ERASMUS UNIVERSITY ROTTERDAM

Erasmus School of Economics

Master Thesis Quantitative Finance

Modelling the leverage effect with stochastic volatility models with an application for volatility forecasting

August 1, 2022

Author	Max Stolwerk	(457523)
Supervisor	Dr. R-J. Lange	
Second Assessor	Prof. Dr. D.J.C. van Dijk	

Abstract

In this paper we investigate if adding an extra relation between the volatility shocks and return shocks in SV models improves the models ability to forecast volatility. Squared returns and realised variance are used as a proxy for the “true” variance. The added leverage effect is the correlation between the current volatility shock and the future return shock. We show that on US index data this effect is generally estimated to be small and negative and including it into a SV model improves its volatility forecasts. We show that the volatility feedback effect and inter-temporal leverage effects are also present in the data and should also be included in modelling financial log returns.

Contents

1 Introduction

Modelling the dependence in a time series is at the core of econometrics in finance but this is not only limited to finance as many other fields also use time series analysis. While the increases in data availability and computational power open up opportunities for more accurate models and methods one problem remains persistent, that of noisy data. Extracting information about a signal from this noise-contaminated data remains a difficult task. The workhorse in finance have long been univariate ARIMA models, these simple models have shown surprisingly good results however they are limited by their assumptions. The main two points of criticism have been the requirement of the observations to be linear functions of past observations and the inability to have stable and persistent components.

A way to deal with these limitations and noise problems has been the introduction of state space modelling. In state space models it is assumed the "true" state of an observation is unknown as it is contaminated with noise and the distribution of the observations depends on the underlying "true" state. As the distribution of the observations now relies on an underlying stochastic process it can now vary over time. It is also not required that this relationship is linear. The hardest part of state space models is determining the value of this unobservable latent state. This filtering of the observed data has been an interesting topic as exact filtering methods are not always available. The exception is the Kalman filter, however this filter is only exact for linear Gaussian state space models. So lacking an exact filtering solution for non-Gaussian state space models we need to turn to approximate solutions.

State space models are used to try and solve the "leverage puzzle", the notion that financial returns and volatility shocks are correlated is at the base of this conundrum. The common thought in past literature is that an increase in volatility follows the day after a negative return, however recent studies using state space models have shown an indication that this may not be true. Intuitively arguments can be made that a negative return actually follows a volatility shock. Using state space modelling to test if this idea holds could change how we view the causality and timing of volatility and returns and could provide a new insight into past research.

This research follows the paper of Catania (2022) closely as he has developed a stochastic volatility (SV) model that incorporates a leverage effect and is nicely suited the explore the leverage puzzle. In a simulation study estimating several versions of this model we find that the newly developed Bellman filter of Lange (2020), which relies on Bellman's dynamic programming principle, outperforms the continuous resampling importance resampling (CSIR) method of Malik and Pitt (2011) originally used in Catania (2022) from both an accuracy and a computational point of view. The Bellman filter and CSIR method are then used to estimate the Catania model with different lags on data obtained from two major US stock indices, the Standard & Poor 500 (S&P 500) and the National Association of Securities Dealers Automated Quotations (Nasdaq) Composite with data ranging from the 3^{rd} of January to the 31^{st} of December.

To evaluate the predictive power of the models we compare their predicted volatility to two proxies, squared returns and realised variance. We find that the original model estimated with the CSIR method outperforms the one estimated with the Bellman filter from a predictive volatility point of view. Even though the simulation study finds that the Bellman filter delivers more accurate parameter estimates.

The Bellman filter is also used to estimate an extended version of the model that includes a correlation between the future return shock and the current volatility shock. The models estimated with this extra leverage effect provide better volatility forecasts than those without it. The best performing models find a negative but small estimate for this extra leverage effect and also find large negative values for the contemporaneous leverage effect. Different inter-temporal leverage effects are also found to have significant estimates. This indicates the presence of the volatility feedback effect and leverage effect in the log returns of US indices.

2 State space models

The focus of this research is on nonlinear non-Gaussian state space models where a latent state $(\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_t)$ affects the distribution of the observations $(y_1, \ldots, y_t)$. For convenience the following general form data-generating process is considered:

$$\begin{aligned} \text{observation equation}: \quad & y_t = g^{\boldsymbol{\theta}}\left(\boldsymbol{\alpha}_t, \boldsymbol{\varepsilon}_t\right), & \boldsymbol{\varepsilon}_t \sim p^{\boldsymbol{\theta}}\left(\boldsymbol{\varepsilon}_t\right), \\ \text{state equation}: \quad & \boldsymbol{\alpha}_t = h^{\boldsymbol{\theta}}\left(\boldsymbol{\alpha}_{t-1}, \boldsymbol{\eta}_t\right), & \boldsymbol{\eta}_t \sim p^{\boldsymbol{\theta}}\left(\boldsymbol{\eta}_t\right). \end{aligned} \tag{1}$$

The observation vector $y_t \in Y$ has dimensions (1×1) and the latent state vector $\boldsymbol{\alpha}_t \in A$ has dimensions $(m \times 1)$ and functions $g_{\boldsymbol{\theta}}$ and $h_{\boldsymbol{\theta}}$ are functions mapping into Y and A. $\boldsymbol{\varepsilon}_t$ and $\boldsymbol{\eta}_t$ are i.i.d disturbances with corresponding distributions $p^{\boldsymbol{\theta}}\left(\boldsymbol{\varepsilon}_t\right)$ and $p^{\boldsymbol{\theta}}\left(\boldsymbol{\eta}_t\right)$. All functions depend on the constant parameter $\boldsymbol{\theta}$.

From the definition in (1) the implied conditional distributions can be written as

$$y_t \sim p^{\boldsymbol{\theta}}\left(y_t \middle| \boldsymbol{\alpha}_t\right) \qquad \boldsymbol{\alpha}_t \sim p^{\boldsymbol{\theta}}\left(\boldsymbol{\alpha}_t \middle| \boldsymbol{\alpha}_{t-1}\right). \tag{2}$$

The difficulty lies in filtering the latent true states $\boldsymbol{\alpha}$ out of the observations $\boldsymbol{y}$. As these states are unknown they cannot be used in the log-likelihood function. The estimated states are defined as $(\boldsymbol{a}_1, \ldots, \boldsymbol{a}_t)$. Finding the states that maximise the log-likelihood is the common method of estimating the most likely states. However there are two options, computing the mean or the mode. If the choice would be made from a loss function point of view the mean would be optimal, but generally speaking computing the mean would be computationally infeasible (an exception would be for linear Gaussian models using the Kalman (1960) filter). In state space literature it is common to calculate the posterior mode instead of the mean caused by these computational problems. This posterior mode estimation often called *the maximum a posteriori estimate* might be sub-optimal from a loss function perspective but does not outweigh the benefits of having to solve an optimisation instead of a high dimensional integral. The posterior mode is the most likely value of the states given the data and can thus be found by maximising the log-likelihood conditional on the data. By deconstructing the conditional likelihood it can be seen that the posterior mode is also the maximiser of the joint log-likelihood:

$$\ell(\boldsymbol{a}_1, \ldots, \boldsymbol{a}_t | y_1, \ldots, y_t) = \ell(\boldsymbol{a}_1, \ldots, \boldsymbol{a}_t, y_1, \ldots, y_t) - \ell(y_1, \ldots, y_t), \tag{3}$$

where $\ell()$ is the log-likelihood function. The second term does not depend on the states and is known at the moment of estimation so can be left out of the maximisation problem. The posterior mode will then be

$$\left(\widehat{\boldsymbol{a}}_{1|t}, \ldots, \widehat{\boldsymbol{a}}_{t|t}\right) := \underset{(\boldsymbol{a}_1, \ldots, \boldsymbol{a}_t)}{\arg\max} \quad \ell\left(\boldsymbol{a}_1, \ldots, \boldsymbol{a}_t, y_1, \ldots, y_t\right). \tag{4}$$

Now for model (1) and a prior distribution $p(\boldsymbol{a}_1)$ the joint log-likelihood can be written as

$$\ell\left(\boldsymbol{a}_1, \ldots, \boldsymbol{a}_t, y_1, \ldots, y_t\right) = \ell\left(\boldsymbol{a}_1\right) + \ell\left(y_1 \mid \boldsymbol{a}_1\right) + \sum_{i=2}^{t} \left[\ell\left(y_i \mid \boldsymbol{a}_i\right) + \ell\left(\boldsymbol{a}_i \mid \boldsymbol{a}_{i-1}\right)\right]. \tag{5}$$

A method to find an exact solution to the optimisation problem in (4) does not exist for most state space models, e.g. nonlinear non-Gaussian state space models. Several approximate methods have been developed to tackle the filtering and parameter estimation problem of these state space models. The accuracy and computational times of these approximate methods heavily depend on the type of model used. In this article we will consider three types of stochastic volatility models to examine the leverage effect present in stock returns.

3 The leverage puzzle and stochastic volatility

The relation between negative returns and an increase in volatility in financial data has been the subject of many research papers and debate in the economic/econometric world ever since Black (1976) first described it. Black and other earlier literature attributed this asymmetric return–volatility relationship to the changing leverage of a company and hence coined this effect the leverage effect. The general idea is that a reduction in price decreases the equity value of a firm and as the debt of this company remains the same the companies leverage ratio increases. This will increase the perceived risk and thus increase the variance of the returns.

While this explanation seems theoretically sound from an economical point of view, later research shows that this is not the sole explanation. On top of that the exact causality and timing has also been subject to debate. It used to be assumed that a fall in price was accompanied by a delayed volatility spike as this was in line with the leverage explanation. Catania (2022) find evidence that the effect may rather be simultaneous than delayed. If negative returns and increases in volatility occur simultaneously this would go against previous literature and may lead to very different conclusions and notions about leverage that are currently accepted as the truth. Teulings, Lange, and van der Kroft (2020) even find evidence of a reverse causality where expected future volatility shocks cause the price to drop today.

3.1 Stochastic volatility

The need for stochastic volatility models arose when econometricians needed to model the dynamic evolution of volatility over time. Adding a stochastic process for the volatility instead of a time invariant one adds flexibility to a model that was needed to model financial time series. In SV models the volatility is an unobservable component following some latent stochastic process (Ghysels, Harvey, and Renault (1996), Taylor (1994)). SV models are a wide array of models that contain an observation equation that is not necessarily linear in the volatility process. Many can be written as a version of this nonlinear state space model without correlation:

$$
\begin{aligned}
y_t &= \mu + \sigma_t \varepsilon_t, \qquad \varepsilon_t \overset{\text{iid}}{\sim} N(0,1), \\
\sigma_t &= \exp\left(\frac{h_t}{2}\right), \\
h_t &= c + \varphi h_{t-1} + \kappa \eta_t, \qquad \eta_t \overset{\text{iid}}{\sim} N(0,1),
\end{aligned} \tag{6}
$$

where y_t is a financial log return with mean μ and intercept c to control the level of the log-volatility equation, φ as the persistence of the log-volatility process and κ as the variance of the volatility shock.

When a relation between the shocks is included into SV models some statistical characteristics are also introduced into the model, for example with the asymmetric SV model of Jacquier, Polson, and Rossi (2004) which allows for a contemporaneous relation would lead to returns being skewed and serial correlated. Serial correlation is against the efficient market hypothesis which states that market prices should only react to new information and the price now reflects all information available. Assuming that asset returns follow a SV model with a contemporaneous relation one would have to admit markets do not adhere to the efficient market hypothesis, but from a intuitive standpoint with the ever increasing speed of information processing it is making more sense that the volatility now would react now to return shocks and not only have a delayed reaction. If one takes an asset pricing point of view and views volatility as a measure of risk and that the price of an asset is dependant on its risk, then an expected return unequal to zero makes sense. To avoid this issue researchers can also choose to only include an inter-temporal lag as done by Harvey and Shephard (1996).

Another drawback of SV models is the complicated estimation procedure as the likelihood is usually not easy to evaluate. There are some exceptions but in many cases there is no closed-form derivation possible. These characteristics of many SV models limit the empirical application in the financial literature. But luckily the growing literature continuously develops new methods and improvements on existing methods for estimating or simulating the likelihood which alleviates the estimation problem. We are now able to estimate parameters based on maximum likelihood for the quasi- or simulated likelihood functions more accurate than ever before. With this improved accuracy past research can be redone and potentially yield different results or SV models can now be considered for more empirical applications. This could cause the financial world to realise they may have underestimated or even wrongly assumed the effects or causality of variables. An example of this can be the leverage effect mentioned in the section above which was investigated in Catania (2022) with a newly developed model that can be found in the next Section.

3.2 SV model of Catania

Catania (2022) introduces a SV model that allows for correlation of return and volatility shocks at different temporal lags instead of only one. This model encompasses many familiar SV models such as the simple model in equation (6), the contemporaneous specification of Jacquier et al. (2004) and the inter-temporal specification of Harvey and Shephard (1996) that can retrieved by setting parameter restrictions and varying the amount of lags allowed. They assume that the log returns y_t are generated from the following SV model

$$\begin{aligned} y_t &= \mu + \sigma_t \varepsilon_t, \qquad \varepsilon_t \overset{\text{iid}}{\sim} N(0,1), \\ \sigma_t &= \exp\left(\frac{h_t}{2}\right), \\ h_t &= c + \varphi h_{t-1} + \kappa \eta_t, \\ \eta_t &= \sum_{j=0}^{m} \rho_j \varepsilon_{t-j} + \sqrt{1 - \sum_{j=0}^{m} \rho_j^2} * b_t, \quad b_t \overset{\text{iid}}{\sim} N(0,1) \end{aligned} \tag{7}$$

where to control the variance and persistence of the volatility process, $\kappa > 0$ and $|\varphi| < 1$. For identification is required that $\sum_{j=0}^{m} \rho_j < 1$. In the original Catania model σ_t is only multiplied with a constant parameter β to control for the level but this is replaced with the intercept c in the log-volatility equation. Both options control the level and using either is a matter of (mathematical) preference.

The volatility shock η_t is a linear function of the current and lagged return shocks up till m periods of time ago. Because of this the log-volatility at time t and the return shock at the same time will be dependent when $\rho_0 \neq 0$. This will not be a problem for modelling financial returns as past research such as Carr and Wu (2017) have shown that $\rho_0 < 0$ is typically found in financial returns, known as the "volatility-feedback effect" and causes the distribution of y_t to be negatively skewed. A consequence of this will be that μ in equation (7) is the median instead of the usual mean. Catania (2022) sets $\mu = 0$, however this limits the estimation of ρ_0 as it can be more accurate if μ is estimated. So in this paper μ will also be estimated.

Catania (2022) estimates the parameters with two different estimation methods, a simulated log-likelihood estimation using the CSIR method of Malik and Pitt (2011) and with a quasi log-likelihood estimation (with the assumption $\rho_0 = 0$) using the log of the squared returns and then running the Kalman filter to obtain a likelihood. They find that from a likelihood perspective it is worth to estimate the contemporaneous correlation of the returns and log-volatility (ρ_0) and an increase of m increases the precision of the volatility predictions.

Catania's model implies the unconditional distribution of the log-volatility as

$$h \sim N\left(\frac{c}{1-\varphi}, \frac{\kappa^2}{1-\varphi^2}\left[1 + 2\sum_{s=1}^{m}\varphi^s\sum_{j=s}^{m}\rho_j\rho_{j-s}\right]\right), \tag{8}$$

causing the expected value of tomorrows return to be

$$\begin{aligned} E\left(y_{t+1}\right) &= E\left(\mu + \exp\left(\frac{h_{t+1}}{2}\right)\varepsilon_{t+1}\right) \\ &= \mu + \frac{1}{2}\kappa\rho_0 \exp\left(\frac{c}{2(1-\varphi)} + \frac{\kappa^2}{8\left(1-\varphi^2\right)}\left[1 + 2\sum_{s=1}^{m}\varphi^s\sum_{j=s}^{m}\rho_j\rho_{j-s}\right]\right), \end{aligned} \tag{9}$$

this is non-zero if $\mu = 0$ and $\rho_0 \neq 0$. However by not setting $\mu = 0$ we can now still get an expected return of zero when $\rho_0 \neq 0$ if μ is chosen to exactly resemble the second term. Since $\kappa > 0$ the sign of μ then depends on the sign of ρ_0. Because of the volatility feedback effect generally present in financial returns the expected estimated median of the returns will be positive.

