbol{\alpha} = \boldsymbol{\alpha}_{t|t}^{(i)}}$$

4. *Update step:*
   Set $\boldsymbol{\alpha}_{t|t} = \boldsymbol{\alpha}_{t|t}^{i_{max}}$.

   Newton :

$$\boldsymbol{I}_{t|t} = \boldsymbol{I}_{t|t-1} - \frac{\mathrm{d}^2\ell(\boldsymbol{y}_t|\,\boldsymbol{\alpha})}{\mathrm{d}\boldsymbol{\alpha}\mathrm{d}\boldsymbol{\alpha}'} \Bigg|_{\boldsymbol{\alpha} = \boldsymbol{\alpha}_{t|t}}$$

   Fisher :

$$\boldsymbol{I}_{t|t} = \boldsymbol{I}_{t|t-1} + \mathbb{E}\left[- \frac{\mathrm{d}^2\ell(\boldsymbol{y}_t|\,\boldsymbol{\alpha})}{\mathrm{d}\boldsymbol{\alpha}\mathrm{d}\boldsymbol{\alpha}'} \,|\, \boldsymbol{\alpha} \right]_{\boldsymbol{\alpha} = \boldsymbol{\alpha}_{t|t}}$$

   BHHH :

$$\boldsymbol{I}_{t|t} = \boldsymbol{I}_{t|t-1} + \frac{\mathrm{d}\ell(\boldsymbol{y}_t|\,\boldsymbol{\alpha})}{\mathrm{d}\boldsymbol{\alpha}} \frac{\ell(\boldsymbol{y}_t|\,\boldsymbol{\alpha})}{d\boldsymbol{\alpha}'} \Bigg|_{\boldsymbol{\alpha} = \boldsymbol{\alpha}_{t|t}}$$

5 Simulation Study

In this section, we investigate the finite sample performance of the likelihood-based estimators by performing Monte Carlo simulations. Particularly, we focus on validating the consistency of the parameter estimates obtained with the aforementioned estimation methods. For this purpose, we repeatedly generate artifical observations $\tilde{y}_1, \ldots, \tilde{y}_T$ from two data generating processes which are given in equations (6) and (9). We then estimate the model parameters by using the replicated data sets. This is done for different sample sizes T and different values m, the number of return shock lags. For comparison with the true parameter values, the root mean squared error (RMSE) is used. This indicator should decreases as we increase the sample size T.

5.1 Design

Since not every estimation approach is equally efficient, we need different numbers of replication R for each estimator. Following Catania (2020), we perform $R = 100$ MC replication for SML estimation and $R = 1000$ replications for QML. For the Bellman Filtering method we set $R = 100$ as done in the simulation study of Lange (2020). Furthermore, we consider sample sizes $T \in \{500, 1000, 2000, 5000\}$. The model parameters of the SV model of Catania (2020) are denoted by $\boldsymbol{\theta} = (\beta, \phi, \kappa, \boldsymbol{\rho})$. We set $\beta = 1$, $\phi = 0.975$, $\kappa = 0.1$, $\mu = 0$ and consider three cases for the SML and QML estimators:

Case I: $m = 0$ ***for SML and*** $m = 1$ ***for QML****:*
$\rho_0^{SML} = -0.8$, $\rho_1^{QML} = -0.8$

Case II: $m = 1$ ***for SML and*** $m = 2$ ***for QML****:*
$\rho_0^{SML} = -0.8, \rho_1^{SML} = -0.5,\ \rho_1^{QML} = -0.8,\ \rho_2^{QML} = -0.5$

Case III: $m = 2$ ***for SML and*** $m = 3$ ***for QML***:
$\rho_0^{SML} = -0.8, \rho_1^{SML} = -0.5, \rho_2^{SML} = -0.3$, $\rho_1^{QML} = -0.8, \rho_2^{QML} = -0.5,\ \rho_3^{QML} = -0.3$

We use the above DGP values, just to replicate the simulation results of Catania (2020). However, the focus is more on the model parameter of the SV model with non-causal leverage for which we estimate the model parameters using the Bellman filter (BF).

Similarly, we consider three cases for the SV model with non-causal leverage. We set $c = 0$, $\varphi = 0.975$, $\sigma_\eta^2 = 0.01$ and $\mu = 0$. We estimate in Case I, the standard SV model without leverage. We estimate in Case II, the 2-dimensional version of the anti-causal SV: $\rho_1 = -0.5$. Lastly, we estimate in Case III, the 4-dimensional anti-causal SV model where the correlation parameter are set to $\rho_{-1} = -0.3$, $\rho_0 = -0.8$, $\rho_1 = -0.5$.

The steps that we take to generate log returns from the anti-causal SV model for the monte carlo simulations are:

(i) Draw T random return shocks $\varepsilon_1, \ldots, \varepsilon_T$ and $T + 2$ random volatility shocks $\eta_0, \ldots, \eta_{T+1}$ from a standard normal distribution.

(ii) Construct new return shocks $\tilde{\varepsilon}_1, \ldots, \tilde{\varepsilon}_T$ by using the non-causal formula

$$\tilde{\varepsilon}_t \;=\; \left[\sum_{i\,=-1}^{+1} \rho_i \, \eta_{t+i} \right] + \sigma_\varepsilon \, \varepsilon_t,$$

where $\sigma_\varepsilon = \sqrt{1 - \sum_{i=-1}^{+1} \rho_i^2}$.

(iii) Furthermore, recursively construct a series of log-volatilities $h_1, \ldots, h_T$ by using the state transition equation:

$$h_t \;=\; c + \varphi h_{t-1} + \sigma_\eta \, \eta_t, \tag{26}$$

where the log-volatility process starts with $h_0 = E(h_t) = 0$.

(iv) Afterwards, the set of artificial observations $\tilde{y}_1, \ldots, \tilde{y}_T$, are then obtained with the observation equation:

$$\tilde{y}_t \;=\; \exp\left\{ \frac{h_t}{2} \right\} \tilde{\varepsilon}_t. \tag{27}$$

(v) Apply the BF estimation method on the generated data set $\{\tilde{y}_t\}_{t=1}^T$ and store the obtained model parameter estimates $\hat{\boldsymbol{\theta}}_{\mathrm{BF}}$.

(vi) Repeat steps (i) till (v) R times.

Simulating log-returns from the SV model of Catania (2020) is analogue. The steps that we take to perform MC simulations for the SV model with general leverage structure are:

(i) Draw $T + m$ random return shocks $\varepsilon_{1-m}, \ldots \varepsilon_0, \varepsilon_1, \ldots, \varepsilon_T$ and T random shocks $b_1, \ldots, b_T$ from a standard normal distribution.

(ii) Construct the volatility shocks $\eta_t, \ldots, \eta_T$ by using the general leverage formula:

$$\eta_t \;=\; \sum_{j\,=\,0}^{m} \rho_j\, \varepsilon_{t-j} \;+ \sigma_b\, b_t,$$

where $\sigma_b = \sqrt{1 - \sum_{j\,=\,0}^{m} \rho_j^2}$.

(iii) Furthermore, recursively construct a series of log-volatilities $h_1, \ldots, h_T$ by using the state transition equation:

$$h_t \;=\; \phi h_{t-1} + \kappa\, \eta_t,$$

where the log-volatility process starts with $h_0 = E(h_t) = 0$.

(iv) Afterwards, the set of artificial observations $\tilde{y}_1, \ldots, \tilde{y}_T$, are then obtained with the observation equation:

$$\tilde{y}_t \;=\; \beta\, \exp\left\{\frac{h_t}{2}\right\} \varepsilon_t.$$

(v) Apply the QML and SML estimation methods on the generated data set $\{\tilde{y}_t\}_{t\,=1}^{T}$ and store the obtained model parameter estimates $\boldsymbol{\theta}_{\text{QML}}$ and $\boldsymbol{\theta}_{\text{SML}}$

(vi) Repeat steps (i) till (v) R times.

5.2 Results

By applying the three parameterization cases for the QML and SML methods described in Section 5.1, we obtain Monte Carlo estimates for the parameters of six different SV models. These estimates including their RMSE are shown in tables 1 and 2. Furthermore, the use of the Bellman filter regarding the SV model with anti-causal leverage rendered valid Monte Carlo estimates as well. The results are formatted similarly in table 3. From tables 1 to 3 we can confirm that all three method provide valid and pretty consistent estimates when T increases. From computational point, the QML approach is the fastest and the SML approach the slowest whereas BF is in between and is the most stable.

Table 1: Monte Carlo simulation results for QML estimators without mean μ.