Since both the return shock and the log-volatility are related to the volatility shock the sign of ρ_0 determines the way the returns are skewed. With the in financial returns typically found $\rho_0 < 0$ a positive volatility shock is accompanied by a contemporaneous negative return shock implying a negatively skewed distribution of returns.

In this model the volatility shocks itself are correlated, which can be argued for. As it is now assumed the market is not efficient so the shock of yesterday can still influence the market today.

The model could be extended to include a negative lag such that it is possible to test for a reverse causality in the shocks. If the new model including the negative lag performs better from a likelihood perspective or can perform better forecasts this would indicate the reverse causality would be worth to investigate further.

3.3 Extended SV model

Teulings et al. (2020) perform a study where they use the VIX, an index for the expected future volatility, as a proxy for future volatility. They find that daily changes (squared to different powers) of the VIX explain up to 60% of the daily changes of the S&P 500 with a negative correlation. Although this is a small and simple setup using a simple regression model this is a promising enough result to investigate this different relation between return and volatility shocks as not just the notion that a volatility shock is determined by past return shocks. This inspired us to investigate a reverse causality where is also a relation between the current volatility shock and the future return shock. Catania (2022) mentions in the conclusion that the model can be extended to include a negative lag and would be optimal to test this idea with. The extended model of Catania (2022) includes one future return shock in the volatility shock equation and will be

$$\begin{aligned} y_t &= \mu + \sigma_t \varepsilon_t, \qquad \varepsilon_t \overset{\text{iid}}{\sim} N(0,1), \\ \sigma_t &= \exp\left(\frac{h_t}{2}\right), \\ h_t &= c + \varphi h_{t-1} + \kappa \eta_t, \\ \eta_t &= \sum_{j=-1}^{m} \rho_j \varepsilon_{t-j} + \sqrt{1 - \sum_{j=-1}^{m} \rho_j^2} * b_t, \quad b_t, \overset{\text{iid}}{\sim} N(0,1), \end{aligned} \tag{10}$$

with the same restrictions placed on $\kappa > 0$, $|\varphi| < 1$ and $\sum_{j=-1}^{m} \rho_j < 1$ as before. Including this extra relation assumes that the current volatility shock is correlated with the future return shock. From an economical point of view this could be argued for, investors could fear the future expected return shock and adjust for it today. Following this reasoning we would expect the value of ρ_{-1} to be negative.

The extended model shares very similar statistical characteristics with the original model only now slightly altered to account for the incorporation of the future return shock in the equation of the current volatility shock. As can be seen in the unconditional distribution of the log-volatility,

$$h \sim \mathrm{N}\left(\frac{c}{1-\varphi}, \frac{\kappa^2}{1-\varphi^2}\left[1 + 2\sum_{s=1}^{m} \varphi^s \sum_{j=s-1}^{m} \rho_j \rho_{j-s}\right]\right). \tag{11}$$

The expected return of tomorrow given today will now also be different as it will now be

$$E\left[y_{t+1}\right]=\mu+\frac{1}{2}\kappa\rho_0\exp\left(\frac{c}{2(1-\varphi)}+\frac{\kappa^2}{8\left(1-\varphi^2\right)}\left[1+2\sum_{s=0}^{m}\varphi^s\sum_{j=s-1}^{m}\rho_j\rho_{j-s}\right]\right). \tag{12}$$

The same as for the original Catania model the expected value will be zero if $\mu=0$ and $\rho_0\neq 0$. For the expected return to be zero if $\rho_0\neq 0$ the value of μ will now be different to account for the extra ρ_{-1} term. If $\rho_{-1}=0$ the original model is retrieved with the same characteristics.

To estimate these SV models two different methods are used to address the filtering and estimation problem.

4 Bellman Filter

The first method is a new method developed for the filtering of state space models, that of Lange (2020). The Bellman filter is a generalisation of the Kalman filter based on Bellman's dynamic programming principle and is based on the mode. The goal is to find good filtered states $\boldsymbol{a}_{1|1},\ldots,\boldsymbol{a}_{t|t}$ that maximise the likelihood $\ell\left(\boldsymbol{a}_1,\ldots,\boldsymbol{a}_t,y_1,\ldots,y_t\right)$.

4.1 Filtering

Assume the existence of the mode and for now that the fixed hyper-parameters $\boldsymbol{\theta}$ are known. All following equations are indexed on $\boldsymbol{\theta}$ however for notational purposes this is suppressed.
The function $h^{\boldsymbol{\theta}}\left(\boldsymbol{a}_{t-1},\boldsymbol{\eta}_t\right)$ in the state equation of (1) takes the general form

$$\boldsymbol{a}_t=\boldsymbol{c}+\boldsymbol{T}\boldsymbol{a}_{t-1}+\boldsymbol{\eta}_t,\quad \boldsymbol{\eta}_t\overset{\text{iid}}{\sim}N(0,\boldsymbol{Q}). \tag{13}$$

Define a value function that maximises the likelihood with respect to all states apart from the most recent state $\boldsymbol{a}_t$ as,

$$V_t\left(\boldsymbol{a}_t\right):=\max_{\boldsymbol{a}_1,\ldots,\boldsymbol{a}_{t-1}}\quad\ell\left(\boldsymbol{a}_1,\ldots,\boldsymbol{a}_t,y_1,\ldots,y_t\right), \tag{14}$$

or rewritten like equation (3) such that the value function satisfies the recursive relation known as Bellman's equation,

$$V_t\left(\boldsymbol{a}_t\right)=\ell\left(y_t\mid\boldsymbol{a}_t\right)+\max_{\boldsymbol{a}_{t-1}}\left\{\ell\left(\boldsymbol{a}_t\mid\boldsymbol{a}_{t-1}\right)+V_{t-1}\left(\boldsymbol{a}_{t-1}\right)\right\}. \tag{15}$$

The filtered states will then be

$$\begin{bmatrix} \boldsymbol{a}_{t|t} \\ \boldsymbol{a}_{t-1|t} \end{bmatrix} := \underset{\begin{bmatrix} \boldsymbol{a}_{t} \\ \boldsymbol{a}_{t-1} \end{bmatrix} \in \mathbb{R}^{2m}}{\arg\max} \left\{ \ell \left(y_t \mid \boldsymbol{a}_t \right) + \ell \left(\boldsymbol{a}_t \mid \boldsymbol{a}_{t-1} \right) + V_{t-1} \left(\boldsymbol{a}_{t-1} \right) \right\}, \tag{16}$$

where the estimate of the previous state is revised now that there is more information available.

The value function $V_{t-1}()$ is usually not available in closed form but the only thing needed is its behaviour around the peak. Lange (2020) use a quadratic approximation that is parameterised by the location of its maximum $\boldsymbol{a}_{t-1|t-1}$ and the negative Hessian, also known as the information matrix $\boldsymbol{I}_{t-1|t-1}$, as the researcher's knowledge will be approximated by a quadratic function. This approximation will be

$$V_{t-1}(\boldsymbol{a}_{t-1}) = -\frac{1}{2}(\boldsymbol{a}_{t-1} - \boldsymbol{a}_{t-1|t-1})' \boldsymbol{I}_{t-1|t-1} (\boldsymbol{a}_{t-1} - \boldsymbol{a}_{t-1|t-1}) + \text{constants}. \tag{17}$$

Plugging this equation into the maximisation a viable function-space is created where standard optimisation methods such as Newton's method (Nocedal & Wright, 1999) can be deployed

$$\begin{bmatrix} \boldsymbol{a}_{t} \\ \boldsymbol{a}_{t-1} \end{bmatrix} \leftarrow \begin{bmatrix} \boldsymbol{a}_{t} \\ \boldsymbol{a}_{t-1} \end{bmatrix} + \begin{bmatrix} \boldsymbol{J}_t^{11} - \frac{\mathrm{d}^2 \ell(y_t|\boldsymbol{a})}{\mathrm{d}\boldsymbol{a}\mathrm{d}\boldsymbol{a}'} & \boldsymbol{J}_t^{12} \\ \boldsymbol{J}_t^{21} & \boldsymbol{I}_{t-1|t-1} + \boldsymbol{J}_t^{22} \end{bmatrix}^{-1} \begin{bmatrix} \boldsymbol{J}_t^{1} + \frac{\mathrm{d} \ell(y_t|\boldsymbol{a}_t)}{\mathrm{d}\boldsymbol{a}_t} \\ \boldsymbol{J}_t^{2} - \boldsymbol{I}_{t|t-1}(\boldsymbol{a}_{t-1} - \boldsymbol{a}_{t-1|t-1}) \end{bmatrix}, \tag{18}$$

where

$$\begin{bmatrix} \boldsymbol{J}_t^{1} \\ \boldsymbol{J}_t^{2} \end{bmatrix} := \begin{bmatrix} \frac{\mathrm{d} \ell(\boldsymbol{a}_t|\boldsymbol{a}_{t-1})}{\mathrm{d}\boldsymbol{a}_t} \\ \frac{\mathrm{d} \ell(\boldsymbol{a}_t|\boldsymbol{a}_{t-1})}{\mathrm{d}\boldsymbol{a}_{t-1}} \end{bmatrix}, \quad \begin{bmatrix} \boldsymbol{J}_t^{11} & \boldsymbol{J}_t^{12} \\ \boldsymbol{J}_t^{21} & \boldsymbol{J}_t^{22} \end{bmatrix} := - \begin{bmatrix} \frac{\mathrm{d}^2 \ell(\boldsymbol{a}_t|\boldsymbol{a}_{t-1})}{\mathrm{d}\boldsymbol{a}_t \mathrm{d}\boldsymbol{a}_t'} & \frac{\mathrm{d}^2 \ell(\boldsymbol{a}_t|\boldsymbol{a}_{t-1})}{\mathrm{d}\boldsymbol{a}_t \mathrm{d}\boldsymbol{a}_{t-1}'} \\ \frac{\mathrm{d}^2 \ell(\boldsymbol{a}_t|\boldsymbol{a}_{t-1})}{\mathrm{d}\boldsymbol{a}_{t-1} \mathrm{d}\boldsymbol{a}_t'} & \frac{\mathrm{d}^2 \ell(\boldsymbol{a}_t|\boldsymbol{a}_{t-1})}{\mathrm{d}\boldsymbol{a}_{t-1} \mathrm{d}\boldsymbol{a}_{t-1}'} \end{bmatrix} \tag{19}$$

Another well known optimisation method, that of Fisher, is achieved by replacing $\frac{\mathrm{d}^2 \ell(y_t|\boldsymbol{a}_t)}{\mathrm{d}\boldsymbol{a}_t \mathrm{d}\boldsymbol{a}_t'}$ with its expectation. These optimisation methods can also be combined where $\frac{\mathrm{d}^2 \ell(y_t|\boldsymbol{a}_t)}{\mathrm{d}\boldsymbol{a}_t \mathrm{d}\boldsymbol{a}_t'}$ is replaced with a weighted sum of $\frac{\mathrm{d}^2 \ell(y_t|\boldsymbol{a}_t)}{\mathrm{d}\boldsymbol{a}_t \mathrm{d}\boldsymbol{a}_t'}$ and $E\left(\frac{\mathrm{d}^2 \ell(y_t|\boldsymbol{a}_t)}{\mathrm{d}\boldsymbol{a}_t \mathrm{d}\boldsymbol{a}_t'}\right)$. To start the optimisation (after initialising $\boldsymbol{a}_{0|0}, \boldsymbol{I}_{0|0}$ as the unconditional mean and a sufficiently large multiple of the identity matrix) predict the next state and information matrix as

$$\boldsymbol{a}_{t|t-1} = \underset{\boldsymbol{a}_t \in \mathbb{R}}{\arg\max}\, \ell(\boldsymbol{a}_t \mid \boldsymbol{a}_{t-1|t-1}) \tag{20}$$

and

$$\boldsymbol{I}_{t|t-1} = \boldsymbol{J}_t^{11} - \boldsymbol{J}_t^{12}(\boldsymbol{I}_{t-1|t-1} + \boldsymbol{I}_t^{22})^{-1}\boldsymbol{J}_t^{21} \mid_{\boldsymbol{a}_t=\boldsymbol{a}_{t|t-1}, \boldsymbol{a}_{t-1}=\boldsymbol{a}_{t-1|t-1}} . \tag{21}$$

Now using Newtons step for the optimisation as

$$\begin{aligned}
&\boldsymbol{S}_t \leftarrow \boldsymbol{J}_t^{11} - \boldsymbol{J}_t^{12}(\boldsymbol{I}_{t-1|t-1} + \boldsymbol{J}_t^{22})^{-1}\boldsymbol{J}_t^{21} - \frac{\mathrm{d}^2\ell\left(y_t \mid \boldsymbol{a}_t\right)}{\mathrm{d}\boldsymbol{a}_t \mathrm{d}\boldsymbol{a}_t'}, \\
&\boldsymbol{D}_t \leftarrow \boldsymbol{I}_{t-1|t-1} + \boldsymbol{J}_t^{22}, \\
&\boldsymbol{G}_t^1 \leftarrow \boldsymbol{J}_t^1 + \frac{\mathrm{d}\ell\left(y_t \mid \boldsymbol{a}_t\right)}{\mathrm{d}\boldsymbol{a}_t}, \\
&\boldsymbol{G}_t^2 \leftarrow \boldsymbol{J}_t^2 - \boldsymbol{I}_{t-1|t-1}(\boldsymbol{a}_{t-1} - \boldsymbol{a}_{t-1|t-1}), \\
&\boldsymbol{a}_t \leftarrow \boldsymbol{a}_t + \boldsymbol{S}_t^{-1}\boldsymbol{G}_t^1 - \boldsymbol{S}_t^{-1}\boldsymbol{J}_t^{12}\boldsymbol{D}_t^{-1}\boldsymbol{G}_t^2, \\
&\boldsymbol{a}_{t-1} \leftarrow \boldsymbol{a}_{t-1} - \boldsymbol{D}_t^{-1}\boldsymbol{J}_t^{21}\boldsymbol{S}_t^{-1}\boldsymbol{G}_t^1 + (\boldsymbol{D}_t^{-1} + \boldsymbol{D}_t^{-1}\boldsymbol{J}_t^{21}\boldsymbol{S}_t^{-1}\boldsymbol{J}_t^{12}\boldsymbol{D}_t^{-1})\boldsymbol{G}_t^2.
\end{aligned} \tag{22}$$

After this optimisation set $\boldsymbol{a}_{t|t} = \boldsymbol{a}_t$, $\boldsymbol{a}_{t-1|t} = \boldsymbol{a}_{t-1}$ and $\boldsymbol{I}_{t|t-1} = \boldsymbol{J}_t^{11} - \boldsymbol{J}_t^{12}(\boldsymbol{I}_{t-1|t-1} + \boldsymbol{J}_t^{22})^{-1}\boldsymbol{J}_t^{21} \mid_{\boldsymbol{a}_t=\boldsymbol{a}_{t|t-1}, \boldsymbol{a}_{t-1}=\boldsymbol{a}_{t-1|t-1}}$. And move to the next time step to repeat the entire process again.