T	β	ϕ	κ	ρ_1	ρ_2	ρ_3	Computing Time
Case I: $m = 1$							
500	1.006	0.728	0.133	-0.871	-	-	2.43
	(0.140)	(0.473)	(0.209)	(0.289)	-	-	-
1000	1.001	0.928	0.120	-0.846	-	-	3.98
	(0.065)	(0.183)	(0.110)	(0.224)	-	-	-
2000	1.001	0.971	0.103	-0.829	-	-	6.71
	(0.041)	(0.014)	(0.029)	(0.142)	-	-	-
5000	1.001	0.974	0.101	-0.807	-	-	15.15
	(0.025)	(0.006)	(0.016)	(0.080)	-	-	-
Case II: $m = 2$							
500	1.001	0.945	0.105	-0.899	-0.501	-	4.88
	(0.122)	(0.135)	(0.061)	(0.134)	(0.139)	-	-
1000	1.001	0.970	0.104	-0.894	-0.450	-	9.26
	(0.068)	(0.036)	(0.036)	(0.152)	(0.210)	-	-
2000	1.002	0.974	0.105	-0.874	-0.415	-	17.87
	(0.040)	(0.007)	(0.027)	(0.187)	(0.265)	-	-
5000	1.000	0.975	0.110	-0.852	-0.349	-	43.83
	(0.023)	(0.004)	(0.023)	(0.189)	(0.335)	-	-
Case III: $m = 3$							
500	0.998	0.963	0.099	-0.854	-0.523	-0.296	6.542
	(0.084)	(0.078)	(0.040)	(0.068)	(0.079)	(0.076)	-
1000	1.001	0.972	0.101	-0.851	-0.513	-0.285	12.902
	(0.070)	(0.012)	(0.027)	(0.081)	(0.098)	(0.094)	-
2000	1.002	0.974	0.101	-0.843	-0.510	-0.264	25.706
	(0.046)	(0.006)	(0.020)	(0.082)	(0.110)	(0.120)	-
5000	1.000	0.975	0.103	-0.827	-0.517	-0.227	62.676
	(0.023)	(0.003)	(0.016)	(0.108)	(0.117)	(0.153)	-

Note: The results of the Monte Carlo simulations with QML method are based on 1000 replicates. The true values are $\beta = 1, \phi = 0.975$, $\kappa = 0.1$ and $\mu = 0$. For case I: $\rho_1 = -0.8$, case II: $\rho_1 = -0.8, \rho_2 = -0.5$, and case III: $\rho_1 = -0.8, \rho_2 = -0.5, \rho_3 = -0.3$. The entries in brackets are the root mean squared errors. The last column represents the average computing time per sample.

Table 2: Monte Carlo simulation results for SML estimators without mean μ.

T	β	ϕ	κ	ρ_0	ρ_1	ρ_2	Computing Time
Case I: $m = 0$							
500	0.997	0.964	0.105	-0.804	-	-	200.85
	(0.029)	(0.032)	(0.031)	(0.028)	-	-	-
1000	0.998	0.973	0.098	-0.802	-	-	375.66
	(0.025)	(0.009)	(0.014)	(0.013)	-	-	-
2000	0.999	0.975	0.100	-0.802	-	-	750.17
	(0.016)	(0.005)	(0.012)	(0.014)	-	-	-
5000	0.998	0.975	0.100	-0.800	-	-	1925.90
	(0.009)	(0.003)	(0.007)	(0.006)	-	-	-
Case II: $m = 1$							
500	1.001	0.973	0.096	-0.799	-0.500	-	225.62
	(0.006)	(0.014)	(0.011)	(0.005)	(0.005)	-	-
1000	1.000	0.976	0.096	-0.800	-0.500	-	425.58
	(0.005)	(0.004)	(0.008)	(0.004)	(0.004)	-	-
2000	1.000	0.976	0.098	-0.800	-0.500	-	875.56
	(0.002)	(0.003)	(0.006)	(0.003)	(0.002)	-	-
5000	1.000	0.976	0.098	-0.799	-0.499	-	2125.80
	(0.003)	(0.002)	(0.004)	(0.004)	(0.003)	-	-
Case III: $m = 2$							
500	0.999	0.976	0.094	-0.792	-0.495	-0.296	250.49
	(0.012)	(0.006)	(0.012	(0.021)	(0.013)	(0.013)	-
1000	1.000	0.978	0.094	-0.794	-0.497	-0.297	500.79
	(0.006)	(0.004)	(0.009)	(0.013)	(0.009)	(0.007)	-
2000	1.001	0.977	0.096	-0.794	-0.496	-0.298	975.16
	(0.004)	(0.003)	(0.007)	(0.011)	(0.008)	(0.005)	-
5000	1.000	0.977	0.095	-0.793	-0.496	-0.297	2473.30
	(0.004)	(0.003)	(0.007)	(0.012)	(0.008)	(0.006)	-

Note: The results of the Monte Carlo simulations with SML method are based on 100 replicates. The true values are $\beta = 1, \phi = 0.975$, $\kappa = 0.1$ and $\mu = 0$. For case I: $\rho_0 = -0.8$, case II: $\rho_0 = -0.8, \rho_1 = -0.5$, and case III: $\rho_0 = -0.8, \rho_1 = -0.5, \rho_2 = -0.3$. The entries in brackets are the root mean squared errors. The last column represents the average computing time in seconds per sample.

Table 3: Monte Carlo simulation results for BF estimators without mean μ.

T	c	φ	σ_η^2	ρ_{-1}	ρ_0	ρ_1	Computing Time
Case I: $\rho_1 = \rho_0 = \rho_{-1} = 0$							
500	-0.009	0.915	0.027	-	-	-	0.48
	(0.070)	(0.199)	(0.068)	-	-	-	-
1000	0.002	0.953	0.016	-	-	-	0.51
	(0.007)	(0.060)	(0.028)	-	-	-	-
2000	0.002	0.966	0.014	-	-	-	0.76
	(0.004)	(0.020)	(0.010)	-	-	-	-
5000	0.003	0.970	0.013	-	-	-	1.56
	(0.003)	(0.009)	(0.005)	-	-	-	-
Case II: $\rho_0 = \rho_{-1} = 0$							
500	0.000	0.956	0.017	-	-	-0.568	18.14
	(0.012)	(0.045)	(0.090)	-	-	(0.285)	-
1000	0.001	0.957	0.012	-	-	-0.501	33.41
	(0.006)	(0.072)	(0.090)	-	-	(0.197)	-
2000	0.001	0.970	0.012	-	-	-0.478	61.77
	(0.003)	(0.011)	(0.089)	-	-	(0.116)	-
5000	0.002	0.972	0.012	-	-	-0.491	140.19
	(0.002)	(0.007)	(0.088)	-	-	(0.063)	-
Case III: $\rho_i \neq 0$ for $i \in \{-1, 0, 1\}$							
500	0.000	0.969	0.011	-0.310	-0.790	-0.500	58.14
	(0.003)	(0.014)	(0.089)	(0.070)	(0.028)	(0.083)	-
1000	0.000	0.973	0.010	-0.319	-0.801	-0.481	59.25
	(0.002)	(0.006)	(0.090)	(0.052)	(0.022)	(0.069)	-
2000	0.000	0.974	0.010	-0.314	-0.801	-0.484	174.36
	(0.001)	(0.004)	(0.090)	(0.042)	(0.018)	(0.056)	-
5000	0.000	0.975	0.010	-0.305	-0.801	-0.493	391.97
	(0.001)	(0.002)	(0.090)	(0.026)	(0.011)	(0.035)	-

Note: The results of the Monte Carlo simulations with BF method are based on 100 replicates. The true values are $c = 0, \varphi = 0.975$, $\sigma_\eta^2 = 0.01$ and $\mu = 0$. For case I: no leverage, case II: $\rho_1 = -0.5$, and case III: $\rho_{-1} = -0.3, \rho_0 = -0.8, \rho_1 = -0.5$. The entries in brackets are the root mean squared errors. The last column represents the average computing time in seconds per sample.

6 Empirical Study

The empirical study that we perform is similar to that of Catania (2020) and is two-fold. Firstly, we consider the in-sample results where we estimate our univariate SV models on the whole data sample. Secondly, we perform out-of-sample volatility prediction analysis to compare the forecast accuracy of the SV models with leverage against the benchmark GARCH(1,1) model.

6.1 Data

For the purpose of the in-sample study and out-of-sample forecasting analysis, equity index returns are used. Particularly, we consider two indices: the Standard & Poor's 500 index (SP500) and the Amsterdam Exchange index (AEX). The raw data set consists of daily prices for the two index return series, where the AEX prices are obtained from the Oxford-Man institute's realized library and SP500 from the Wharton research data services. As the raw data are closing prices, the index returns are transformed to log-returns by calculating the differences in the natural logarithm of the closing prices. The data range in this case is 2010/01/04 - 2020/12/31. This leaves us with a data set of $T = 2618$ observations for SP500 and $T = 2730$ observations for AEX.

Table 4: Summary statistics of the log returns of the AEX and S&P500 indices.

	S&P 500	AEX
Mean	0.056	0.020
Std.Dev.	1.106	1.134
Skewness	−0.679	−0.512
Kurtosis	17.986	10.188

Note : This table shows descriptive statistics of the AEX and S&P 500 indices over the full sample period: 2010/01/04 - 2020/12/31.

In table 4, we present summary statistics for both indices. Moreover, in Figures 1 and 2 the development of the price and log-returns of the S&P 500 and AEX indices are shown. From the plots and table, we confirm our notion that the stylized facts of returns hold, i.e. high kurtosis, negative skewness, volatility clustering. Interestingly, sharp drops of the price of both indices in for example 2011 and 2020, correspond with periods of large return swings, indicating the presence of the leverage effect.

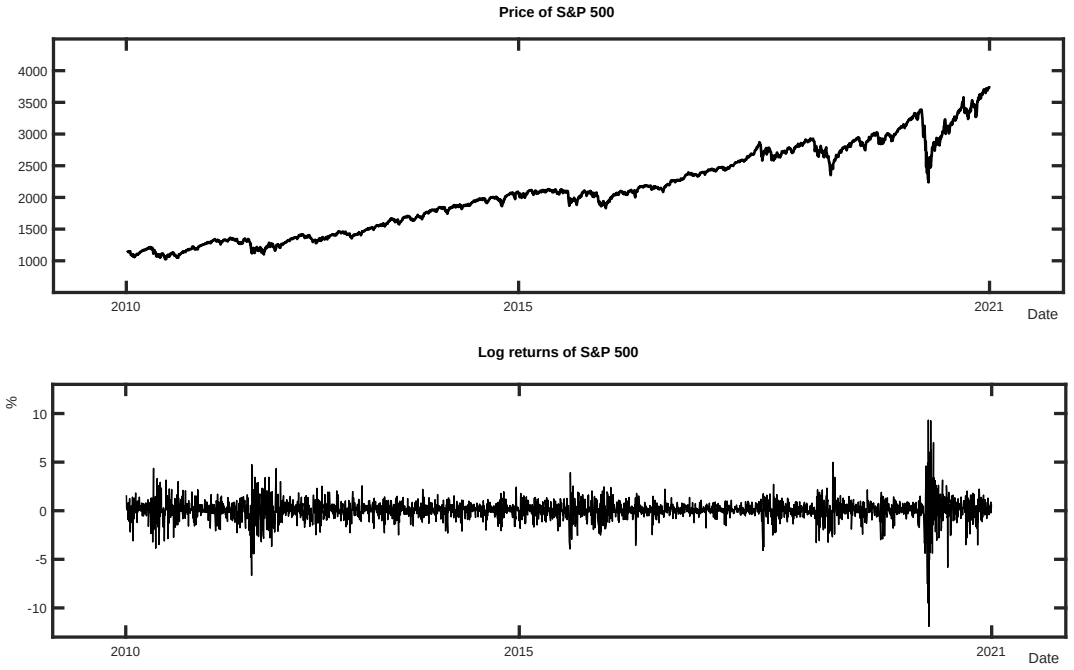

Figure 1: The level and log returns of the S&P 500 index over sample period 2010/01/04 - 2020/12/31.

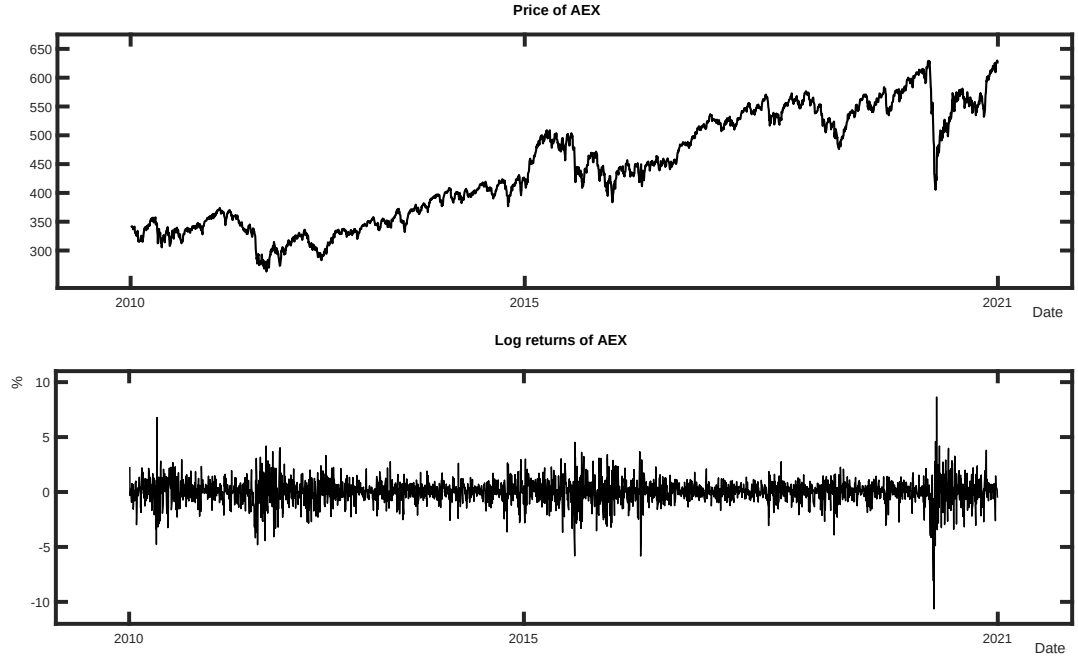

Figure 2: The level and log returns of the AEX over sample period 2010/01/04 - 2020/12/31.

6.2 Full sample performance

We estimate the model parameters of the SV models with general and non-causal leverage structures using the three methods on the full data set. This is done with and without a mean term μ, since we expect that the mean parameter μ will off-set the poorly estimated leverage parameter ρ_0. Especially for the SV model with non-causal, this should make a difference. We determine the SV model with the best fit to our data set by considering the BIC model selection criterion.

The full sample estimation results without μ are given in table 5. It show the maximum likelihood estimates of the SV model with general leverage (QML/SML) and SV model with non-causal leverage (BF) on both the S&P 500 and the AEX index. Furthermore the in-sample fit criterion BIC is included. For each method the model with the lowest BIC value is chosen to be shown in the table. This means that for SML, the BIC criterion chooses Catania's SV model with $m = 1$ and $m = 3$ lags for the S&P 500 and AEX index, respectively. For QML, we obtain the best fit when $m = 1$ for both the S&P 500 and AEX index but the restriction ρ_0 is imposed causing the BIC values to be way higher than the BIC values obtained with SML. For the BF method, the 4-dimensional non-causal SV model gives the second to lowest BIC values. The best fit of all three approaches are obtained when we use the SML approach. For both the S&P 500 index and AEX index, the SML approach renders the best fit of BIC values of 6331 and 7384, respectively. This means that neither of the intertemporal SVL model and contemporaneous SVL model is a good fit to the returns of the S&P 500 and AEX indices. Despite that, the SV model with non-causal leverage still give a comparable full-sample fit value of BIC = 6347 for S&P 500 and BIC = 7408 for AEX.

Table 6 similarly shows parameter estimates of the three methods on the full sample but now we estimate with an additional mean parameter μ. What we generally observe is that the full sample fit slightly increases for each model (and method). Also, the ranking stays unchanged (SML gives best BIC values for both S&P 500 and AEX). More importantly, we see that in table 6, the leverage parameters ρ_0 of the non-causal SV model for both S&P 500 and AEX are significantly estimated as -0.736 and -0.731. In table 5 (without a mean term), the leverage parameters ρ_0 are -0.013 and -0.093. This confirms the suspicion that the mean parameter μ indeed influences the estimate of leverage parameter ρ_0. The inclusion of a mean term μ therefore actually improves the ability of the non-causal SV model to capture the leverage effect. For SML and QML, the leverage parameter parameters do not change considerably.