4.1.1 Dealing with degenerate densities

Lange (2020) also provide a slightly altered version of the Bellman filter to deal with model filtering and estimation of models with degenerate observation and/or state-transition densities. If the state-transition density is degenerate some elements of the current state are deterministic functions of the previous state and can be dropped from the optimisation. To account for this the prediction of the state and information matrix will be changed to

$$\begin{aligned}
&\boldsymbol{a}_{t|t-1} = \boldsymbol{c} + \boldsymbol{T}\boldsymbol{a}_{t-1|t-1}, \\
&\boldsymbol{I}_{t|t-1} = \boldsymbol{Q}^{-1} - \boldsymbol{Q}^{-1}\boldsymbol{T}\left(\boldsymbol{I}_{t-1|t-1} + \boldsymbol{T}'\boldsymbol{Q}^{-1}\boldsymbol{T}\right)^{-1}\boldsymbol{T}'\boldsymbol{Q}^{-1} = \left(\boldsymbol{T}\boldsymbol{I}_{t-1|t-1}^{-1}\boldsymbol{T}' + \boldsymbol{Q}\right)^{-1},
\end{aligned} \tag{23}$$

with $\boldsymbol{Q}$ as the covariance matrix of the shocks in the state equation. Also the optimisation step will now be

$$\boldsymbol{a}_t \leftarrow \boldsymbol{a}_t + \left\{\boldsymbol{I}_{t|t-1} - \frac{\mathrm{d}^2\ell\left(y_t \mid \boldsymbol{a}_t\right)}{\mathrm{d}\boldsymbol{a}_t \mathrm{d}\boldsymbol{a}_t'}\right\}^{-1}\left\{\frac{\mathrm{d}\ell\left(y_t \mid \boldsymbol{a}_t\right)}{\mathrm{d}\boldsymbol{a}_t} - \boldsymbol{I}_{t|t-1}\left(\boldsymbol{a}_t - \boldsymbol{a}_{t|t-1}\right)\right\}, \tag{24}$$

and a slightly different updating step of the information matrix as this will now be

$$\boldsymbol{I}_{t|t} = \boldsymbol{I}_{t|t-1} - \left.\frac{\mathrm{d}^2\ell\left(y_t \mid \boldsymbol{a}_t\right)}{\mathrm{d}\boldsymbol{a}_t\mathrm{d}\boldsymbol{a}_t'}\right|_{\boldsymbol{a_t}=\boldsymbol{a}_{t|t}}. \tag{25}$$

4.2 Parameter estimation

So far it was assumed the set of hyper parameters $\boldsymbol{\theta}$ was known, however this will obviously not be the case and along with the state will also need to be estimated. This is done by decomposing the likelihood. A single observation's contribution to the likelihood will be

$$\ell(y_t \mid \mathcal{F}_{t-1}) = \ell(y_t, \boldsymbol{\alpha}_t \mid \mathcal{F}_{t-1}) - \ell(\boldsymbol{\alpha}_t \mid y_t, \mathcal{F}_{t-1}) = \ell(y_t \mid \boldsymbol{\alpha}_t) + \ell(\boldsymbol{\alpha}_t \mid \mathcal{F}_{t-1}) - \ell(\boldsymbol{\alpha}_t \mid \mathcal{F}_t). \tag{26}$$

But as the exact state $\boldsymbol{\alpha}_t$ is not known it is impossible to calculate this quantity exactly, the closest approximation will be to evaluate the expression at the filtered state $\boldsymbol{a}_{t|t}$ found by the Bellman filter. Swapping the last two terms and evaluating the likelihood at $\boldsymbol{a}_{t|t}$ leaves a nice optimisation function where the first part represents the "fit" which we needs to be maximised, penalised by the realised Kullback-Leibler (KL) divergence between the filtered and predicted state,

$$\ell(y_t \mid \mathcal{F}_{t-1}) = \ell(y_t \mid \boldsymbol{a}_{t|t}) - \left\{\ell(\boldsymbol{a}_{t|t} \mid \mathcal{F}_t) - \ell(\boldsymbol{a}_{t|t} \mid \mathcal{F}_{t-1})\right\}. \tag{27}$$

However the exact expression to calculate the KL divergence is also generally unknown, but can be approximated. Like the researcher's knowledge before it is once again presumed the function can be approximated by a multivariate quadratic function, such that the two terms in the KL divergence can be approximated by

$$\begin{aligned}
\ell\left(\boldsymbol{\alpha}_t \mid \mathcal{F}_t\right) &\approx \frac{1}{2}\log\det\left\{\boldsymbol{I}_{t|t}/(2\pi)\right\} - \frac{1}{2}\left(\boldsymbol{\alpha}_t - \boldsymbol{a}_{t|t}\right)'\boldsymbol{I}_{t|t}\left(\boldsymbol{\alpha}_t - \boldsymbol{a}_{t|t}\right)\\
\ell\left(\boldsymbol{\alpha}_t \mid \mathcal{F}_{t-1}\right) &\approx \frac{1}{2}\log\det\left\{\boldsymbol{I}_{t|t-1}/(2\pi)\right\} - \frac{1}{2}\left(\boldsymbol{\alpha}_t - \boldsymbol{a}_{t|t-1}\right)'\boldsymbol{I}_{t|t-1}\left(\boldsymbol{\alpha}_t - \boldsymbol{a}_{t|t-1}\right),
\end{aligned} \tag{28}$$

where $\boldsymbol{\alpha}_t$ will be replaced with its closest approximation $\boldsymbol{a}_{t|t}$. Using the output of the Bellman filter all terms are now either known or computable using known quantities and thus allow one to calculate an approximate likelihood. To obtain the estimates $\widetilde{\boldsymbol{\theta}}$, a maximum-likelihood estimator can be used where

$$\widetilde{\boldsymbol{\theta}} := \underset{\boldsymbol{\theta}}{\operatorname{argmax}} \sum_{t=t_0+1}^{N} \left\{ \ell(y_t \mid \boldsymbol{a_{t|t}}) + \frac{1}{2} \log\det\left\{\boldsymbol{I}_{t|t-1}\right\} - \frac{1}{2} \log\det\left\{\boldsymbol{I}_{t|t}\right\} - \frac{1}{2}\left(\boldsymbol{a}_{t|t} - \boldsymbol{a}_{t|t-1}\right)' \boldsymbol{I}_{t|t-1} \left(\boldsymbol{a}_{t|t} - \boldsymbol{a}_{t|t-1}\right) \right\}. \tag{29}$$

4.3 Filtering and estimating the SV models with the Bellman filter

As the basic SV model in equation (6) is actually nested in the model of equation (7) and can be retrieved with $m = 0, \rho_0 = 0$ it is only needed to workout how to filter and estimate the model in (7) and then set $m = 0$ and $\rho_0 = 0$ to retrieve the needed formulas for the simple SV model. In his paper Catania (2022) found that a m between 2 and 5 performs better than previously defined models. But to keep notation general we will continue with m and once again all distributions in the filtering steps are indexed with $\boldsymbol{\theta}$ but this will be suppressed for convenience.

To start first define the state as $\boldsymbol{a}_t = (h_t, h_{t-1}, \ldots, h_{t-m})'$ In state space representation conditional on the information at $t-m-1$, $\mathcal{F}_{t-m-1}$, the shock will be jointly normally distributed as

$$\begin{bmatrix} \eta_t \\ \varepsilon_t \\ \varepsilon_{t-1} \\ \vdots \\ \varepsilon_{t-m} \end{bmatrix} \mid \mathcal{F}_{t-m-1} \sim \mathrm{N}\left(\begin{bmatrix} 0 \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & \rho_0 & \rho_1 & \ldots & \rho_m \\ \rho_0 & 1 & 0 & \ldots & 0 \\ \rho_1 & 0 & 1 & \ldots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho_m & 0 & 0 & 0 & 1 \end{bmatrix} \right). \tag{30}$$

Now, conditional on all information known at time $t-1$, where $\mathcal{F}_{t-1}, \boldsymbol{a}_{t-1}$ imply $\varepsilon_{t-1}, \ldots, \varepsilon_{t-m}$, the joint normal distribution of the current shocks η_t, ε_t will be

$$\begin{bmatrix} \eta_t \\ \varepsilon_t \end{bmatrix} \mid \mathcal{F}_{t-1}, \boldsymbol{a}_{t-1} \sim \mathrm{N}\left(\begin{bmatrix} \sum_{j=1}^{m} \rho_j \varepsilon_{t-j} \\ 0 \end{bmatrix}, \begin{bmatrix} 1 - \sum_{j=1}^{m} \rho_j^2 & \rho_0 \\ \rho_0 & 1 \end{bmatrix} \right). \tag{31}$$

which follows from the Normal lemma. From rewriting $\varepsilon_{t-j} = (y_{t-j} - \mu)\exp(-h_{t-j}/2)$ together with that the marginal distribution of η_t is Gaussian and that the state-transition equation is a linear transformation of η_t it implies that

$$
\begin{aligned}
&h_t \mid \mathcal{F}_{t-1}, \boldsymbol{a}_{t-1} \sim \mathrm{N}\left(\mu_{h,t}, \sigma^2_{h,t}\right), \quad \text{where} \\
&\mu_{h,t} = c + \varphi h_{t-1} + \kappa \sum_{j=1}^{m} \rho_j \frac{y_{t-j} - \mu}{\exp\left(h_{t-j}/2\right)}, \quad \sigma_{h,t} = \kappa \sqrt{1 - \sum_{j=1}^{m} \rho_j^2}.
\end{aligned} \tag{32}
$$

Using the conditional-Gaussian lemma again, the distribution of ε_t conditional on $\mathcal{F}_{t-1}, \boldsymbol{a}_{t-1}$ and η_t will be

$$
\begin{aligned}
&\varepsilon_t \mid \mathcal{F}_{t-1}, \boldsymbol{a}_{t-1}, \eta_t \sim N\left(\mu_{\varepsilon,t}, \sigma^2_{\varepsilon,t}\right), \quad \text{where} \\
&\mu_{\varepsilon,t} = \frac{\rho_0}{1 - \sum_{j=1}^{m} \rho_j^2} \left(\eta_t - \sum_{j=1}^{m} \rho_j \frac{y_{t-j} - \mu}{\exp(h_{t-j}/2)} \right), \quad \sigma_{\varepsilon,t} = \sqrt{1 - \frac{\rho_0^2}{1 - \sum_{j=1}^{m} \rho_j^2}}.
\end{aligned} \tag{33}
$$

Now since $\boldsymbol{a}_{t-1}$ and η_t together imply $\boldsymbol{a}_t$, the distribution of y_t conditional on $\mathcal{F}_{t-1}, \boldsymbol{a}_t$ will be $\mathrm{N}(\mu_{y,t}, \sigma^2_{y,t})$ with

$$
\mu_{y,t} = \mu + \exp(h_t/2)\mu_{\varepsilon,t}, \tag{34}
$$

and

$$
\sigma^2_{y,t} = \exp(h_t)\sigma^2_{\varepsilon,t}. \tag{35}
$$

The state-transition is a degenerate Gaussian with pdf

$$
p\left(\boldsymbol{a}_t \mid \boldsymbol{a}_{t-1}, \mathcal{F}_{t-1}\right) = \frac{1}{\sigma_{h,t}\sqrt{2\pi}} \exp\left(-\frac{\left(h_t - \mu_{h,t}\right)^2}{2\sigma^2_{h,t}} \right) \times \prod_{j=1}^{m} \delta\left(a_{j+1,t} - a_{j,t-1}\right), \tag{36}
$$

where $a_{j,t}$ is the j-th element in $\boldsymbol{a}_t = (h_t, h_{t-1}, \ldots, h_{t-m})$ and δ is the Dirac delta function with mass at zero such that the second element of $\boldsymbol{a}_t$ equals that first element of $\boldsymbol{a}_{t-1}$. Because of these variables that the current state and past state have in common it is a degenerate density. Because this model has a degenerate state-transition density we will need to use the modified filter of Subsection 4.1.1.

The optimisation to be solved will be

$$\begin{aligned}
&\underset{h_t,h_{t-1},\ldots,h_{t-m-1}}{\text{argmax}} \left\{\ell\left(y_t \mid \boldsymbol{a}_t, \mathcal{F}_{t-1}\right) + \ell\left(h_t \mid \boldsymbol{a}_{t-1}, \mathcal{F}_{t-1}\right) + V_{t-1}(\boldsymbol{a}_{t-1})\right\}, \\
&\underset{h_t,h_{t-1},\ldots,h_{t-m-1}}{\text{argmax}} \{-\frac{1}{2}\log(2\pi) - \log(\sigma_{y,t}) + \frac{(y_t-\mu_{y,t})^2}{\sigma_{y,t}^2} + -\frac{1}{2}\log(2\pi) - \log(\sigma_{h,t}) + \frac{(h_t-\mu_{h,t})^2}{\sigma_{h,t}^2} + \\
&\frac{1}{2}\log\det\left\{\boldsymbol{I}_{t|t-1}\right\} - \frac{1}{2}\log\det\left\{\boldsymbol{I}_{t|t}\right\} - \frac{1}{2}\left(\boldsymbol{a}_{t|t} - \boldsymbol{a}_{t|t-1}\right)' \boldsymbol{I}_{t|t-1}\left(\boldsymbol{a}_{t|t} - \boldsymbol{a}_{t|t-1}\right)\},
\end{aligned} \tag{37}$$

instead of the usual optimisation in equation (16).

In the optimisation step we will use an equally weighted sum of the Newton updating steps and the Fisher updating steps. The exact optimisation and updating steps of the state can be found in Appendix A.

For the parameter estimation the distribution we use the maximisation in equation (29) knowing that $y_t \mid \boldsymbol{a}_t, \mathcal{F}_{t-1} \sim N(\mu_{y,t}, \sigma^2_{y,t})$ and using the Gaussian log-likelihood with these parameters.

Now as the extended model is very similar to the original model of Catania (2022) the derivation will be very similar only now altered to include the negative lag. In state space representation conditional on the information at $t-m-1$, $\mathcal{F}_{t-m-1}$, the shocks will be jointly normally distributed as

$$\begin{bmatrix} \eta_t \\ \varepsilon_{t+1} \\ \varepsilon_t \\ \varepsilon_{t-1} \\ \vdots \\ \varepsilon_{t-m} \end{bmatrix} \mid \mathcal{F}_{t-m-1} \sim \mathrm{N}\left(\begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & \rho_{-1} & \rho_0 & \rho_1 & \ldots & \rho_m \\ \rho_{-1} & 1 & 0 & 0 & 0 & 0 \\ \rho_0 & 0 & 1 & 0 & 0 & 0 \\ \rho_1 & 0 & 0 & 1 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho_m & 0 & 0 & 0 & \ldots & 1 \end{bmatrix}\right). \tag{38}$$

Now, conditional on the past shocks $\varepsilon_{t-1}, \ldots, \varepsilon_{t-m}$ implied by $\mathcal{F}_{t-1}, \boldsymbol{a}_{t-1}$ the joint normal distribution of the current and future shocks $\eta_t, \varepsilon_{t+1}, \varepsilon_t$ will be

$$\begin{bmatrix} \eta_t \\ \varepsilon_{t+1} \\ \varepsilon_t \end{bmatrix} \mid \mathcal{F}_{t-1}, h_{t-1}, h_{t-2}, \ldots, h_{t-m} \sim \mathrm{N}\left(\begin{bmatrix} \sum_{j=1}^m \rho_j \varepsilon_{t-j} \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1-\sum_{j=1}^m \rho_j^2 & \rho_{-1} & \rho_0 \\ \rho_{-1} & 1 & 0 \\ \rho_0 & 0 & 1 \end{bmatrix}\right). \tag{39}$$

Declare the state to be $\boldsymbol{a}_t = \varepsilon_{t+1}, h_t, h_{t-1}, \ldots, h_{t-m}$ and the distribution of h_t conditional on $\mathcal{F}_{t-1}$,

$\varepsilon_t, h_{t-1}, h_{t-2}, \ldots, h_{t-m}$ defined as $\boldsymbol{a}_{t-1}$ can be derived to be

$$h_t \mid \mathcal{F}_{t-1}, \boldsymbol{a}_{t-1} \sim \mathrm{N}\left(\mu^*_{h,t}, {\sigma^*_{h,t}}^2\right), \quad \text{where}$$
$$\mu^*_{h,t} = c + \varphi h_{t-1} + \kappa \sum_{j=1}^{m} \rho_j \frac{y_{t-j} - \mu}{\exp\left(h_{t-j}/2\right)} + \kappa \frac{\rho_0}{1 - \sum_{j=1}^{m} \rho_j^2} \varepsilon_t, \quad \sigma^*_{h,t} = \kappa \sqrt{1 - \sum_{j=0}^{m} \rho_j^2}, \tag{40}$$

as h_t is a linear transformation of η_t and using the conditional-Gaussian Lemma. The difference with the original model is that the mean equation now contains the extra term $\kappa\rho_0\varepsilon_t$ and the sum of the correlation parameters now starts at zero instead of one.