Table 5: Parameter estimates of the univariate SV models with leverage from the SML/QML/BF estimators

	β	ϕ	κ	ρ_0	ρ_1	ρ_2	ρ_3	ρ_4	ρ_5	BIC
SML General SV										
S&P 500	1.012	0.946	0.314	−0.336	−0.669	———	———	———	———	6331
	(0.001)	(0.001)	(0.001)	(0.002)	(0.001)	———	———	———	———	–
AEX	1.051	0.972	0.242	−0.403	−0.698	0.047	0.273	———	———	7384
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	———	———	–
QML General SV										
S&P 500	0.921	0.948	0.379	–	−0.731	———	———	———	———	12225
	(0.060)	(0.008)	(0.034)	–	(0.045)	———	———	———	———	–
AEX	1.015	0.979	0.211	–	−0.862	———	———	———	———	12406
	(0.067)	(0.004)	(0.022)	–	(0.046)	———	———	———	———	–
BF Non - Causal SV										
	c	ϕ	σ_η	ρ_{-1}	ρ_0	ρ_1				BIC
S&P 500	0.005	0.939	0.157	−0.137	−0.013	−0.639	———	———	———	6347
	(0.007)	(0.008)	(0.022)	(0.036)	(0.030)	(0.051)	———	———	———	–
AEX	0.004	0.951	0.090	−0.044	−0.093	−0.647	———	———	———	7408
	(0.004)	(0.006)	(0.010)	(0.040)	(0.046)	(0.041)	———	———	———	–

Note : This table shows parameter estimates of the univariate SV models with leverage selected by BIC. Entries in parenthesis are the asymptotic standard errors. Empirical log returns on the S&P 500 and AEX are used over the full sample period: 2010/01/04 - 2020/12/31 ($T = 2618$ for SP500 and $T = 2730$ for AEX). For SML the standard errors are unstable. For QML we set $\rho_0 = 0$.

Table 6: Parameter estimates of the univariate Leverage SV models with mean μ from SML/QML/BF estimators

	β	ϕ	κ	μ	ρ_0	ρ_1	ρ_2	ρ_3	ρ_4	ρ_5	BIC
SML General SV											
S&P 500	0.831	0.962	0.293	0.065	−0.428	−0.519	——	——	——	——	6306
	(0.001)	(0.001)	(0.001)	(0.001)	(0.002)	(0.001)	——	——	——	——	–
AEX	1.000	0.978	0.250	0.033	−0.420	−0.693	0.082	0.295	——	——	7383
	(0.001)	(0.001)	(0.001)	(0.001)	(0.004)	(0.001)	(0.001)	(0.001)	——	——	–
QML General SV											
S&P 500	0.337	0.980	0.253	0.255	–	−0.587	——	——	——	——	11673
	(0.076)	(0.004)	(0.025)	(0.001)	–	(0.069)	——	——	——	——	–
AEX	1.109	0.980	0.196	−0.016	–	−0.884	0.009	——	——	——	12311
	(0.073)	(0.004)	(0.012)	(0.001)	–	(0.043)	(0.030)	——	——	——	–
BF Non - Causal SV											
	c	ϕ	σ_η	μ	ρ_{-1}	ρ_0	ρ_1				BIC
S&P 500	−0.0012	0.968	0.096	0.091	0.204	−0.736	−0.126	——	——	——	6319
	(0.006)	(0.006)	(0.012)	(0.011)	(0.039)	(0.035)	(0.040)	——	——	——	–
AEX	0.002	0.972	0.062	0.058	0.155	−0.731	−0.184	——	——	——	7397
	(0.005)	(0.006)	(0.011)	(0.015)	(0.041)	(0.045)	(0.043)	——	——	——	–

Note : This table shows parameter estimates of the univariate SV models with leverage selected by BIC. The SV models now contain a mean term μ. Entries in parenthesis are the asymptotic standard errors. Empirical log returns on the S&P 500 and AEX are used over the full sample period: 2010/01/04 - 2020/12/31 ($T = 2618$ for SP500 and $T = 2730$ for AEX). For QML: $\rho_0 = 0$.

6.3 Forecasting Performance

For the comparison of the predictive performance of the different SV models, we do the following. The total data sample is equally split up such that the first $T/2$ observations are used for in-sample (IS) estimations and the last $T/2$ observations for out-of-sample (OOS) volatility evaluations. We construct one-day ahead volatility forecasts $\hat{\sigma}^2_{t\,|\,t-1} = \text{var}(y_t\,|\mathcal{F}_{t-1})$ using the outputs of the particle filter (SML), Kalman filter (QML) and the Bellman filter (BF). The forecast criteria that we employ are the robust Quasi-likelihood (QLIKE) and the usual mean squared forecast error (MSFE):

$$\text{QLIKE} \;=\; \sum_{t\,=\,T/2}^{T} \left(\log(\,\hat{\sigma}^2_{t\,|\,t-1}\,) \;+\; \frac{\tilde{\sigma}^2_{t+1}}{\hat{\sigma}^2_{t\,|\,t-1}} \right),$$

$$\text{MSFE} \;=\; \sum_{t\,=\,T/2}^{T} \left(\hat{\sigma}^2_{t\,|\,t-1} \;-\; \tilde{\sigma}^2_{t+1}\right)^2,$$

where $\tilde{\sigma}^2_{t+1}$ is the "observed" volatility. For $\tilde{\sigma}^2_{t+1}$, we use two proxies: the squared returns and the 5-min. intra-day realized variances. The latter can be obtained from Oxford-Man institute's realized library.[2]

In order to confirm that improvements in volatility predictions are significant, we make use of a test for equal predictive ability proposed by Diebold-Mariano (1995). For the two loss functions (QLIKE and MSFE) denoted L, the out-of-sample average of the loss differentials between models i and j is computed as:

$$\bar{d} \;=\; \frac{1}{(T/2)} \sum_{t\,=\,T/2}^{T} \left(L_{i,t} - L_{j,t}\right).$$

We use the formula mentioned in Harvey et al. (1997) to obtain the asymptotic variance of $\bar{d}$. Additionally, we also take into account the fact that the Diebold-Mariano statistic (DM) is large for small sample sizes. Given the forecast horizon $h = 1$ and sample size $T/2$, the adjusted Diebold-Mariano statistic becomes:

$$S_1^* \;=\; \left[\,\frac{T-2}{2}\,\right]^{1/2} \frac{\bar{d}}{\sqrt{\hat{\gamma}_0}}.$$

Here, $\hat{\gamma}_0$ is the out-of-sample variance of the loss differentials. The asymptotic distribution of the adjusted DM test is a t-distribution with $(T-2)/2$ degrees of freedom.

[2] https://realized.oxford-man.ox.ac.uk/data/download

6.4 Results

In this section, we present the results of the out-of-sample volatility prediction analysis. Since two proxies are used for the "true" volatility with and without a mean term μ, we consider four separate cases. We start off with results of the 5-min. intra-day realized variance proxy and end with the noisy squared returns proxy.

Table 7 shows the absolute and relative QLIKE and MSFE values for all three methods. The relative values are used using the absolute QLIKE and MSFE value of the benchmark GARCH(1,1) model. For QML, we estimate and predict with the SV general leverage models for values of $m = 1$ up to $m = 10$ since it is quickly done due to the use of the Kalman filter. However, for SML, we only use values $m = 0$ till $m = 5$ due to the long computation time. For BF, we consider the 2-dimensional and 4-dimensional SV model with non-causal leverage[3]. Generally, what we can see from table 6 is that the SV model with non-causal leverage significantly outperforms all other leverage SV model, including the GARCH(1,1), with the exception for the QLIKE on the AEX index. In that case, the QML approach with $m = 6$ seems to be the best at predicting the volatility, but the QLIKE value does not differ significantly from the QLIKE values obtained from the non-causal SV models (0.436 vs. 0.452). Another neat thing of table 6 is that all three estimation approaches deliver significantly better volatility predictions than the asymmetrics GARCH(1,1). This also holds for the standard GARCH(1,1) model, but QML approach seems to produce worse volatility prediction for the AEX index when the MSFE measure is used. This could be due to possible outliers in the AEX index (e.g. corona-crisis). These observations provide evidence for the claim that stochastic volatility models are significantly superior to the observation driven models. Table 8 displays similar results, but now the models are estimated with an additional mean term μ. The ranking of the model remains largely unchanged with the exception for the QLIKE measure on the AEX index. Here the best predictive model now becomes the 4-dimensional non-causal SV model instead of the 2-dim. non-causal SV model. The volatility predictions of the non-causal SV models seem to improve from adding a mean term (e.g. from 0.059 to 0.046 for QLIKE on SP500). On the other hand the QLIKE and MSFE values (volatility predictions) obtained with QML seem to become larger (worse). The same holds for the SML approach.