For the conditional distribution of ε_t, we use the multivariate conditional-Gaussian lemma such that

$$\varepsilon_t \mid \mathcal{F}_{t-1}, h_{t-1}, h_{t-2}, \ldots, h_{t-m}, \eta_t, \varepsilon_{t+1} \sim \mathrm{N}\left(\mu^*_{\varepsilon,t}, \sigma^{2*}_{\varepsilon,t}\right), \quad \text{where}$$
$$\mu^*_{\varepsilon,t} = \begin{bmatrix} \rho_0 & 0 \end{bmatrix} \begin{bmatrix} 1 - \sum_{j=1}^{m} \rho_j^2 & \rho_{-1} \\ \rho_{-1} & 1 \end{bmatrix}^{-1} \left(\begin{bmatrix} \frac{h_t - c - \varphi h_{t-1}}{\kappa} \\ \varepsilon_{t+1} \end{bmatrix} - \begin{bmatrix} \sum_{j=1}^{m} \rho_j \frac{y_{t-j}-\mu}{\exp(h_{t-j}/2)} \\ 0 \end{bmatrix} \right), \tag{41}$$
$$\sigma^*_{\varepsilon,t} = \sqrt{1 - \begin{bmatrix} \rho_0 & 0 \end{bmatrix} \begin{bmatrix} 1 - \sum_{j=1}^{m} \rho_j^2 & \rho_{-1} \\ \rho_{-1} & 1 \end{bmatrix}^{-1} \begin{bmatrix} \rho_0 \\ 0 \end{bmatrix}},$$

where η is replaced with $\frac{h_t - c - \varphi h_{t-1}}{\kappa}$ and ε_{t-j} with $\frac{y_{t-j}-\mu}{\exp(h_{t-j}/2)}$ for $j = 1, 2, \ldots, m$. Now $\mathcal{F}_{t-1}, h_{t-1}, h_{t-2}, \ldots, h_{t-m}$, $\eta_t, \varepsilon_{t+1}$ imply $\mathcal{F}_{t-1}, \boldsymbol{a}_t$.

The distribution of y_t conditional on the state and available information will be $N(\mu^*_{y,t}, \sigma^{2*}_{y,t})$ with

$$\mu^*_{y,t} = \mu + \exp(h_t/2)\mu^*_{\varepsilon,t}$$
$$\mu + \frac{\rho_0 \exp(h_t/2)}{1 - \sum_{j=1}^{m} \rho_j^2 - \rho_{-1}^2} \left[\left(\frac{h_t - c - \varphi h_{t-1}}{\kappa} - \sum_{j=1}^{m} \rho_j \frac{y_{t-j} - \mu}{\exp(h_{t-j}/2)} \right) - \rho_{-1}\varepsilon_{t+1} \right], \tag{42}$$

and

$$\sigma^*_{y,t} = \exp(h_t/2)\sigma^*_{\varepsilon,t},$$
$$= \exp(h_t/2)\sqrt{1 - \frac{\rho_0^2}{1 - \sum_{j=1}^{m} \rho_j^2 - \rho_{-1}^2}}. \tag{43}$$

The differences between this distribution of y_t and that of equation (34), (35) are the extra term $\rho_{-1}\varepsilon_{t+1}$ in (42) and the replacement of $\sum_{j=1}^{m} \rho_j^2$ with $\sum_{j=1}^{m} \rho_j^2 - \rho_{-1}^2$.

Since the state $\boldsymbol{a}_t$ includes the future shock ε_{t+1} the state transition density also changes to

$$
\begin{aligned}
p\left(\boldsymbol{a}_t \mid \boldsymbol{a}_{t-1}, \mathcal{F}_{t-1}\right) = & \frac{1}{\sigma_{h,t}^{*}\sqrt{2\pi}} \exp\left(-\frac{\left(h_t - \mu_{h,t}^{*}\right)^2}{2{\sigma_{h,t}^{*}}^2}\right) \times \prod_{j=1}^{m-1} \delta\left(a_{j+1,t} - a_{j,t-1}\right) \\
& \times \frac{1}{\sigma_{\varepsilon,t+1}\sqrt{2\pi}} \exp\left(-\frac{\left(\varepsilon_{t+1} - \mu_{\varepsilon,t+1}\right)^2}{2\sigma_{\varepsilon,t+1}^2}\right).
\end{aligned} \tag{44}
$$

as the element $j+1$ in $a_{j+1,t}$ is equal to the j-th element in $a_{j,t-1}$ for $j = 1, \ldots, m-1$. The last term comes from the distribution of ε_{t+1} conditional on the past state $\boldsymbol{a}_{t-1}$ and h_t. Because h_t and ε_{t+1} are correlated you can not draw them independently. Here we assume h_t is drawn first and then calculate the distribution of ε_{t+1} given this h_t. Since h_t is implied by η_t this distribution can be derived by using the Multivariate conditional-Gaussian lemma on equation (39),

$$
\begin{aligned}
& \varepsilon_{t+1} \mid \mathcal{F}_{t-1}, h_t, h_{t-1}, h_{t-2}, \ldots, h_{t-m}, \eta_t, \varepsilon_t \sim \mathrm{N}\left(\mu_{\varepsilon,t+1}, \sigma_{\varepsilon,t+1}^2\right), \qquad \text{where} \\
\mu_{\varepsilon,t+1} &= \begin{bmatrix} \rho_{-1} & 0 \end{bmatrix} \begin{bmatrix} 1 - \sum_{j=1}^{m} \rho_j^2 & \rho_0 \\ \rho_0 & 1 \end{bmatrix}^{-1} \left(\begin{bmatrix} \frac{h_t - c - \varphi h_{t-1}}{\kappa} \\ \varepsilon_t \end{bmatrix} - \begin{bmatrix} \sum_{j=1}^{m} \rho_j \frac{y_{t-j} - \mu}{\exp(h_{t-j}/2)} \\ 0 \end{bmatrix} \right), \\
&= \frac{\rho_{-1}\left(\frac{h_t - c - \varphi h_{t-1}}{\kappa} - \sum_{j=1}^{m} \rho_j \frac{y_{t-j} - \mu}{\exp(h_{t-j}/2)}\right) - \rho_{-1}\rho_0\varepsilon_t}{1 - \sum_{j=0}^{m} \rho_j^2}, \\
\sigma_{\varepsilon,t+1} &= \sqrt{1 - \begin{bmatrix} \rho_{-1} & 0 \end{bmatrix} \begin{bmatrix} 1 - \sum_{j=1}^{m} \rho_j^2 & \rho_0 \\ \rho_0 & 1 \end{bmatrix}^{-1} \begin{bmatrix} \rho_{-1} \\ 0 \end{bmatrix}}, \\
&= \sqrt{1 - \frac{\rho_{-1}^2}{1 - \sum_{j=0}^{m} \rho_j^2}}.
\end{aligned} \tag{45}
$$

If $\rho_{-1} = 0$ then ε_{t+1} conditional on the past state and h_t it will simply be a standard normal variable.

The optimisation to be solved will be

$$
\begin{aligned}
&\underset{\varepsilon_{t+1},h_t,h_{t-1},\dots,h_{t-m-1}}{\operatorname{argmax}} \left\{\ell\left(y_t \mid \boldsymbol{a}_t, \mathcal{F}_{t-1}\right) + \ell\left(h_t \mid \boldsymbol{a}_{t-1}, \mathcal{F}_{t-1}\right) + \ell\left(\varepsilon_{t+1} \mid \boldsymbol{a}_{t-1}, h_t, \mathcal{F}_{t-1}\right) + V_{t-1}(\boldsymbol{a}_{t-1})\right\}, \\
&\underset{\varepsilon_{t+1},h_t,h_{t-1},\dots,h_{t-m-1}}{\operatorname{argmax}} \{-\frac{1}{2}\log(2\pi) - \log(\sigma^*_{y,t}) - \frac{(y_t-\mu^*_{y,t})^2}{2\sigma^{2*}_{y,t}} + -\frac{1}{2}\log(2\pi) - \log(\sigma^*_{h,t}) - \frac{(h_t-\mu^*_{h,t})^2}{2\sigma^{2*}_{h,t}} + \\
&\qquad - \frac{1}{2}\log(2\pi) - \log(\sigma_{\varepsilon,t+1}) - \frac{(\varepsilon_{t+1}-\mu_{\varepsilon,t+1})^2}{2\sigma^2_{\varepsilon,t+1}} \\
&\qquad \frac{1}{2}\log\det\left\{\boldsymbol{I}_{t|t-1}\right\} - \frac{1}{2}\log\det\left\{\boldsymbol{I}_{t|t}\right\} - \frac{1}{2}\left(\boldsymbol{a}_{t|t} - \boldsymbol{a}_{t|t-1}\right)' \boldsymbol{I}_{t|t-1}\left(\boldsymbol{a}_{t|t} - \boldsymbol{a}_{t|t-1}\right)\},
\end{aligned}
\tag{46}
$$

instead of the usual optimisation in equation (16).

The optimisation step of equation (18) and the gradient and Hessian of equation (19) will now be be slightly different as the state now also contains ε_{t+1} and you will now also need to take the derivatives with regarding to ε_{t+1} leading to an extra dimension. This will also be the case for the state transition density or seen as the log sum of its relevant parts, $\log\{p(h_t \mid \boldsymbol{a}_{t-1}, \mathcal{F}_{t-1})p(\varepsilon_{t+1} \mid \boldsymbol{a}_{t-1}, h_t, \mathcal{F}_{t-1})\}$ as it will also have an extra dimension. The altered optimisation and updating steps can be found in Appendix B.

For the parameter estimation the distribution we use is the maximisation in equation (29) knowing that $y_t \mid \boldsymbol{a}_t, \mathcal{F}_{t-1} \sim N(\mu^*_{y,t}, \sigma^{2*}_{y,t})$ and using the Gaussian log-likelihood with these parameters.

5 CSIR of Malik and Pitt (2011)

Particle filters are another relatively easy to implement and widely applicable methods to solve the filtering and estimation problem. Particle filters are based on the principle of Bayesian updating, combining a prior with a likelihood can construct the density of the state conditional on all available information.

The most commonly used particle filter in econometric literature is that of Malik and Pitt (2011). They develop the continuous sampling importance resampling (CSIR) method. This method is popular as it handles a common drawback of particle filters, that of parameter estimation. This is also the filter Catania (2022) used in their paper to obtain their results and will be a useful benchmark to include.

5.1 Filtering with SIR

The CSIR method is a continuous time extension of the sampling importance resampling (SIR) algorithm of Gordon, Salmond, and Smith (1993). In this Section again it is assumed all densities are indexed by the set of hyper-parameters $\boldsymbol{\theta}$ which is suppressed for notational convenience.
A particle filter uses that the measurement density $p(y_t|\boldsymbol{a}_t)$ can be calculated and one can simulate from the Markov transition density $p(\boldsymbol{a}_{t+1}|\boldsymbol{a}_t)$. Particle filters are based on recursively simulating from the approximate prediction density

$$p(\boldsymbol{a}_{t+1} \mid \mathcal{F}_t) = \int p(\boldsymbol{a}_{t+1} \mid \boldsymbol{a}_t) p(\boldsymbol{a_t} \mid \mathcal{F}_t) d\boldsymbol{a}_t, \tag{47}$$

and then Bayesian updating as

$$p(\boldsymbol{a}_{t+1} \mid \mathcal{F}_{t+1}) = \frac{p(\boldsymbol{y}_{t+1} \mid \boldsymbol{a}_{t+1}) p(\boldsymbol{a}_{t+1} \mid \mathcal{F}_t)}{p(\boldsymbol{y}_{t+1} \mid \mathcal{F}_{t+1})}. \tag{48}$$

This results in a sample which is approximately distributed as the true filtering density

$$p(\boldsymbol{a}_{t+1} \mid \mathcal{F}_{t+1}) \propto p(\boldsymbol{y}_{t+1} \mid \boldsymbol{a}_{t+1}) \int p(\boldsymbol{a}_{t+1} \mid \boldsymbol{a}_t) p(\boldsymbol{a_t} \mid \mathcal{F}_t) d\boldsymbol{a}_t. \tag{49}$$

To sample from this density, the SIR algorithm starts with a random set of $\boldsymbol{a}_t^1, \ldots, \boldsymbol{a}_t^N$ "particles" with probability masses $\pi_t^1, ..., \pi_t^N$ drawn from $p(\boldsymbol{a}_t \mid \mathcal{F}_t)$ then for each time point t in $t = 0, \ldots, T-1$ follows the following algorithm

Algorithm 1 SIR

1: It will sample $\tilde{\boldsymbol{a}}_{t+1}^k \sim p(\boldsymbol{a}_{t+1}|\boldsymbol{a}_t^k)$ for $k = 1, ..., N$
2: For each sample $\tilde{\boldsymbol{a}}_{t+1}^k$ calculate the normalised weights $\pi_{t+1}^k = \frac{\omega_{t+1}^k}{\sum_{i=1}^N \omega_{t+1}^i}$, where $\omega_{t+1}^k = p\left(y_{t+1} \mid \tilde{\boldsymbol{a}}_{t+1}^k\right)$.
3: Sample (from the mixture) $\boldsymbol{a}_{t+1}^k \sim \sum_{k=1}^N \pi_{t+1}^k \delta\left(\boldsymbol{a}_{t+1} - \tilde{\boldsymbol{a}}_{t+1}^k\right)$ for each $k = 1, ..., N$. With δ the Dirac-delta function with mass at zero.

In this last step the original samples are resampled to achieve an equally weighted sample. Malik and Pitt (2011) show that the first sample in step 1 will converge to samples from $p(\boldsymbol{a}_{t+1}|\mathcal{F}_t)$ with N large and the samples from step 3 converge to $p(\boldsymbol{a}_{t+1}|\mathcal{F}_{t+1})$. To obtain the filtered estimates of the state now calculate the average of all particles k from step 3, $\boldsymbol{a}_t = \frac{1}{N} \sum_{k=1}^N \boldsymbol{a}_t^k$.

5.2 Parameter estimation and CSIR

For maximum-likelihood parameter estimation it is required to be able to calculate or approximate the likelihood. Here the contribution to the likelihood of a single observation can be estimated with the output of step 2 as $p(y_t \mid \mathcal{F}_{t-1}) = \int p(y_t \mid \boldsymbol{a}_t, y_{t-1}) p(\boldsymbol{a}_t \mid \mathcal{F}_{t-1}) d\boldsymbol{a}_t \approx \frac{1}{N}\sum_{k=1}^{N} p(y_t \mid \boldsymbol{a}_t^k, y_{t-1}) = \frac{1}{N}\sum_{k=1}^{N}\omega_t^k$ and the log-likelihood will thus be $\log\left(\frac{1}{N}\sum_{k=1}^{N}\omega_t^k\right)$ However the original SIR algorithm has a discontinuous likelihood estimator as it uses the approximating empirical distribution function (EDF)

$$\widehat{F}_N(\boldsymbol{a}_{t+1}) = \frac{\frac{1}{N}\sum_{k=1}^{N} p\left(y \mid \boldsymbol{a}_{t+1}^k\right) I\left(\boldsymbol{a}_{t+1}^k < \boldsymbol{a}_{t+1}\right)}{\frac{1}{N}\sum_{i=1}^{N} p\left(y \mid \boldsymbol{a}_{t+1}^i\right)} \tag{50}$$

in step 3 of SIR the samples come from

$$\widehat{F}_N(\boldsymbol{a}_{t+1}) = \sum_{k=0}^{N} \pi^k I\left(\boldsymbol{a}_{t+1} - \boldsymbol{a}_{t+1}^k\right). \tag{51}$$

As this is not a continuous function common gradient-based methods to obtain parameter estimates fail. To circumvent this issue Malik and Pitt (2011) introduce a new way to sample in step 3. This only works for one-dimensional models so now considering the one dimensional state a_{t+1}, to get a continuous likelihood-function they first sort all particles in ascending order such that $a_{t+1}^1 \leq \ldots, \leq a_{t+1}^N$ and replace the EDF with

$$\widetilde{F}_N(\boldsymbol{a}_{t+1}) = \sum_{k=0}^{N} \lambda^k \mathrm{G}_k\left(\frac{\boldsymbol{a}_{t+1} - \boldsymbol{a}_{t+1}^{(k)}}{\boldsymbol{a}_{t+1}^{(k+1)} - \boldsymbol{a}_{t+1}^{(k)}}\right), \tag{52}$$

with $a_{t+1}^0 = -\infty$ and $a_{t+1}^{N+1} = \infty$. Here $\lambda_{t+1}^0 = \pi_{t+1}^1/2, \lambda_{t+1}^N = \pi_{t+1}^N/2$ and $\lambda_{t+1}^k = (\pi_{t+1}^{k+1} + \pi_{t+1}^k)/2$ for $k = 1, \ldots, N-1$ and G_k is a uniform distribution on [0,1]. To reduce sample impoverishment they use a stratification scheme instead of generating a set of N uniform variates. Simply simulate one $u \sim UID(0,1)$ and generate

$$u_j = \frac{(j-1)+u}{N} \tag{53}$$

for $j = 1, \ldots, N$. And then follow Algorithm 2 to obtain the samples $a_{t+1}^{1*}, \ldots, a_{t+1}^{N*}$ from $\widetilde{F}_N(\boldsymbol{a}_{t+1})$.

This resampling scheme combined with the SIR algorithm yields the CSIR algorithm where the likelihood is now approximated as a smooth function in $\boldsymbol{\theta}$ and thus is also suited to a standard maximum-likelihood estimator to obtain estimates for $\boldsymbol{\theta}$.

Algorithm 2 Continuous resampling

Set $s = 0, j = 1$ and initialise a vector $r^1, \dots, r^N$.
for i=0,..., N **do** s = s + λ^i_{t+1}
  while $\mathrm{u}_j \leq s, j \leq N$ **do** $\mathrm{r}^j = i$
    $u^*_j = (u_j - (s - \lambda^i_{t+1}))/\lambda^i_{t+1}$
    $j = j + 1$ =0
for j=1,..., N **do** **if** $\mathrm{r}^j = 0$ **then**
  set $a^{j*}_{t+1} = a^1_{t+1}$
  if $r^j = N$ **then**
    set $a^{j*}_{t+1} = a^N_{t+1}$
  else
    set $a^{j*}_{t+1} = (a^{r^{j+1}}_{t+1} - a^{r^j}_{t+1}) \times u^*_j + a^{r^j}_{t+1}$

5.3 Estimating the SV models with CSIR

Before one can estimate of the SV model in equation (6) it is required to derive several distributions such as the unconditional distribution of the state, where the state $\boldsymbol{a}_t$ will simply be h_t, and $h_t \sim \mathrm{N}(\frac{c}{1-\varphi}, \frac{\kappa^2}{1-\varphi^2})$. The state-transition distribution will be

$$\begin{aligned} &h_t \mid h_{t-1}, \mathcal{F}_{t-1} \sim \mathrm{N}(\mu_{h,t}, \sigma^2_{h,t}), \quad \text{where} \\ &\mu_{h,t} = c + \varphi h_{t-1}, \\ &\sigma^2_{h,t} = \kappa^2. \end{aligned} \tag{54}$$

And lastly

$$\begin{aligned} &y_t \mid h_t, \mathcal{F}_{t-1} \sim \mathrm{N}(\mu_{y,t}, \sigma^2_{y,t}), \quad \text{where} \\ &\mu_{y,t} = \mu, \quad \sigma^2_{y,t} = \exp(h_t). \end{aligned} \tag{55}$$

Plugging these three distributions into the CSIR algorithm will give the desired filtered densities and allow for parameter estimation with the maximum likelihood estimator.

For the CSIR method to be able to estimate the model in equation (7) needs to be rewritten in

Gaussian nonlinear state space representation with uncorrelated errors, as

$$
\begin{aligned}
y_t - \mu = & \exp\left(\frac{h_t}{2}\right)\left[\sqrt{1-\frac{\rho_0^2}{1-\sum_{j=1}^m \rho_j^2}}u_t + \frac{\rho_0}{\kappa\left(1-\sum_{j=1}^m \rho_j^2\right)}\right. \\
& \left.\times\left[h_t - c - \varphi h_{t-1} - \kappa\sum_{j=1}^m \rho_j \frac{y_{t-j}}{\exp(h_{t-j}/2)}\right]\right], \qquad (56) \\
h_t = & c + \varphi h_{t-1} + \kappa\sum_{j=1}^m \rho_j \frac{y_{t-j}}{\exp(h_{t-j}/2)} + \kappa\sqrt{1-\sum_{j=1}^m \rho_j^2 \varpi_t},
\end{aligned}
$$

where u_t, ϖ_t are jointly standard Gaussian variables with $E[u_t \varpi_t] = 0$. This can be done by using that the disturbances η, ε are correlated and conditionally Gaussian as can be seen in equation (30), η_t can be rewritten as $\sum_{j=1}^m \rho_j \varepsilon_{t-j} + \sqrt{1-\sum_{j=1}^m \rho_j^2 \varpi_t}$ and $\varepsilon_{t-j} = \frac{y_{t-j}}{\exp(h_{t-j}/2)}$ for $j = 1, \ldots, m$. Since the correlation between ε_t and η_t equals ρ_0 and ε_t is independent of ε_{t-j}, for $j = 1, \ldots, m$ $Corr(\varepsilon_t, \varpi_t) = \frac{\rho_0}{\sqrt{1-\sum_{j=1}^m \rho_j^2}}$ and ε_t can thus be rewritten as $\varepsilon_t = \sqrt{1-\frac{\rho_0^2}{1-\sum_{j=1}^m \rho_j^2}}u_t + \frac{\rho_0}{\kappa\left(1-\sum_{j=1}^m \rho_j^2\right)}\varpi_t$ yielding the equation above.

The unconditional distribution of h_t is $\sim \mathrm{N}(\frac{c}{1-\varphi}, r^2)$ with

$$
r^2 = \frac{\kappa^2}{1-\varphi^2}\left[1 + 2\sum_{s=1}^m \varphi^s \sum_{j=s}^m \rho_j \rho_{j-s}\right]. \qquad (57)
$$

Now the distribution of y_t given the state $\boldsymbol{a}_t = (h_t, h_{t-1}, \ldots, h_{t-m})$ and the available information $\mathcal{F}_{t-1}$, will be $\mathrm{N}(\mu_{y,t}, \sigma_{y,t}^2)$, where

$$
\mu_{y,t} = \exp\left(\frac{h_t}{2}\right)\frac{\rho_0}{\kappa\left(1-\sum_{j=1}^m \rho_j^2\right)} \times \left(h_t - c - \varphi h_{t-1} - \kappa\sum_{j=1}^m \rho_j \frac{y_{t-j}}{\exp(h_{t-j}/2)}\right), \qquad (58)
$$

and

$$
\sigma_{y,t}^2 = \exp(h_t)\left(1 - \frac{\rho_0^2}{1-\sum_{j=1}^m \rho_j^2}\right). \qquad (59)
$$

And $(h_t \mid \boldsymbol{a}_{t-1}, \mathcal{F}_{t-1}) \sim \mathrm{N}(\mu_{h,t}, \sigma_{h,t}^2)$ with

$$
\mu_{h,t} = c + \varphi h_{t-1} + \kappa\sum_{j=1}^m \rho_j \frac{y_{t-j}}{\exp(h_{t-j}/2)}, \qquad (60)
$$

and

$$\sigma_{h,t}^2 = \kappa^2 \left(1 - \sum_{j=1}^{m} \rho_j^2 \right). \tag{61}$$

So the CSIR algorithm will look as follows,

Algorithm 3 CSIR

For $k = 1 : N$ to obtain the samples $\boldsymbol{a}_0^k = (h_0^k, h_{-1}^k, \ldots, h_{-m}^k)'$ draw from the unconditional distribution of $h \sim \mathrm{N}(\frac{c}{1-\varphi}, r^2)$.
For $t = 1 : T - 1$:
Samples $\boldsymbol{a}_t^k \sim p(\boldsymbol{a}_t \mid \mathcal{F}_t)$ are known for $k = 1, \ldots, N$.
=0

1: For $k = 1 : N$, sample h_{t+1}^k from $p(h_{t+1} \mid \boldsymbol{a}_t^k, \mathcal{F}_t)$ and set the $j+1$-th element of $\boldsymbol{a}_{t+1}^k$ equal to the j-th element of $\boldsymbol{a}_t^k$ for $j = 1, \ldots, m-1$.
2: For $k = 1 : N$, calculate the normalised weights as $\pi_{t+1}^k = \frac{\omega_{t+1}^k}{\sum_{i=1}^N \omega_{t+1}^i}$, where $\omega_{t+1}^k = p\left(y_{t+1} \mid \boldsymbol{a}_{t+1}^k, \mathcal{F}_t\right)$.
3: For $k = 1 : N$, follow Algorithm 2

6 Simulation study

To evaluate the performance of the filtering and estimation methods, we first compare the mean absolute errors (MAE) of the one-step ahead predictions of the state and parameter estimates for the model of Section 3.2 with two inter-temporal leverage effects.

We simulate our data with the parameters $c = 0, \varphi = 0.975$ and $\kappa = 0.1$, $\rho_0 = -0.30$, $\rho_1 = -0.70$ and $\rho_2 = 0.30$. These values are similar to those used in past simulation studies or found in empirical applications. We use 20 samples each containing 5000 simulated data points. For the Bellman filter we set the optimisation steps to be at most 100 and for the particle filter we set $N = 5000$ as more particles do not lead to an enough increase of performance to warrant the extra computational time required.

For each sample we split the data into two parts, where we use the first half to train the model and estimate the parameters and the second half to calculate the MAE of the one-step-ahead predictions. On top of this, we also calculate the MAE based on the true parameter to investigate the effect of filtering inaccuracy.

As we see in Table 1, the Bellman filter delivers stable and accurate parameter estimations apart from the estimated variance of the volatility shock κ. This parameter is consistently estimated too low as can be seen by the small standard deviation. The estimated leverage effect parameters are accurate compared to those of the particle filter, but the ρ_2 estimate is much too low. Also all estimates are pretty unstable as can be seen by the large standard deviation in the sample and numerical standard estimates.

There are differences between the Bellman filter and the particle filter, the particle filter estimate of the persistence in the log-volatility equation φ is less accurate and has an standard deviation ten times as high as that of the Bellman filter. The leverage effect estimates are also inaccurate with a small standard deviation, which would indicate the particle filter estimates do not vary a lot across samples, however their mean is far from the true value. The particle filter also produces in most ($\sim 80\%$) cases a singular and non-invertible Hessian matrix as the likelihood function can be not perfectly smooth, which makes it impossible to calculate the numerical standard errors.

From a filtering aspect using the true parameters the Bellman filter produces slightly better one-step-ahead predictions of the log-volatility at more than half of the filtering time of the particle filter. This difference in computational time becomes even more prominent when estimating the parameters as the particle filter its estimation time is almost three times as large as that of the Bellman filter. This limits the amount of particles that we can use even though an even larger amount of particles may be needed to increase the accuracy of the predictions.

This result is mostly in line with the results in Lange (2020) where he finds that using the Bellman filter to obtain parameter estimates leads to better one-step-ahead predictions and more accurate parameter estimates.

Comparing the numerical standard errors with the standard deviation of the sample means hints at incorrectly calculated standard errors in Catania (2022) as he uses the particle filter to obtain the results. This could be explained by looking at the disadvantages of using the CSIR method for multivariate state estimation. Malik and Pitt (2011) suggest a univariate approach to obtain a continuous likelihood function that maximises only the first element in the current state and does not update the past estimates based on the new information, it only updates the first element of the current state. The resampling step is introduced to ensure a continuous likelihood but this does not exactly do as promised as this function will not be perfectly smooth and can contain kinks in the likelihood. A non-perfectly smooth function could lead to poorly estimated Hessian matrices when they are evaluated near kinks or on linear pieces causing the negative Hessian to not be invertible or equal to zero. This would lead to inaccurate standard errors calculated by the CSIR method as they use a piecewise linear approximation to calculate the standard errors. This combined with the fact that gradient-based optimisers could fail for a non-perfectly smooth function causing the user to resort to non-gradient based optimisers which have a considerably bigger computational burden. This most likely plays a big part in the difference in the computational times seen in Table 1.

Table 1: Average parameters estimates with standard deviations and the average of numerical standard errors, MAE of the one-step-ahead predictions and the computational time taken for data simulated from the model in (7) with $m = 2$.

	Parameters							MAE (true)	MAE (est)	Time taken (seconds)	
	μ	c	φ	κ	ρ_0	ρ_1	ρ_2	$h_{t\|t-1}$	$h_{t\|t-1}$	Filtering	Estimation
True value	0.05	0	0.975	0.1	-0.30	-0.70	0.30	h_t	h_t		
Bellman filter	0.0420	0.0002	0.9632	0.0443	-0.2787	-0.6496	0.0056	0.2064	0.2526	2.2631	893.06
St dev	0.0261	0.0018	0.0339	0.0149	0.3049	0.2956	0.1374	0.0181	0.0304	0.1951	562.44
Num St dev	0.0230	0.0015	0.0095	0.0216	0.3195	0.4926	0.6594				
Particle filter	0.0421	0.0003	0.9278	0.0507	-0.1495	-0.0017	-0.1668	0.2066	0.2840	5.2312	2104.6
St dev	0.0081	0.0160	0.3047	0.0753	0.0361	0.0371	0.1651	0.0186	0.0338	0.5608	865.02
Num st dev	0.0133	0.0010	0.0074	0.0096	0.0247	0.0242	0.0229				

Note: MAE = mean absolute error. Given are the averages of 20 samples of 5000 simulated data points where the first 2500 observations are used to estimate the parameters and are then used to calculate the MAE of the predicted one-step-ahead state for the last 2500 observations. Two types of standard errors are given, the first is standard deviation for means of the samples and the second is a numerical standard deviation calculated by taking the average of the square root of the diagonal of the inverted Hessian matrix if it is invertible otherwise it will be ignored for calculating the mean (For the particle filter 80% of the cases). For the particle filter the amount of particles is $N = 5000$ and the Bellman filter has at most 100 optimisation steps.