Tables 9 and 10 are similar to tables 7 and 8. They show absolute and relative QLIKE/MSFE values with and without mean μ, but now with the noisy squared returns as proxy. Here, the results are different as expected. There seems to be (significant)

[3]See appendix B for more details on the 2-dim. SV model.

dominance of Catania's SV model used with SML for the QLIKE values on both indices. However, for the MSFE measure on the AEX, only the SML approach renders significant volatility predictions. For the S&P 500 index, the asymmetric GARCH(1,1) model seems to do the best in that case with a MSFE value of 29.010. This applies to both the cases with and without a mean term. The worst volatility predictions are obtained with the QML method since all the relative QLIKE and MSFE values are larger than 1 (both with and without mean term). This means the SV model with non-causal leverage (BF) performs relatively worse to Catania's SV model with QML when using the noisy squared returns. However, the difference again are not (significantly) large. The other relative QLIKE and MSFE values are interpreted similarly. Adding a mean term results in significant volatility predictions only for the QLIKE measure on the S&P 500 index. The important thing to note is that the results for the squared returns are very unreliable due to the high variance of this noisy estimator. Therefore, the results with the 5-min. realized variances should be preferred over the noisy squared returns.

Table 7: Out-of-sample volatility forecasting results with 5 - min. realized variance - $\mu = 0$.

	QLIKE				MSFE			
	SP500		AEX		SP500		AEX	
Benchmark								
GARCH(1,1)	0.162	(1.000)	0.525	(1.000)	4.796	(1.000)	2.116	(1.000)
A-GARCH(1,1)	0.138	(0.852)	0.515	(0.981)	7.971	(1.662)	2.703	(1.277)
QML General SV								
$m = 1$	0.082	(0.504)	0.437	(0.832)	*1.549*	*(0.323)*	*2.441*	*(1.154)*
2	0.081	(0.501)	0.437	(0.831)	1.551	(0.323)	2.443	(1.154)
3	0.081	(0.498)	0.436	(0.831)	1.551	(0.323)	2.465	(1.165)
4	*0.080*	*(0.495)*	0.436	(0.830)	1.560	(0.325)	2.464	(1.165)
5	0.082	(0.507)	0.436	(0.830)	1.576	(0.329)	2.467	(1.166)
6	0.082	(0.504)	***0.436****	*(0.829)*	1.579	(0.329)	2.464	(1.165)
7	0.082	(0.504)	0.436	(0.830)	1.580	(0.329)	2.464	(1.164)
8	0.082	(0.507)	0.436	(0.830)	1.577	(0.329)	2.462	(1.164)
9	0.082	(0.504)	0.436	(0.831)	1.579	(0.329)	2.455	(1.160)
10	0.082	(0.505)	0.436	(0.831)	1.577	(0.329)	2.446	(1.156)
SML General SV								
$m = 0$	*0.093*	*(0.574)*	0.479	(0.912)	*1.375*	*(0.287)*	1.636	(0.773)
1	0.123	(0.759)	0.475	(0.905)	1.524	(0.318)	*1.381*	*(0.652)*
2	0.122	(0.753)	0.476	(0.907)	1.682	(0.351)	1.440	(0.680)
3	0.114	(0.704)	*0.466*	*(0.888)*	1.603	(0.334)	1.531	(0.724)
4	0.117	(0.722)	0.477	(0.909)	1.688	(0.352)	1.748	(0.826)
5	0.111	(0.685)	0.474	(0.903)	1.708	(0.356)	1.671	(0.790)
BF Non - Causal SV								
2-dim	0.062	(0.383)	*0.452*	*(0.861)*	***1.344****	*(0.280)*	***1.340****	*(0.633)*
4-dim	***0.059****	*(0.364)*	0.453	(0.863)	1.358	(0.283)	1.361	(0.643)

Note: The entries are the average Quasi-likelihood (QLIKE) and Mean Squared Forecast Error (MSFE) values computed using one-day ahead volatility predictions over the out-of-sample period. For SP500: 2015/06/05 - 2020/12/31 and AEX: 2015/06/25 - 2020/12/31. The numbers in parenthesis denote relative QLIKE and MSFE values where the benchmark is GARCH(1,1). Entries < 1 indicates outperformance of the benchmark. Numbers in bold are QLIKE- and MSFE- best and * indicates significant improvement of best model over the benchmark GARCH(1,1) at 1% level.

Table 8: Out-of-sample volatility forecasting results with 5 - min. realized variance - $\mu \neq 0$.

	QLIKE				MSFE			
	SP500		AEX		SP500		AEX	
Benchmark								
GARCH(1,1)	0.162	(1.000)	0.525	(1.000)	4.796	(1.000)	2.116	(1.000)
A-GARCH(1,1)	0.138	(0.852)	0.515	(0.981)	7.971	(1.662)	2.703	(1.277)
QML General SV								
$m = 1$	0.951	(5.872)	0.438	(0.833)	5.309	(1.107)	*2.446*	*(1.160)*
2	0.942	(5.817)	0.437	(0.833)	5.461	(1.139)	2.458	(1.161)
3	0.942	(5.817)	0.437	(0.832)	5.283	(1.102)	2.479	(1.172)
4	0.942	(5.814)	0.437	(0.832)	5.280	(1.101)	2.479	(1.172)
5	*0.939*	*(5.795)*	0.437	(0.832)	*5.228*	*(1.090)*	2.482	(1.173)
6	0.994	(6.133)	***0.437****	*(0.831)*	5.346	(1.115)	2.479	(1.172)
7	0.984	(6.077)	0.437	(0.832)	5.282	(1.101)	2.478	(1.171)
8	0.970	(5.985)	0.437	(0.832)	5.292	(1.103)	2.477	(1.171)
9	0.974	(6.015)	0.437	(0.832)	5.445	(1.135)	2.555	(1.207)
10	0.978	(6.035)	0.472	(0.898)	5.461	(1.139)	2.852	(1.348)
SML General SV								
$m = 0$	0.119	(0.735)	0.484	(0.922)	1.814	(0.378)	1.697	(0.802)
1	*0.116*	*(0.716)*	*0.469*	*(0.893)*	*1.652*	*(0.351)*	*1.379*	*(0.652)*
2	0.119	(0.735)	0.469	(0.893)	1.748	(0.364)	1.432	(0.677)
3	0.143	(0.883)	0.476	(0.907)	2.073	(0.432)	1.567	(0.805)
4	0.132	(0.815)	0.478	(0.911)	2.016	(0.420)	1.702	(0.805)
5	0.123	(0.759)	0.487	(0.928)	1.934	(0.403)	1.846	(0.872)
BF Non - Causal SV								
2-dim	0.059	(0.364)	0.452	(0.861)	1.328	(0.277)	***1.341****	*(0.634)*
4-dim	***0.046****	*(0.284)*	*0.441*	*(0.840)*	***1.262****	*(0.263)*	1.356	(0.641)

Note: The entries are the average Quasi-likelihood (QLIKE) and Mean Squared Forecast Error (MSFE) values computed using one-day ahead volatility predictions over the out-of-sample period. For SP500: 2015/06/05 - 2020/12/31 and AEX: 2015/06/25 - 2020/12/31. The numbers in parenthesis denote relative QLIKE and MSFE values where the benchmark is GARCH(1,1). Entries < 1 indicates outperformance of the benchmark. Numbers in bold are QLIKE- and MSFE- best and * indicates significant improvement of best model over the benchmark GARCH(1,1) at 1% level.

Table 9: Out-of-sample volatility forecasting results with squared returns - $\mu = 0$.

	QLIKE				MSFE			
	SP500		AEX		SP500		AEX	
Benchmark								
GARCH(1,1)	0.637	(1.000)	0.935	(1.000)	30.084	(1.000)	20.675	(1.000)
A-GARCH(1,1)	0.640	(1.005)	0.898	(0.960)	***29.010***	(0.964)	20.623	(0.997)
QML General SV								
$m = 1$	*0.748*	*(1.175)*	*0.987*	*(1.056)*	35.947	(1.195)	*21.388*	*(1.034)*
2	0.750	(1.177)	0.987	(1.056)	35.962	(1.195)	21.390	(1.035)
3	0.751	(1.180)	0.989	(1.058)	35.950	(1.195)	21.426	(1.036)
4	0.752	(1.181)	0.990	(1.058)	36.003	(1.197)	21.430	(1.037)
5	0.763	(1.198)	0.990	(1.059)	35.730	(1.188)	21.438	(1.037)
6	0.763	(1.197)	0.990	(1.059)	*35.716*	*(1.187)*	21.435	(1.037)
7	0.763	(1.197)	0.990	(1.058)	35.724	(1.187)	21.431	(1.037)
8	0.763	(1.198)	0.990	(1.058)	35.822	(1.191)	21.430	(1.037)
9	0.762	(1.197)	0.989	(1.058)	35.808	(1.190)	21.414	(1.038)
10	0.762	(1.197)	0.988	(1.057)	35.767	(1.189)	21.399	(1.035)
SML General SV								
$m = 0$	0.599	(0.940)	0.890	(0.952)	33.618	(1.117)	20.071	(0.971)
1	0.609	(0.956)	0.884	(0.945)	31.690	(1.053)	19.687	(0.952)
2	0.607	(0.953)	***0.881****	*(0.943)*	31.727	(1.055)	***19.684****	(0.952)
3	***0.596***	*(0.936)*	0.882	(0.943)	31.489	(1.047)	19.883	(0.962)
4	0.597	(0.937)	0.884	(0.945)	*31.408*	*(1.044)*	19.986	(0.967)
5	0.600	(0.942)	0.883	(0.945)	31.727	(1.055)	19.939	(0.964)
BF Non - Causal SV								
2-dim	0.633	(0.994)	*0.897*	*(0.959)*	*31.781*	*(1.056)*	*19.993*	*(0.967)*
4-dim	*0.631*	*(0.991)*	0.897	(0.960)	31.789	(1.057)	20.041	(0.969)

Note: The entries are the average Quasi-likelihood (QLIKE) and Mean Squared Forecast Error (MSFE) values computed using one-day ahead volatility predictions over the out-of-sample period. For SP500: 2015/06/05 - 2020/12/31 and AEX: 2015/06/25 - 2020/12/31. The numbers in parenthesis denote relative QLIKE and MSFE values with benchmark GARCH(1,1). Entries < 1 indicate outperformance of the benchmark. Numbers in bold are QLIKE- and MSFE-best. A star * indicates significant improvement of the best model over the benchmark GARCH(1,1) at 5% level.

Table 10: Out-of-sample volatility forecasting results with squared returns - $\mu \neq 0$.

	QLIKE				MSFE			
	SP500		AEX		SP500		AEX	
Benchmark								
GARCH(1,1)	0.637	(1.000)	0.935	(1.000)	30.084	(1.000)	20.675	(1.000)
A-GARCH(1,1)	0.640	(1.005)	0.898	(0.960)	***29.010***	(0.964)	20.623	(0.997)
QML General SV								
$m = 1$	1.320	(2.073)	*0.989*	*(1.058)*	43.777	(1.455)	*21.413*	*(1.035)*
2	1.324	(2.078)	0.989	(1.058)	43.849	(1.458)	21.415	(1.036)
3	1.324	(2.078)	0.991	(1.059)	43.848	(1.458)	21.451	(1.038)
4	1.323	(2.077)	0.991	(1.060)	43.844	(1.457)	21.456	(1.038)
5	*1.317*	*(2.067)*	0.992	(1.061)	*43.769*	*(1.455)*	21.463	(1.038)
6	1.398	(2.194)	0.991	(1.060)	44.249	(1.471)	21.460	(1.038)
7	1.394	(2.189)	0.991	(1.060)	44.225	(1.470)	21.456	(1.038)
8	1.404	(2.204)	0.991	(1.060)	44.319	(1.473)	21.455	(1.038)
9	1.406	(2.207)	1.003	(1.073)	44.408	(1.476)	21.537	(1.042)
10	1.405	(2.205)	1.034	(1.106)	44.399	(1.476)	21.388	(1.035)
SML General SV								
$m = 0$	***0.582****	*(0.914)*	0.887	(0.949)	31.851	(1.059)	19.864	(0.961)
1	0.600	(0.942)	0.883	(0.944)	31.159	(1.036)	***19.691****	*(0.952)*
2	0.597	(0.937)	***0.882****	*(0.943)*	31.257	(1.039)	19.790	(0.957)
3	0.604	(0.948)	*0.882*	(0.943)	30.555	*(1.016)*	19.942	(0.965)
4	0.597	(0.937)	0.882	(0.943)	30.855	(1.026)	19.889	(0.962)
5	0.598	(0.939)	0.885	(0.947)	31.086	(1.033)	20.000	(0.967)
BF Non - Causal SV								
2-dim	0.632	(0.992)	0.897	(0.959)	*31.630*	*(1.051)*	20.009	(0.968)
4-dim	*0.589*	*(0.925)*	*0.893*	*(0.955)*	32.517	(1.081)	*19.932*	*(0.964)*

Note: The entries are the average Quasi-likelihood (QLIKE) and Mean Squared Forecast Error (MSFE) values computed using one-day ahead volatility predictions over the out-of-sample period. For SP500: 2015/06/05 - 2020/12/31 and AEX: 2015/06/25 - 2020/12/31. The numbers in parenthesis denote relative QLIKE and MSFE values with benchmark GARCH(1,1). Entries < 1 indicate outperformance of the benchmark. Numbers in bold are QLIKE- and MSFE- best. A star * indicates significant improvement of the best model over the benchmark GARCH(1,1) at 5% level.

7 Conclusion

In this paper, we propose a novel stochastic volatility model which is based on the stochastic volatility model of Catania (2020). The major innovation of this new model lies in the inclusion of a leading volatility shock instead of only considering lagging volatility shocks in the general leverage structure. The aim is to find out how we can feasibly estimate the new SV model and use it for improving the volatility predictions on the S&P 500 and AEX index return series. Three difficult, yet elegant filtering methods are applied in conjunction with maximum likelihood to realize this goal. Lastly, we extend the SV model of Catania (2020) by adding an additional mean term to correct for the bias in capturing the leverage effect.

From the simulation and empirical study, we conclude with the following. The new SV model with non-causal leverage indeed significantly outperforms the SV model of Catania (2020). However, this only holds when we use the 5-min. intra-day realized variance proxy for the true volatility. If we use the noisy squared returns, the novel SV models fails to beat the SV model of Catania (2020). On the other hand, we find that the leverage effect can be properly accounted for when adding an additional mean term in the non-causal SV model. For the SV model of Catania (2020) the effect of adding a mean term is less pronounced. The question whether the contemporaneous or the intertemporal SV leverage model gives a better full sample fit can now be answered. We can argue that for both the S&P 500 and AEX indices a SV model with leverage of at least one lag (intertemporal) suits the data better according to BIC.

In this research, we could not investigate more due to time constraints. However, there is room for future research. A good first idea would be to consider an alternative filtering method to estimate the SV model of Catania (2020) given the fact that the particle filter is computationally very slow. Perhaps applying the Bellman filter of Lange (2021) would be worth the effort to investigate in the future. Second neat extension would be to estimate a bi-variate SV leverage model using a Bayesian MCMC approach. This requires the researcher to find a strong prior for the correlation matrix of the return shock and volatility shock vector. Kirby et al. (2006) is recommended in this case.

8 References

Ait-Sahalia, Y., Fan, J., Li, Y. (2013). The leverage effect puzzle: Disentangling sources of bias at high frequency. Journal of Financial Economics, 109(1), 224-249.

Anderson, B. D., Moore, J. B. (2012). Optimal filtering. Courier Corporation.

Asai, M., McAleer, M. (2011). Alternative asymmetric stochastic volatility models. Econometric Reviews, 30(5), 548-564.

Bellman, R. (1956). Dynamic programming and Lagrange multipliers. Proceedings of the National Academy of Sciences of the United States of America, 42(10), 767.

Black, F., 1976, "Studies of Stock Price Volatility Changes", Proceedings of the Business and Economics Section of the American Statistical Association , 177–181.

Broto, C., Ruiz, E. (2004). Estimation methods for stochastic volatility models: a survey. Journal of Economic Surveys, 18(5), 613-649.

Catania, L. (2020). A Stochastic Volatility Model with a General Leverage Specification. Journal of Business Economic Statistics, 1-23.

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, J. (1998). Multivariate stochastic volatility models: estimation and a comparison with VGARCH models. Journal of Empirical Finance, 5(2), 155-173.

Diebold, F. X., Mariano, R. S. (1995). Com paring predictive accu racy. Journal of Business and Economic Statistics, 13(3), 253-263.

Durbin, J., Koopman, S. J. (1997). Monte Carlo maximum likelihood estimation for non-Gaussian state space models. Biometrika, 84(3), 669-684.

Gordon, N. J., Salmond, D. J., Smith, A. F. (1993, April). Novel approach to nonlinear/non-Gaussian Bayesian state estimation. In IEE proceedings F (radar and signal processing)

(Vol. 140, No. 2, pp. 107-113). IET Digital Library.

Harvey, A. C., Shephard, N. (1996). Estimation of an asymmetric stochastic volatility model for asset returns. Journal of Business Economic Statistics, 14(4), 429-434.