Since the DGP is not always known in advance we investigate the results of estimating an incorrectly specified model that does not exactly capture the reality. To do so both filters are used to estimate models based on simulated data generated from the following DGPs:

- The basic SV model in equation 6,
- The SV model in equation (7) with $m = 0$ and $\rho_0 = -0.7$, such that it is equal to the asymmetric SV model of Jacquier et al. (2004), that only contains a contemporaneous leverage effect,
- The SV model of Catania (2022) with $m = 1$ but $\rho_0 = 0$ and $\rho_1 = -0.7$ to get the asymmetric SV model Harvey and Shephard (1996) with only an inter-temporal leverage effect,
- The SV model of Catania (2022) with $m = 1$, $\rho_0 = -0.3$ and $\rho_1 = -0.7$,
- The extended SV model found in equation (10) with $m = 0$, $\rho_{-1} = -0.1$ and $\rho_0 = -0.7$,
- The extended SV model now with $m = 1$, $\rho_{-1} = -0.1$, $\rho_0 = -0.3$ and $\rho_1 = -0.7$,

where all again use $\mu = 0.05$ $c = 0$, $\varphi = 0.975$ and $\kappa = 0.1$.

For each DGP first the Bellman filter is used to estimate the parameters of the Catania model of Section 3.2 with $m = 0, 1, 2$ and for $m = 1$ with $\rho_0 = 0$. The same is then repeated using the particle filter.
The parameter estimates in Tables 2:7 are the means of the estimated parameters for each sample based on the first 2500 observations. The last 2500 observations are used to compute the MAE of the predicted state for each sample based on the estimated parameters.

For the basic SV model of equation (6) the Bellman filter delivers slightly more accurate parameters estimates and also provides a better prediction of the log-volatility as can be seen in Table 2. All models with different leverage effects, apart from the one with only a contemporaneous leverage effect, have similar parameter estimates to those found using the correctly specified model without any leverage effects. The estimates of the leverage effects are estimated to be very small which is as expected as they should be zero. They do have very large numerical standard errors indicating that the estimates are unstable. The model with the contemporaneous timing estimates the median of the return μ to be too low and the persistence of the log-volatility process φ also too low, on top of that it also assigns an average ρ_0 of 0.2033 which is much higher than expected. But because it does estimate the level c and the variance of the volatility shock κ the most accurate it surprisingly performs the second best of all models considering the MAE of the predicted log-volatility. The correctly specified model performs the best.

The estimates using the particle filter do differ from those obtained with the Bellman filter, μ is estimated too high for all models and c and φ estimates are slightly more inaccurate. The main difference is in the κ and leverage effect estimates. The variance of the volatility shock κ estimate is even more inaccurate, where the model with the contemporaneous leverage timing even estimates a value that differs more than 80% with its true value. The leverage effect parameters have much larger values than expected as they are not close to 0 but can reach values up to 0.2561. This all results in a worse MAE of log-volatility predictions than with the Bellman filter. Considering the MAE of the predictions, the SV model of (7) with $m = 1$ performs the best which is not what we expected as this is not the correctly specified model. Even more surprising is that the correctly specified model has the second worst MAE of the predictive log-volatility, which is most likely a result of the poor parameter estimation of κ.

Table 2: Bellman filter parameter estimates and MAE for simulated data from the basic SV model in equation (6).

	Parameter estimates								MAE
	μ	c	φ	κ	ρ_{-1}	ρ_0	ρ_1	ρ_2	$h_{t\|t-1}$
True value	0.05	0	0.975	0.1	-	-	-	-	h_t
Bellman filter	0.0525	0.0014	0.9637	0.0624	-	-	-	-	**0.3144**
	(0.0194)	(0.0021)	(0.0278)	(0.0304)	-	-	-	-	[0.0306]
	0.0339	-0.0004	0.9558	0.0682	-	0.2033	-	-	0.3158
	(0.0209)	(0.0016)	(0.0189)	(0.0224)	-	(0.1687)	-	-	[0.0305]
	0.0568	0.0009	0.9695	0.0533	-	-	0.0215	-	0.3168
	(0.0205)	(0.0016)	(0.0193)	(0.0229)	-	-	(0.4490)	-	[0.0309]
	0.0570	0.0008	0.9722	0.0517	-	-0.0360	0.0430	-	0.3165
	(0.0225)	(0.0016)	(0.0201)	(0.0240)	-	(0.2886)	(0.3419)	-	[0.0302]
	0.0571	0.0008	0.9725	0.0565	-	-0.0222	0.0621	-0.03117	0.3176
	(0.0222)	(0.0015)	(0.0183)	(0.0302)	-	(0.2213)	(0.4490)	(0.4518)	[0.0322]
Particle filter	0.0799	-0.0006	0.9899	0.0210	-	-	-	-	0.3553
	(0.0194)	(0.0006)	(0.0064)	(0.0092)	-	-	-	-	[0.0381]
	0.0794	-0.0003	0.9946	0.0155	-	0.1921	-	-	0.3559
	(0.0193)	(0.0004)	(0.0075)	(0.0075)	-	(0.1921)	-	-	[0.0384]
	0.0780	-0.0019	0.9705	0.0430	-	-	0.2561	-	0.3540
	(0.0188)	(0.0014)	(0.0116)	(0.0120)	-	-	(0.1428)	-	[0.0394]
	0.0854	-0.0029	0.9559	0.0602	-	-0.1355	0.1240	-	**0.3505**
	(0.0152)	(0.0019)	(0.0069)	(0.0077)	-	(0.0168)	(0.0412)	-	[0.0368]
	0.0869	-0.0022	0.9652	0.0526	-	-0.2088	0.0544	0.1433	0.3531
	(0.0122)	(0.0014)	(0.0065)	(0.0071)	-	(0.0128)	(0.0335)	(0.0303)	[0.0354]

Note: MAE = mean absolute error. Given are the averages of 20 samples of 5000 simulated data points where the first 2500 observations are used to estimate the parameters and are then used to calculate the MAE of the predicted one-step-ahead state for the last 2500 observations. Two types of standard errors are given, the first is standard deviation for means of the samples and the second is a numerical standard deviation calculated by taking the average of the square root of the diagonal of the inverted Hessian matrix if it is invertible otherwise it will be ignored for calculating the mean. For the particle filter the amount of particles is $N = 1000$ and the Bellman filter has at most 100 optimisation steps.

Parameter estimates for the model of Jacquier et al. (2004) with only a contemporaneous correlation found in Table 3 are more similar using the different filtering and estimation methods. Both estimate the μ parameter too low with the particle filter estimating it even lower. The Bellman filter provides accurate c and φ estimates but they are slightly less accurate for the correctly specified model. The particle filter also performs well for these parameters with only one less accurate estimate for φ for the simple SV model.

The estimation of κ and the leverage effect parameters is again the most troublesome. The Bellman filter estimates the κ parameter $\sim 50\%$ too low for all models. The particle filter performs even worse, for the simple model its estimate is close to zero and all others are around 0.3.

The ρ_0 estimate in the correctly specified model estimated with the Bellman filter is less than 0.01 of its true value compared to the 0.15 when estimated with the particle filter. When the leverage effect timings are incorrectly specified the methods do find large correlations for the incorrect timings. For the Bellman filter these sum up to the true value of ρ_0. For the particle filter this pattern can also be seen although a little less close to the true value. Apart from the model with only an inter-temporal correlation, where both methods assign a value of less than 0.05 of the true value of ρ_0 to ρ_1.

The two estimation methods do differ more considering the MAE of the predicted log-volatility, all models estimated with the Bellman filter outperform the best performing model estimated with the particle filter. This best performing model is those without leverage effects. For the Bellman filter estimated models the models with leverage effects do perform much better than the simple model without the leverage effects. But the best performing model is not the correctly specified one but the one with both a contemporaneous and one inter-temporal leverage effect, possibly due to the less accurate estimates of c and φ.

Table 3: Bellman filter parameter estimates and MAE for simulated data from the asymmetric SV model of Jacquier et al. (2004) found in equation (7) when setting $m = 0$.

	Parameter estimates								
	μ	c	φ	κ	ρ_{-1}	ρ_0	ρ_1	ρ_2	$h_{t\|t-1}$
True value	0.05	0	0.975	0.1	-	-0.7	-	-	h_t
Bellman filter	0.0342	0.0006	0.9707	0.0530	-	-	-	-	0.3239
	(0.0192)	(0.0014)	(0.0178)	(0.0211)	-	-	-	-	[0.0287]
	0.0407	0.0040	0.9603	0.0445	-	-0.6916	-	-	0.2905
	(0.0200)	(0.0015)	(0.0109)	(0.0141)	-	(0.2099)	-	-	[0.0438]
	0.0280	0.0002	0.9738	0.0468	-	-	-0.6959	-	0.2815
	(0.0201)	(0.0012)	(0.0104)	(0.0133)	-	-	(0.1717)	-	[0.0307]
	0.0362	0.0001	0.9713	0.0486	-	-0.4898	-0.2861	-	**0.2798**
	(0.0222)	(0.0015)	(0.0138)	(0.0165)	-	(0.2629)	(0.2763)	-	[0.0297]
	0.0362	0.0001	0.9709	0.0506	-	-0.4845	-0.2131	-0.0134	0.2808
	(0.0236)	(0.0017)	(0.0227)	(0.0370)	-	(0.4933)	(0.7692)	(0.9328)	[0.0306]
Particle filter	0.0191	-0.0000	0.9971	0.0067	-	-	-	-	**0.3557**
	(0.0200)	(0.0002)	(0.0090)	(0.0123)	-	-	-	-	[0.0296]
	0.0239	-0.0004	0.9734	0.0355	-	-0.5532	-	-	0.3602
	(0.0063)	(0.0009)	(0.0035)	(0.0038)	-	(0.0059)	-	-	[0.0276]
	0.0164	-0.0002	0.9811	0.0263	-	-	-0.6983	-	0.3599
	(0.0198)	(0.0015)	(0.0079)	(0.0122)	-	-	(0.1129)	-	[0.0290]
	0.0243	-0.0003	0.9743	0.0337	-	-0.5215	-0.0896	-	0.3602
	(0.0079)	(0.0009)	(0.0038)	(0.0041)	-	(0.0090)	(0.0104)	-	[0.0280]
	0.0196	-0.0002	0.9760	0.0313	-	-0.3861	-0.3374	0.0856	0.3593
	(0.0078)	(0.0008)	(0.0039)	(0.0045)	-	(0.0090)	(0.0109)	(0.0105)	[0.0293]

Note: MAE = mean absolute error. Given are the averages of 20 samples of 5000 simulated data points where the first 2500 observations are used to estimate the parameters and are then used to calculate the MAE of the predicted one-step-ahead state for the last 2500 observations. Two types of standard errors are given, the first is standard deviation for means of the samples and the second is a numerical standard deviation calculated by taking the average of the square root of the diagonal of the inverted Hessian matrix if it is invertible otherwise it will be ignored for calculating the mean. For the particle filter the amount of particles is $N = 1000$ and the Bellman filter has at most 100 optimisation steps.

Table 4 shows that for the asymmetric model of Harvey and Shephard (1996) the Bellman filter delivers accurate parameter estimates for μ, c and φ and the κ estimates are again about 50% too low. For the correctly specified model the correlation estimates are fairly accurate, -0.7397 against the true value of -0.7. The models with multiple leverage effects assign a value a little smaller than -0.7 to ρ_1 but compensate for this by assigning a weight of about -0.1 to ρ_0. When ρ_2 is also estimated it correctly get assigned a small value however with a huge numerical standard error of almost 1.3. This large standard error causes ρ_1 to also have a standard error three times larger than when it is estimated without ρ_2. So while the average estimates of ρ_0, ρ_1 remain roughly the same and ρ_2 is also estimated close to its true value they now have large standard errors. In the model with only the contemporaneous leverage effect we now find a large value for $\rho_0 = -0.6608$ similar to what happened before when the model with only one inter-temporal leverage was estimated on the data generated from the contemporaneous model, only now reversed. Here the correctly specified model does predict the log-volatility the best according to its MAE.

The particle filter estimates μ and φ to be a little too high and produces accurate c estimates. For the κ parameter the estimates are again too low, for all models the estimates are more than

50% too small. The correlation estimates show a similar pattern to that of the Bellman filter only they are less accurate in estimating the true value of ρ_1 and estimate it too big, apart from the model of equation (7) with $m = 2$ with divides the effect over the three correlation parameters as $\rho_0 = -0.1337, \rho_1 = -0.4829$ and $\rho_2 = -0.3836$. However the simple SV model without leverage effects does deliver the best log-volatility predictions according to the MAE.

Parameter estimates in Table 5 based on data simulated from the SV model of Catania with $m = 1$ show a similar pattern as found before. Fairly accurate parameter predictions for c and φ. Estimated κ values that are much too low, again underestimating the variance of the volatility shock. Here the particle filter also estimates μ to be too low and produces on average inaccurate leverage effect estimates. When estimating the correctly specified model the filter estimates $\rho_0 = 0.0822, \rho_1 = -0.8981$ when the true values are $\rho_0 = -0.3$ and $\rho_1 = -0.7$. When ρ_2 is also estimated it correctly get assigned a small value but the estimates of ρ_0, ρ_1 remain roughly the same. This is also the case for the Bellman filter estimates only its $\rho_0 = -0.2475, \rho_1 = -0.6737$ estimates in the correctly specified model are much closer to their true values. The estimated leverage effect parameters in the Catania model with $m = 2$ are also fairly accurate and even produce the best forecast of the log-volatility according to the MAE, possibly to the more precise estimates of c and φ. For the particle filter the best performing model once again is the simple SV model without leverage effects.

Table 4: Bellman filter parameter estimates and MAE for simulated data from the asymmetric SV model of Harvey and Shephard (1996) found in equation (7) setting $m = 1$ and $\rho_0 = 0$.

	Parameter estimates								
	μ	c	φ	κ	ρ_{-1}	ρ_0	ρ_1	ρ_2	$h_{t\|t-1}$
True value	0.05	0	0.975	0.1	-	-	-0.7	-	h_t
Bellman filter	0.0588	0.0004	0.9685	0.0597	-	-	-	-	0.3180
	(0.0191)	(0.0018)	(0.0185)	(0.0234)	-	-	-	-	[0.0208]
	0.0634	0.0002	0.9780	0.0432	-	-0.6608	-	-	0.2802
	(0.0204)	(0.0012)	(0.0103)	(0.0136)	-	(0.1838)	-	-	[0.0223]
	0.0529	0.0001	0.9734	0.0524	-	-	-0.7397	-	**0.2664**
	(0.0200)	(0.0014)	(0.0115)	(0.0151)	-	-	(0.1559)	-	[0.0195]
	0.0541	0.0000	0.9734	0.0512	-	-0.1120	-0.6608	-	0.2689
	(0.0258)	(0.0016)	(0.0130)	(0.0205)	-	(0.3587)	(0.2815)	-	[0.0233]
	0.0545	0.0000	0.9742	0.0538	-	-0.1007	-0.6532	-0.0111	0.2691
	(0.0224)	(0.0014)	(0.0158)	(0.0523)	-	(0.2759)	(0.8796)	(1.2965)	[0.0204]
Particle filter	0.0689	0.0003	0.9816	0.0454	-	-	-	-	**0.3607**
	(0.0198)	(0.0010)	(0.080)	(0.0115)	-	-	-	-	[0.0252]
	0.0734	0.0003	0.9875	0.0412	-	-0.6584	-	-	0.3741
	(0.0026)	(0.0007)	(0.0017)	(0.0020)	-	(0.0025)	-	-	[0.0347]
	0.0586	0.0003	0.9875	0.0460	-	-	-0.8271	-	0.3850
	(0.0196)	(0.0010)	(0.0043)	(0.0089)	-	-	(0.0551)	-	[0.0414]
	0.0592	0.0003	0.9878	0.0461	-	-0.0583	-0.7780	-	0.3849
	(0.0139)	(0.0008)	(0.0035)	(0.0068)	-	(0.0167)	(0.0215)	-	[0.0421]
	0.0662	0.0003	0.9876	0.0403	-	-0.1337	-0.4829	-0.3294	0.3836
	(0.0608)	(0.0006)	(0.0024)	(0.0038)	-	(0.0066)	(0.0078)	(0.0098)	[0.0420]

Note: MAE = mean absolute error. Given are the averages of 20 samples of 5000 simulated data points where the first 2500 observations are used to estimate the parameters and are then used to calculate the MAE of the predicted one-step-ahead state for the last 2500 observations. Two types of standard errors are given, the first is standard deviation for means of the samples and the second is a numerical standard deviation calculated by taking the average of the square root of the diagonal of the inverted Hessian matrix if it is invertible otherwise it will be ignored for calculating the mean. For the particle filter the amount of particles is $N = 1000$ and the Bellman filter has at most 100 optimisation steps.