Harvey, D., Leybourne, S., Newbold, P. (1997). Testing the equality of prediction mean squared errors. International Journal of forecasting, 13(2), 281-291.

Jacquier, E., Polson, N. G., Rossi, P. E. (2004). Bayesian analysis of stochastic volatility models with fat-tails and correlated errors. Journal of Econometrics, 122(1), 185-212.

Kalman, R. E. (1960). A new approach to linear filtering and prediction problems.

Koopman, S. J., Lucas, A., Scharth, M. (2016). Predicting time-varying parameters with parameter-driven and observation-driven models. Review of Economics and Statistics, 98(1), 97-110.

Lange, R. J. (2020). Bellman filtering for state-space models.

Malik, S., Pitt, M. K. (2011). Particle filters for continuous likelihood evaluation and maximisation. Journal of Econometrics, 165(2), 190-209.

Richard, J. F., Zhang, W. (2007). Efficient high-dimensional importance sampling. Journal of Econometrics, 141(2), 1385-1411.

Scharth, M., Kohn, R. (2016). Particle efficient importance sampling. Journal of Econometrics, 190(1), 133-147.

Shepard, N., Pitt, M. K. (1999). Filtering via simulation: auxiliary particle filter. Journal of the American Statistical Association, 94, 590-599.

Taylor, S. J. (1994). Modeling stochastic volatility: A review and comparative study. Mathematical finance, 4(2), 183-204.

Yu, J. (2005). On leverage in a stochastic volatility model. Journal of Econometrics, 127(2), 165-178.

9 Appendix

A: Continuous resampling of algorithm CSIR

The continuous resampling step of algorithm: CSIR requires a set of ordered uniforms: $u_1 < u_2 < \cdots < u_N$. We follow the advice to generate stratified uniforms to prevent sample impoverishment as suggested by Malik and Pitt (2011). That is, we simply generate a single random number $u \sim Unif(0,1)$ and use that to get stratified and sorted uniforms with $u_j = (j-1)/N + u/N$ for $j = 1, \ldots, N$. This is done for each time t during filtering implying the need for T fixed random uniform numbers prior to the application of the resampling steps of the continuous particle filter.

Recall that we had to resample from the continuous mixture density function

$$\tilde{F}_N(h_t) = \sum_{k=0}^{N} \lambda_t^k \, G_K \left(\frac{h_t - h_t^{(k)}}{h_t^{(k+1)} - h_t^{(k)}} \right),$$

where $\lambda_t^k = \pi_t^1/2$ if $k = 0$, $\lambda_t^k = \pi_t^N/2$ if $k = N$ and $\lambda_t^k = \pi_t^{k+1} + \pi_t^k)/2$ otherwise. The pseudo-code for resampling from the continuous function $\tilde{F}_N$ is given below:

Algorithm Continuous resampling pseudo-code

Input: particles: $H_t = \{h_t^k\}_{k=1}^N$, normalized weights: $\Pi_t = \{\pi_t^k\}_{k=1}^N$, uniform: u
Output: Continuously resampled particles: $\tilde{H}_t = \{\tilde{h}_t^k\}_{k=1}^N$

Construct uniforms: $u_1 < u_2 < \cdots < u_N \leftarrow \text{Stratify}(u)$
Sort H_t in ascending order and arrange Π_t accordingly
Set $s = 0$, $j = 1$
for $k = 0$ to N **do**
    Compute λ_t^k
    $s = s + \lambda_t^k$
    while $u_j \leq s$ AND $j \leq N$ **do**
        $r^j = i$
        $u_j^* = (u_j - (s - \lambda_t^k))/\lambda_t^k$
        $j = j + 1$
        $\tilde{h}_t^k = h_t^{(1)}$ if $r^j = 0$,
        $\tilde{h}_t^k = h_t^{(N)}$ if $r^j = N$
        $\tilde{h}_t^k = \left(h_t^{(r^j+1)} - h_t^{(r^j)}\right) u_j^* + h_t^{(r^j)}$ if $r^j = 1, \ldots, N-1$
    end while
end for

B: Derivations score and information matrix for BF

For the optimization step and evaluating the (negative) BF-implied log-likelihood, an analytical expression of the gradient and negative Hessian of $\ell(\, y_t - \mu \,|\, \boldsymbol{\alpha}_t \,) = \log p(\, y_t - \mu \,|\, \boldsymbol{\alpha}_t)$ is necessary. Recall we constructed a linear state transition equation with 4-dimensional state vector $\boldsymbol{\alpha}_t = (h_t, \eta_{t+1}, \eta_t, \eta_{t-1})'$. We assumed that $(y_t - \mu) \,|\, \boldsymbol{\alpha}_t$ is normally distributed with conditional mean and variance:

$$\mathrm{E}\,(\, y_t - \mu \,|\, \boldsymbol{\alpha}_t \,) \;=\; \exp\left\{\frac{h_t}{2}\right\}\left(\sum_{i\,=-1}^{+1} \rho_i\, \eta_{t+i}\right) \;\equiv\; m_t,$$

$$\mathrm{Var}\,(\, y_t - \mu \,|\, \boldsymbol{\alpha}_t \,) \;=\; \exp\{h_t\}\left(1 - \sum_{i=-1}^{+1} \rho_i^2\right) \;\equiv\; s_t^2.$$

The score vector (gradient) and negative information matrix (negative Hessian) of log $p(\, y_t - \mu \,|\, \boldsymbol{\alpha}_t)$ are shown below for the 4-dimensional state vector.[4]

Score vector :

$$\begin{bmatrix} \dfrac{\mathrm{d}\,\ell(\, y_t \,|\, \boldsymbol{\alpha}_t \,)}{\mathrm{d}\, h_t} \\[2ex] \dfrac{\mathrm{d}\,\ell(\, y_t \,|\, \boldsymbol{\alpha}_t \,)}{\mathrm{d}\, \eta_{t+1}} \\[2ex] \dfrac{\mathrm{d}\,\ell(\, y_t \,|\, \boldsymbol{\alpha}_t \,)}{\mathrm{d}\, \eta_t} \\[2ex] \dfrac{\mathrm{d}\,\ell(\, y_t \,|\, \boldsymbol{\alpha}_t \,)}{\mathrm{d}\, \eta_{t-1}} \end{bmatrix} = \begin{bmatrix} \dfrac{y_t(y_t - m_t)}{2\, s_t^2} - \dfrac{1}{2} \\[2ex] \dfrac{\rho_1}{\sqrt{1 - (\rho_{-1}^2 + \rho_0^2 + \rho_1^2)}}\left(\dfrac{y_t - m_t}{s_t}\right) \\[2ex] \dfrac{\rho_0}{\sqrt{1 - (\rho_{-1}^2 + \rho_0^2 + \rho_1^2)}}\left(\dfrac{y_t - m_t}{s_t}\right) \\[2ex] \dfrac{\rho_{-1}}{\sqrt{1 - (\rho_{-1}^2 + \rho_0^2 + \rho_1^2)}}\left(\dfrac{y_t - m_t}{s_t}\right) \end{bmatrix}$$

[4]For the 2-dimensional SV model with non-causal leverage, the scores and information matrix can be computed analogously. The 2-dimensional state vector becomes in that case: $\boldsymbol{\alpha}_t = (h_t, \eta_{t+1})'$ with imaginary correlation parameters $\rho_{-1} = \rho_0 = 0$.

Realized Information matrix :

$$
\begin{bmatrix}
-\dfrac{\mathrm{d}^2\ell(\boldsymbol{y}_t|\,\boldsymbol{\alpha})}{\mathrm{d}\,h_t^2} & & & \\
-\dfrac{\mathrm{d}^2\ell(\boldsymbol{y}_t|\,\boldsymbol{\alpha})}{\mathrm{d}h_t\,\mathrm{d}\eta_{t+1}} & -\dfrac{\mathrm{d}^2\ell(\boldsymbol{y}_t|\,\boldsymbol{\alpha})}{\mathrm{d}\,\eta_{t+1}^2} & & \\
-\dfrac{\mathrm{d}^2\ell(\boldsymbol{y}_t|\,\boldsymbol{\alpha})}{\mathrm{d}h_t\,\mathrm{d}\eta_t} & -\dfrac{\mathrm{d}^2\ell(\boldsymbol{y}_t|\,\boldsymbol{\alpha})}{\mathrm{d}\eta_{t+1}\,\mathrm{d}\eta_t} & -\dfrac{\mathrm{d}^2\ell(\boldsymbol{y}_t|\,\boldsymbol{\alpha})}{\mathrm{d}\,\eta_t^2} & \\
-\dfrac{\mathrm{d}^2\ell(\boldsymbol{y}_t|\,\boldsymbol{\alpha})}{\mathrm{d}h_t\,\mathrm{d}\eta_{t-1}} & -\dfrac{\mathrm{d}^2\ell(\boldsymbol{y}_t|\,\boldsymbol{\alpha})}{\mathrm{d}\eta_{t+1}\,\mathrm{d}\eta_{t-1}} & -\dfrac{\mathrm{d}^2\ell(\boldsymbol{y}_t|\,\boldsymbol{\alpha})}{\mathrm{d}\eta_t\,\mathrm{d}\eta_{t-1}} & -\dfrac{\mathrm{d}^2\ell(\boldsymbol{y}_t|\,\boldsymbol{\alpha})}{\mathrm{d}\,\eta_{t-1}^2}
\end{bmatrix}
$$