Table 5: Bellman filter parameter estimates and MAE for simulated data from the SV model in equation (7) with $m = 1$.

	Parameter estimates								
	μ	c	φ	κ	ρ_{-1}	ρ_0	ρ_1	ρ_2	$h_{t\|t-1}$
True value	0.05	0	0.975	0.1	-	-0.3	-0.7	-	h_t
Bellman filter	0.0532	0.0006	0.9530	0.0824	-	-	-	-	0.3902
	(0.0188)	(0.0030)	(0.0412)	(0.0393)	-	-	-	-	[0.0270]
	0.0634	0.0002	0.9780	0.0432	-	-0.6608	-	-	0.3183
	(0.0204)	(0.0012)	(0.0100)	(0.0138)	-	(0.1947)	-	-	[0.0215]
	0.0398	-0.0001	0.9718	0.0595	-	-	-0.8773	-	0.2994
	(0.0198)	(0.0015)	(0.0079)	(0.0122)	-	-	(0.1129)	-	[0.0305]
	0.0433	-0.0049	0.9646	0.0558	-	-0.2475	-0.6737	-	0.3174
	(0.0190)	(0.0016)	(0.0086)	(0.0144)	-	(0.1973)	(0.1316)	-	[0.0923]
	0.0448	-0.0001	0.9719	0.0593	-	-0.2460	-0.6482	-0.0257	**0.2987**
	(0.0196)	(0.0015)	(0.0080)	(0.0333)	-	(0.1911)	(0.2021)	(0.3794)	[0.0315]
Particle filter	0.0387	0.0008	0.9725	0.0663	-	-	-	-	**0.4506**
	(0.0203)	(0.0016)	(0.0090)	(0.0120)	-	-	-	-	[0.0311]
	0.0152	0.0019	0.9752	0.0519	-	-0.7222	-	-	0.4539
	(0.0032)	(0.0009)	(0.0022)	(0.0022)	-	(0.0034)	-	-	[0.0399]
	0.0308	0.0007	0.9668	0.0634	-	-	-0.8792	-	0.4621
	(0.0198)	(0.0016)	(0.0072)	(0.0103)	-	-	(0.0566)	-	[0.0462]
	0.0241	0.0009	0.9664	0.0696	-	0.0822	-0.8981	-	0.4618
	(0.0160)	(0.0014)	(0.0059)	(0.0082)	-	(0.0202)	(0.0250)	-	[0.0466]
	0.0266	0.0008	0.9677	0.0727	-	0.0568	-0.8203	-0.0380	0.4615
	(0.0094)	(0.0012)	(0.0046)	(0.0065)	-	(0.0144)	(0.0132)	(0.0285)	[0.0463]

Note: MAE = mean absolute error. Given are the averages of 20 samples of 5000 simulated data points where the first 2500 observations are used to estimate the parameters and are then used to calculate the MAE of the predicted one-step-ahead state for the last 2500 observations. Two types of standard errors are given, the first is standard deviation for means of the samples and the second is a numerical standard deviation calculated by taking the average of the square root of the diagonal of the inverted Hessian matrix if it is invertible otherwise it will be ignored for calculating the mean. For the particle filter the amount of particles is $N = 1000$ and the Bellman filter has at most 100 optimisation steps.

Introducing a relation between the future return shock and current volatility shock into the model used to generate the data does have an effect on the estimation results. In Tables 6 and 7 these results can be seen to differ from the results in Tables 3 and 5. The log-volatility level and persistence parameters c and φ are estimated just as well as before apart from the c estimate for the model with two inter-temporal leverage effects estimated with the Bellman filter. The median of the return μ and variance of the volatility shock κ are even estimated slightly better. The particle filter κ estimates are even much better than those estimated for the model without ρ_{-1}. The leverage effect estimates are different now though. Like we could see before the effect of differently timed leverage effects that are present but excluded from the model get compensated by bigger leverage effects on the other timings closest in timing to it. This is exactly what can be seen here, all estimates are bigger to account for the not modelled $\rho_{-1} = -0.1$. For the models with $m = 1$ and $m = 2$ the ρ_0 estimate is now bigger and closer to its true value but still not big enough, ρ_1 is still estimated close to 0.14 while this should be zero. The leverage effect estimates for $m = 2$ are small, only 0.0977 and 0.0517 of their true values but do come with huge numerical standard errors. Indicating these estimates are highly unstable.

The simple SV model remains the best at predicting the log-volatility when the particle filter is the estimation method, although the MAE will be 0.4084 instead of 0.3557 which it is when estimated on the same model with $\rho_{-1} = 0$. If $\rho_{-1} \neq 0$ the particle filter log-volatility predictions worsen a substantial amount.

For the Bellman filter a different model now performs the best, the model containing only one inter-temporal relation. Another interesting difference can be seen for the model of Catania with $m = 2$ which now performs horribly at forecasting the log-volatility, more than doubling the MSE of 0.2808 to 0.6316, where the most likely explanation is the estimate of $c = 0.0514$ which is very large compared to the typical estimates of 0.0004.

Using data generated from the extended model with $m = 1$ we find a similar pattern, μ, c, φ and κ are estimated with the same level of accuracy or sometimes more accurately than when estimated with $\rho_{-1} = 0$ even though they do not model ρ_{-1} explicitly. The main difference lies in the estimated leverage effects and predictive log-volatility performance. The estimated leverage effects with the Bellman filter are fairly similar only now the values of ρ_0 are more negative, around -0.05 up till -0.11. The best performing model according to the MSE is now the model with only one inter-temporal relation. Using the particle filter to estimate the models again results in choosing the simple SV model as the preferred model for predicting the log-volatility.

Table 6: Bellman filter parameter estimates and MAE for simulated data from the extended SV model in equation (10) with $m = 0$.

	Parameter estimates								
	μ	c	φ	κ	ρ_{-1}	ρ_0	ρ_1	ρ_2	$h_{t\|t-1}$
True value	0.05	0	0.975	0.1	-0.1	-0.7	-	-	h_t
Bellman filter	0.0422	0.0005	0.9616	0.0706	-	-	-	-	0.3445
	(0.0188)	(0.0021)	(0.0237)	(0.0281)	-	-	-	-	[0.0347]
	0.0565	0.0008	0.9702	0.0524	-	-0.7726	-	-	0.3077
	(0.0195)	(0.0016)	(0.0083)	(0.0119)	-	(0.1588)	-	-	[0.0807]
	0.0295	-0.0004	0.9669	0.0597	-	-	-0.7521	-	**0.2876**
	(0.0199)	(0.0016)	(0.0113)	(0.0142)	-	-	(0.1328)	-	[0.0341]
	0.0466	-0.0004	0.9689	0.0552	-	-0.6462	-0.1833	-	0.2893
	(0.0221)	(0.0017)	(0.0112)	(0.0161)	-	(0.2831)	(0.3469)	-	[0.0344]
	0.0524	0.0514	0.9635	0.0607	-	-0.6371	-0.0977	-0.0517	0.6316
	(0.0209)	(0.0015)	(0.0122)	(0.0249)	-	(0.2941)	(1.3780)	(1.1967)	[1.5415]
Particle filter	0.0292	0.0001	0.9745	0.0680	-	-	-	-	**0.4084**
	(0.0194)	(0.0015)	(0.0088)	(0.0136)	-	-	-	-	[0.0533]
	0.0362	0.0000	0.9754	0.0690	-	-0.6223	-	-	0.4190
	(0.0041)	(0.0013)	(0.0031)	(0.0034)	-	(0.0042)	-	-	[0.0560]
	0.0176	0.0001	0.9801	0.0554	-	-	-0.6006	-	0.4159
	(0.0197)	(0.0012)	(0.0067)	(0.0112)	-	-	(0.0907)	-	[0.0555]
	0.0412	0.0001	0.9758	0.0782	-	-0.6658	0.1390	-	0.4172
	(0.0023)	(0.0010)	(0.0017)	(0.0020)	-	(0.0023)	(0.0026)	-	[0.0555]
	0.0416	0.0000	0.9756	0.0811	-	-0.6537	0.1380	0.0045	0.4171
	(0.0024)	(0.0010)	(0.0015)	(0.0015)	-	(0.0018)	(0.0019)	(0.0026)	[0.0560]

Note: MAE = mean absolute error. Given are the averages of 20 samples of 5000 simulated data points where the first 2500 observations are used to estimate the parameters and are then used to calculate the MAE of the predicted one-step-ahead state for the last 2500 observations. Two types of standard errors are given, the first is standard deviation for means of the samples and the second is a numerical standard deviation calculated by taking the average of the square root of the diagonal of the inverted Hessian matrix if it is invertible otherwise it will be ignored for calculating the mean. For the particle filter the amount of particles is $N = 1000$ and the Bellman filter has at most 100 optimisation steps.

Table 7: Bellman filter parameter estimates and MAE for simulated data from the extended SV model in equation (10) with $m = 1$.

	Parameter estimates								
	μ	c	φ	κ	ρ_{-1}	ρ_0	ρ_1	ρ_2	$h_{t\|t-1}$
True value	0.05	0	0.975	0.1	-0.1	-0.3	-0.7	-	h_t
Bellman filter	0.0558	0.0007	0.9691	0.0716	-	-	-	-	0.3961
	(0.0187)	(0.0019)	(0.0152)	(0.0225)	-	-	-	-	[0.0301]
	0.0611	0.0000	0.9750	0.0582	-	-0.7758	-	-	0.3186
	(0.0191)	(0.0014)	(0.0053)	(0.0104)	-	(0.1260)	-	-	[0.0279]
	0.0441	-0.0003	0.9717	0.0648	-	-	-0.8502	-	**0.3001**
	(0.0198)	(0.0016)	(0.0080)	(0.0125)	-	-	(0.0992)	-	[0.0289]
	0.0508	-0.0003	0.9718	0.0598	-	-0.3085	-0.6367	-	0.3016
	(0.0217)	(0.0017)	(0.0086)	(0.0158)	-	(0.2285)	(0.1797)	-	[0.0283]
	0.0493	-0.0002	0.9717	0.0636	-	-0.3014	-0.6507	0.0469	0.3010
	(0.0218)	(0.0017)	(0.0097)	(0.0421)	-	(0.2924)	(0.2000)	(0.6069)	[0.0273]
Particle filter	0.0427	0.0008	0.9694	0.0927	-	-	-	-	**0.4780**
	(0.0202)	(0.0020)	(0.0079)	(0.0126)	-	-	-	-	[0.0397]
	0.0526	0.0005	0.9818	0.0593	-	-0.8099	-	-	0.4970
	(0.0027)	(0.0008)	(0.0017)	(0.0026)	-	(0.0029)	-	-	[0.0509]
	0.0303	0.0005	0.9834	0.0485	-	-	-0.9710	-	0.5046
	(0.0199)	(0.0011)	(0.0031)	(0.0064)	-	-	(0.0453)	-	[0.0566]
	0.0481	0.0004	0.9822	0.0488	-	-0.6516	-0.3668	-	0.4994
	(0.0027)	(0.0007)	(0.0018)	(0.0023)	-	(0.0026)	(0.0023)	-	[0.0539]
	0.0449	0.0005	0.9818	0.0456	-	-0.6733	-0.2595	-0.1829	0.5006
	(0.0025)	(0.0007)	(0.0015)	(0.0014)	-	(0.0021)	(0.0021)	(0.0017)	[0.0532]

Note: MAE = mean absolute error. Given are the averages of 20 samples of 5000 simulated data points where the first 2500 observations are used to estimate the parameters and are then used to calculate the MAE of the predicted one-step-ahead state for the last 2500 observations. Two types of standard errors are given, the first is standard deviation for means of the samples and the second is a numerical standard deviation calculated by taking the average of the square root of the diagonal of the inverted Hessian matrix if it is invertible otherwise it will be ignored for calculating the mean. For the particle filter the amount of particles is $N = 1000$ and the Bellman filter has at most 100 optimisation steps.

This simulation study shows that the Bellman filter can generally estimate parameters as accurately as the particle filter with a much smaller computational burden. While both methods consistently underestimate κ, the Bellman filter does estimate the μ and leverage effect parameters more accurately than the particle filter. Both methods estimate big values for the incorrect timing of the leverage effect when the model is misspecified, when the timing from which the leverage effect follows is not included in the model the filters assign big values to the timing that is the closest to the true timing. If too many lags are included the effect gets divided over the multiple legs instead of identifying the correct timing and assigning all other leverage effects to be zero.
The particle filter also fails to estimate the leverage effect parameters accurately even when the model is correctly specified. This results in poorly predicted log-volatility if these leverage effects are included. From this perspective the preferred model will be that without any leverage effects for all generated data sets apart from the data actually generated from the SV model without a leverage effect. Then the preferred model would be that of Catania with a contemporaneous and one inter-temporal leverage effect. This illustrates that we should be careful drawing conclusions from the leverage parameter estimates found with the particle filter.

When the model is correctly specified the Bellman filter leverage estimates are reasonably accurate. In the presence of model misspecification when not enough leverage effects are included this is compensated by assigning a bigger effect to the leverage effect of a different timing. The leverage effect estimates do have a large numerical standard error so while on average they are accurate they are also unstable especially when more inter-temporal lags are added. This needs to be taken into consideration when these estimates are interpreted. Because of the estimation method used in the particle filter it often is unable to calculate the numerical standard errors or incorrectly calculates them, because of this these values do not have a lot of information on the actual standard errors so we can not draw a similar conclusion as done for the Bellman filter.
From a log-volatility predictive point of view the Bellman is also the preferred method of estimation, apart from the poorly estimated model of Catania with $m = 2$ in Table 6, the worst performing model estimated with the Bellman filter always performs better than the best performing model estimated with the particle filter.

7 Empirical study

The estimation and predictive performance of the models that was established in the previous section show that the parameter estimates heavily depend on the underlying DGP and choosing the correct model to estimate the data. Particularly the leverage effects can have large but poorly estimated values while in reality their value should be zero due to the misspecification of the leverage effects in the model. To reduce this risk of model misspecification multiple models are estimated, one without leverage effects and ones that have a contemporaneous leverage effect and up to nine inter-temporal leverage effects.

7.1 Data

For the empirical study we use daily logarithmic returns of two similar financial markets, the Standard & Poor's 500 (S&P 500) and the National Association of Securities Dealers Automated Quotations (Nasdaq) Composite. As they are both based in the US they are expected to share somewhat similar characteristics and should have similar parameter estimates and patterns in the data if the estimator performs well. Another advantage of using US data is that it is the biggest stock market in the world and thus covers a large part of the market capitalisation.
Data ranges from a time period of the 3^{rd} of January 2000 up until the 31^{st} of December 2019 and

will be retrieved from the Oxford-Man Institute of Quantitative Finance. This data contains 5013 data points.