$$
\begin{aligned}
-\frac{\mathrm{d}^2\ell(\boldsymbol{y}_t|\,\boldsymbol{\alpha}_t)}{\mathrm{d}\,h_t^2} &= \frac{y_t^2}{2\,s_t^2} - \frac{y_t\,m_t}{4\,s_t^2} \\
-\frac{\mathrm{d}^2\ell(\boldsymbol{y}_t|\,\boldsymbol{\alpha}_t)}{\mathrm{d}h_t\,\mathrm{d}\eta_{t+i}} &= \frac{\rho_i}{\sqrt{1-(\,\rho_1^2+\rho_0^2+\rho_{-1}^2\,)}}\left(\frac{y_t}{2\,s_t}\right) \quad \text{for} \quad i=-1,0,1 \\
-\frac{\mathrm{d}^2\ell(\boldsymbol{y}_t|\,\boldsymbol{\alpha}_t)}{\mathrm{d}\,\eta_{t+i}^2} &= \frac{\rho_i^2}{1-(\,\rho_1^2+\rho_0^2+\rho_{-1}^2\,)} \quad \text{for} \quad i=-1,0,1 \\
-\frac{\mathrm{d}^2\ell(\boldsymbol{y}_t|\,\boldsymbol{\alpha}_t)}{\mathrm{d}\eta_{t+1}\,\mathrm{d}\eta_t} &= \frac{\rho_1\,\rho_0}{1-(\,\rho_1^2+\rho_0^2+\rho_{-1}^2\,)} \\
-\frac{\mathrm{d}^2\ell(\boldsymbol{y}_t|\,\boldsymbol{\alpha}_t)}{\mathrm{d}\eta_{t+1}\,\mathrm{d}\eta_{t-1}} &= \frac{\rho_1\,\rho_{-1}}{1-(\,\rho_1^2+\rho_0^2+\rho_{-1}^2\,)} \\
-\frac{\mathrm{d}^2\ell(\boldsymbol{y}_t|\,\boldsymbol{\alpha}_t)}{\mathrm{d}\eta_t\,\mathrm{d}\eta_{t-1}} &= \frac{\rho_0\,\rho_{-1}}{1-(\,\rho_1^2+\rho_0^2+\rho_{-1}^2\,)}
\end{aligned}
$$

Expected Information matrix :

$$\begin{bmatrix} -\mathrm{E}\left(\frac{\mathrm{d}^2\ell(\boldsymbol{y}_t|\,\boldsymbol{\alpha}_t)}{\mathrm{d}\,h_t^2}\,|\,\boldsymbol{\alpha}_t\right) & & & \\ -\mathrm{E}\left(\frac{\mathrm{d}^2\ell(\boldsymbol{y}_t|\,\boldsymbol{\alpha}_t)}{\mathrm{d}h_t\,\mathrm{d}\eta_{t+1}}\,|\,\boldsymbol{\alpha}_t\right) & -\mathrm{E}\left(\frac{\mathrm{d}^2\ell(\boldsymbol{y}_t|\,\boldsymbol{\alpha}_t)}{\mathrm{d}\,\eta_{t+1}^2}\,|\,\boldsymbol{\alpha}_t\right) & & \\ -\mathrm{E}\left(\frac{\mathrm{d}^2\ell(\boldsymbol{y}_t|\,\boldsymbol{\alpha}_t)}{\mathrm{d}h_t\,\mathrm{d}\eta_t}\,|\,\boldsymbol{\alpha}_t\right) & -\mathrm{E}\left(\frac{\mathrm{d}^2\ell(\boldsymbol{y}_t|\,\boldsymbol{\alpha}_t)}{\mathrm{d}\eta_{t+i}\,\mathrm{d}\eta_t}\,|\,\boldsymbol{\alpha}_t\right) & -\mathrm{E}\left(\frac{\mathrm{d}^2\ell(\boldsymbol{y}_t|\,\boldsymbol{\alpha}_t)}{\mathrm{d}\,\eta_t^2}\,|\,\boldsymbol{\alpha}_t\right) & \\ -\mathrm{E}\left(\frac{\mathrm{d}^2\ell(\boldsymbol{y}_t|\,\boldsymbol{\alpha}_t)}{\mathrm{d}h_t\,\mathrm{d}\eta_{t-1}}\,|\,\boldsymbol{\alpha}_t\right) & -\mathrm{E}\left(\frac{\mathrm{d}^2\ell(\boldsymbol{y}_t|\,\boldsymbol{\alpha}_t)}{\mathrm{d}\eta_{t+1}\,\mathrm{d}\eta_{t-1}}\,|\,\boldsymbol{\alpha}_t\right) & -\mathrm{E}\left(\frac{\mathrm{d}^2\ell(\boldsymbol{y}_t|\,\boldsymbol{\alpha}_t)}{\mathrm{d}\eta_t\,\mathrm{d}\eta_{t-1}}\,|\,\boldsymbol{\alpha}_t\right) & -\mathrm{E}\left(\frac{\mathrm{d}^2\ell(\boldsymbol{y}_t|\,\boldsymbol{\alpha}_t)}{\mathrm{d}\,\eta_{t-1}^2}\,|\,\boldsymbol{\alpha}_t\right) \end{bmatrix}$$

$$-\mathrm{E}\left(\frac{\mathrm{d}^2\ell(\boldsymbol{y}_t|\,\boldsymbol{\alpha}_t)}{\mathrm{d}\,h_t^2}\,|\,\boldsymbol{\alpha}_t\right) = \frac{1}{2}+\frac{1}{4}\left(\frac{m_t^2}{s_t^2}\right)$$

$$-\mathrm{E}\left(\frac{\mathrm{d}^2\ell(\boldsymbol{y}_t|\,\boldsymbol{\alpha}_t)}{\mathrm{d}h_t\,\mathrm{d}\eta_{t+i}}\,|\,\boldsymbol{\alpha}_t\right) = \frac{\rho_i}{\sqrt{1-(\,\rho_1^2+\rho_0^2+\rho_{-1}^2\,)}}\left(\frac{m_t}{2\,s_t}\right) \quad \text{for} \quad i=-1,0,1$$

$$-\mathrm{E}\left(\frac{\mathrm{d}^2\ell(\boldsymbol{y}_t|\,\boldsymbol{\alpha}_t)}{\mathrm{d}\,\eta_{t+i}^2}\,|\,\boldsymbol{\alpha}_t\right) = \frac{\rho_i^2}{1-(\,\rho_1^2+\rho_0^2+\rho_{-1}^2\,)} \quad \text{for} \quad i=-1,0,1$$

$$-\mathrm{E}\left(\frac{\mathrm{d}^2\ell(\boldsymbol{y}_t|\,\boldsymbol{\alpha}_t)}{\mathrm{d}\eta_{t+1}\,\mathrm{d}\eta_t}\,|\,\boldsymbol{\alpha}_t\right) = \frac{\rho_1\,\rho_0}{1-(\,\rho_1^2+\rho_0^2+\rho_{-1}^2\,)}$$

$$-\mathrm{E}\left(\frac{\mathrm{d}^2\ell(\boldsymbol{y}_t|\,\boldsymbol{\alpha}_t)}{\mathrm{d}\eta_{t+1}\,\mathrm{d}\eta_{t-1}}\,|\,\boldsymbol{\alpha}_t\right) = \frac{\rho_1\,\rho_{-1}}{1-(\,\rho_1^2+\rho_0^2+\rho_{-1}^2\,)}$$

$$-\mathrm{E}\left(\frac{\mathrm{d}^2\ell(\boldsymbol{y}_t|\,\boldsymbol{\alpha}_t)}{\mathrm{d}\eta_t\,\mathrm{d}\eta_{t-1}}\,|\,\boldsymbol{\alpha}_t\right) = \frac{\rho_0\,\rho_{-1}}{1-(\,\rho_1^2+\rho_0^2+\rho_{-1}^2\,)}$$