The first 4013 observations of the data set are used to estimate the parameters and the remaining data is considered the out-of-sample data for which the model tries to predict the volatility. Although we do not know the exact volatility to compare the predictions with we can use a proxy, that of squared returns. In econometric literature this is common practice even though this is generally a noisy estimate. The volatility estimates of the Bellman filter for the model of Catania (2022) are obtained by using the first element of the predicted state $h_{t|t-1}$ to then calculate the predicted volatility as $\hat{\sigma}_t^2 = \exp(h_{t|t-1})$. For the extended model $h_{t|t-1}$ is the second element of the state but follows the same calculation from there. For the particle filter the estimated volatility is also calculated as $\hat{\sigma}_t^2 = \exp(h_{t|t-1})$. The performance of the predictions is measured by taking the average of the mean squared error $MSE = \frac{1}{T}\sum_{t=4014}^{5013}(\hat{\sigma}_t^2 - y_t^2)^2$.

As squared returns are not a perfect volatility proxy the same research is performed with realised volatility (Andersen, Bollerslev, Diebold, & Ebens, 2001) now as the proxy for the "true" volatility. The realised volatility is retrieved from the Oxford-Man Institute of Quantitative Finance which calculates the daily 5-min Sub-sampled realised volatility. The realised volatility is sub-sampled into 5 minute samples and then averaged out to reduce the effect of modest amounts of noise present in the prices of assets.

To help pick the preferred model we calculate the Bayesian Information Criterion (BIC) for each model. This criterion reduces the risk of overfitting by introducing a penalty term for the number of parameters in the model. It is calculated as, $\text{BIC} = -2 * \text{LogLik} + \ln \text{T} * np$, where LogLik is the log-likelihood and np is the number of parameters in the model. The preferred model is the model with the lowest BIC.

7.2 Parameter estimates

The parameter estimates using the Nasdaq data are given in Table 8. They indicate that the leverage effect and volatility feedback do play a big role in financial returns. We find that the model of Catania and the extended model estimated with the Bellman filter find similar values for μ ranging from 0.0243 to 0.0450 and for φ ranging from 0.9837 to 0.9942, which is very close to one indicating a high persistence in the log-volatility process. Estimating the extended model instead of the Catania model does result in different level of the log-volatility c estimates, with the

Catania model (excluding the model specification without leverage effects) having estimates that are fairly stable for different inter-temporal lags.The lowest estimate being 0.0019 and the highest 0.0039. This is not the case for the extended model where the values fluctuate between -0.0043 and 0.0042 when lags are included. Resulting in a either positive of negative unconditional mean of the log-volatility. The estimates of κ for the extended model are also estimated much lower than those of the Catania model. Even though the simulation study already found that the Bellman filter already severely underestimates the variance of the volatility shock.

Comparing the estimates to those found with the particle filter the differences are clear. While the parameter estimates of φ are very similar, the estimates of μ and κ are much more unstable and dependant on the leverage effects included. The lowest μ estimate is 0.0100 while the highest found is 0.0595 and the κ estimates vary between 0.1328 and 0.2771 which is much higher than the highest value found with the Bellman filter being 0.2021.

The leverage parameters estimates in the three panels do differ more than the other estimates. However when estimating the model with only a contemporaneous leverage effect they yield similar estimates, all three find values of ρ_0 close to -0.70 with the Bellman filter assigning a small weight of 0.0530 to ρ_{-1} for the extended model. A similar value is assigned to ρ_{-1} for the model with both a contemporaneous and a inter-temporal leverage effect. But for all models with two or more inter-temporal lags the estimated ρ_{-1} is a small negative number between -0.0549 and -0.0267. We should be careful interpreting the weight of ρ_0 in the contemporaneous model as in the simulation study we have seen that the filters can assign a big weight to the wrong leverage effect if the correct timing is omitted.

If we look at the estimation results of the Bellman filter for the original model we can draw some careful conclusions on the trends that we see. The Bellman filter can estimate the leverage effect somewhat accurate if too many legs were included. While in the simulation study it did assign a weight too small to the correct timing and assign some small values to the incorrect timings it was still somewhat accurate. For all models the estimated value of $\rho_0 < -0.4673$ which indicates the presence of the volatility feedback effect in our data. This is followed by large negative estimates of $\rho_1 < -0.5211$ for the models with more than one inter-temporal leverage effect. Also some positive values that vary between 0.4693 and 0.0632 for ρ_2 are found. The last trend that we see for the Bellman filter is that the models which included at least 6 inter-temporal lags find positive values for ρ_6 and for those with less than 6 lags a noticeable positive value for the last lag included. From the simulation study we know that if a lag is omitted the lag with the closest timing will com-

pensate for that, which leads to the suspicion that positive volatility shocks following a negative contemporaneous and one lagged return shock may be partially reversed on day two and day six.
The leverage parameter estimates for the extended model show a similar pattern only with sometimes much smaller values for ρ_0 and bigger values for ρ_1. Adding the correlation between the current volatility shock and future return shock thus reduces the effect of volatility feedback and increases the impact of the leverage effect represented by ρ_1.
As shown in the simulation study, the particle filter is not able to estimate the leverage parameters accurately and this can be seen in the variation of the estimates for the models with different lags. Where in the model with $m = 2$, $\rho_1 = -0.6037$ it is estimated to be 0.5721 when $m = 4$. Because of this is not worth it to analyse these estimates too much.

Now for models estimated on the S&P 500 data there are some differences compared to the models estimated on the Nasdaq data. In Table 9 it can be seen that the estimated volatility feedback effect using the Bellman filter for the Catania model is smaller, seen by the smaller values of ρ_0. For the extended model however it is now estimated to be bigger.
Regarding the other leverage effects there are also some differences. On this data set ρ_1 is estimated to be bigger when estimated for both the original and extended model, where the extended model finds slightly smaller values. ρ_2 is no longer a small positive number as it is now sometimes even negative but ρ_3 is. The same goes for ρ_6 which is now small and mostly negative but ρ_5 is now found to be consistently estimated with positive values although a bit smaller than previously found for ρ_6. Here we can suspect that positive volatility shocks following a contemporaneous and one lagged negative return shock may be partially reversed on day three and day five.
The estimates of ρ_{-1} do not share a clear result, for the models with a small amount of leverage effects it has estimates between 0.1244 and 0.1881, however when more than four inter-temporal leverage effects are included it has small negative estimates similar to those found in Table 8.
The leverage parameters estimated with the particle filter are again heavily dependant on the lags included in the model as they vary heavily, ρ_1 is estimated to be -0.6503 when $m = 3$ and 0.1975 when $m = 9$. Sometimes the estimates do resemble those of the Bellman filter but as seen in the simulation study we should be careful drawing conclusions as the particle filter is not able to accurately estimate the leverage parameters
In the estimation of the first four parameters there are also some differences. The intercept of the log-volatility equation c is estimated to be negative for most models with for the extended model

even values up to -0.0100. This causes the unconditional mean of the log-volatility to generally be negative instead of positive which is the case for the Nasdaq data set when the Catania model is estimated with the Bellman filter. The Bellman filter now also estimates higher values for κ indicating a higher variance of the volatility shocks.

In summary, both data sets estimate the presence of the volatility feedback effect and find that leverage effects are present in the data. Although they vary slightly on the timing of the these leverage effects.

The extra leverage effect in the extended model is generally estimated to be small and negative, but can have small positive estimates as well when few leverage effects are included. We can not draw a clear conclusion on the exact relation between the current volatility shock and the future return shock. However the extra leverage effect does lead to a differently estimated volatility feedback effect and slightly different leverage effects. The estimation of the variance of the volatility shock is also influenced.

7.3 Model selection

The selection of a clear preferred model remains challenging. If the BIC is used, then for the Nasdaq data set the best Bellman filter estimated Catania model is the model with six inter-temporal leverage effects. While for the extended model $m = 9$ is considered the best. For the particle filter it is the model with only a contemporaneous leverage effect, the asymmetric SV model of Jacquier et al. (2004). However these preferred models do not perform the best if we consider their predictive power for the volatility. If we consider the volatility predictions the extended model with $m = 3$ has a MSE of 2.8646 and 0.3369 when the squared returns and realised variance are used as a proxy. The best performing Catania model estimated with the Bellman filter is with $m = 2$ and has a MSE of 3.0100 and 0.4155. This indicates that introducing and estimating the extra leverage effect into the model of Catania is worthwhile from a volatility forecasting perspective when the estimation method is the Bellman filter.

Although the simulation study finds that the particle filter delivers worse or at most just as good parameter estimates as the Bellman filter it performs better predictive wise. The model with $m = 1$ has the lowest MSE of all estimated models for the squared returns being 2.8597. When few leverage effects are included in the model its predicted volatility is also slightly closer to the realised variance. However the extended model still performs better with the best MSE being almost 19%

smaller.

For the models estimated on the S&P 500 data set a similar conclusion can be drawn. Modelling the returns with the extended model yields the best volatility forecasts for both the squared returns and the realised variance, although the best model estimated with the particle filter matches the MSE of the best extended model when the squared returns are used as the true volatility proxy. The difference in performance for the realised variance proxy is now less, a best value of 0.2696 for the extended model compared to the 0.2797 for the particle filter. This best performing model with the particle filter is now one with seven inter-temporal lags even though these performed poorly with the Nasdaq data.
The BIC does select the best predictive model for the Bellman filter estimated original model, the model with $m = 5$, but this still performs worse than the predictions made with most of the models using the extended model or original model estimated with the particle filter. The best performing model predictive-wise is now the extended model with $m = 1$ outperforming all other specifications. Again confirming that this extended model is worthwhile investigating further.

Table 8: Estimation and volatility forecasting results for the Nasdaq with the Bellman filter (top and middle panel) and the particle filter (bottom panel).

μ	c	φ	κ	ρ_{-1}	ρ_0	ρ_1	ρ_2	ρ_3	ρ_4	ρ_5	ρ_6	ρ_7	ρ_8	ρ_9	MSE y^2	MSE rv	LogL(*10^3)	BIC (*10^4)
0.0450	0.0077	0.9899	0.1486												3.0175	0.4496	-6.0468	1.2127
0.0378	0.0030	0.9895	0.1509		-0.6954										3.0164	0.4299	-5.9980	1.2046
0.0320	0.0028	0.9882	0.1439		-0.6158	-0.1530									3.0188	0.4306	-5.9970	1.2044
0.0303	0.0033	0.9895	0.1992		-0.4673	-0.5211	0.4693								**3.0100**	**0.4155**	-5.9887	1.2036
0.0274	0.0039	0.9907	0.1874		-0.5215	-0.5297	0.2151	0.3164							3.0134	0.4237	-5.9852	1.2037
0.0276	0.0036	0.9913	0.1985		-0.4794	-0.5877	0.3046	0.0224	0.2771						3.0133	0.4206	-5.9796	1.2034
0.0308	0.0032	0.9923	0.1892		-0.5476	-0.5159	0.1721	0.1377	-0.0499	0.3515					3.0147	0.4240	-5.9701	1.2023
0.0338	0.0020	0.9926	0.2021		-0.5012	-0.5658	0.2600	0.0536	0.0122	-0.0372	0.3876				3.0134	0.4174	-5.9607	**1.2013**
0.0337	0.0019	0.9925	0.1999		-0.4973	-0.5548	0.1691	0.1592	-0.0521	0.0119	0.4134	-0.0469			3.0133	0.4174	-5.9578	1.2015
0.0324	0.0024	0.9926	0.2005		-0.4944	-0.5156	0.0632	0.0632	0.0514	-0.0524	0.4655	-0.1302	0.0637		3.0139	0.4192	-5.9561	1.2020
0.0313	0.0029	0.9931	0.1940		-0.5241	-0.5135	0.0655	0.0655	-0.0404	0.0480	0.4126	-0.0901	-0.0440	0.1092	3.0155	0.4237	**−5.9548**	1.2026
0.0375	0.0031	0.9895	0.1563	0.0530	-0.6877										2.8948	0.3997	-5.9986	1.2047
0.0312	0.0028	0.9882	0.1485	0.0734	-0.5987	-0.1599									2.8809	0.3911	-5.9974	1.2053
0.0243	-0.0003	0.9881	0.1373	-0.0549	-0.1283	-0.7229	0.5778								2.9697	0.3440	-5.9836	1.2033
0.0372	0.0042	0.9836	0.1837	-0.0429	-0.0307	-0.8538	0.2223	0.4041							**2.8646**	**0.3369**	-5.9711	1.2017
0.0435	-0.0019	0.9847	0.1267	-0.0324	-0.0729	-0.7718	0.5324	0.0488	0.1858						2.9913	0.3542	-5.9556	1.1994
0.0264	-0.0012	0.9942	0.1146	-0.0450	-0.3413	-0.6451	0.2640	0.2623	-0.0948	0.4083					3.3200	0.5441	-5.9669	12025
0.0351	-0.0043	0.9910	0.1400	-0.0386	-0.2853	-0.6145	0.3629	0.2263	-0.0273	-0.1232	0.4016				3.1813	0.4443	-5.9546	1.2009
0.0397	-0.0026	0.9915	0.1158	-0.0389	-0.1784	-0.6823	0.1227	0.1837	-0.1866	0.0187	0.5416	0.0387			3.1638	0.4433	-5.9502	1.2008
0.0384	-0.0037	0.9892	0.1228	-0.0484	-0.1787	-0.6321	0.2304	0.2107	-0.0203	-0.0978	0.5453	-0.1836	-0.0124		3.1429	0.4306	-5.9492	1.2014
0.0437	-0.0031	0.9918	0.1426	-0.0267	-0.1214	-0.5901	0.3339	0.2772	-0.0848	0.0482	0.4872	-0.1937	-0.1847	0.1423	3.1171	0.4027	**−5.9291**	**1.1983**
0.0365	0.0012	0.9913	0.1328												3.0255	0.4641	-6.0565	1.2146
0.0390	0.0006	0.9877	0.1624		-0.7304										2.8740	**0.4066**	**−5.9812**	**1.2012**
0.0100	0.0010	0.9857	0.1405		-0.2823	-0.5512									**2.8597**	0.4077	-5.9978	1.2054
0.0252	0.0012	0.9921	0.2771		-0.3032	-0.6037	0.6670								2.9452	0.4108	-5.9924	1.2051
0.0275	-0.0003	0.9887	0.2066		-0.3756	0.0141	-0.6200	0.4885							2.8795	0.4081	-5.9896	1.2054
0.0587	0.0002	0.9918	0.2112		-0.4566	0.5721	-0.2289	-0.3144	0.1983						3.0198	0.5003	-6.0239	1.2131
0.0447	0.0011	0.9923	0.2544		-0.3370	0.0555	0.0022	-0.2962	0.7125	-0.3511					2.9495	0.4678	-6.0164	1.2124
0.0320	0.0002	0.9916	0.2630		-0.1106	-0.0876	0.3166	-0.6991	0.4422	-0.1496	0.0865				2.9452	0.4728	-6.0287	1.2157
0.0595	0.0006	0.9921	0.1747		-0.3660	0.0774	0.2795	0.1265	-0.1946	0.2173	-0.3504	0.1656			3.0658	0.4751	-6.0433	1.2194
0.0322	-0.0000	0.9905	0.2142		-0.1831	-0.0915	-0.0509	0.3514	-0.0537	-0.2553	0.1888	0.3884	-0.5529		3.0865	0.5239	-6.0331	1.2182
0.0328	0.0006	0.9901	0.1748		-0.2892	0.3235	-0.1289	0.1215	-0.5111	0.0548	0.0055	-0.1732	0.3145	-0.0898	3.0363	0.4965	-6.0314	1.2187

Note: MSE = mean squared error, LogL = log-likelihood and BIC = Bayesian information criterion. Given are the parameter estimates based on the first 4013 data points. These parameters are used to predict the volatility of the last 1000 data points and then used to calculate the MSE with the squared returns and realised volatility used as 'true' volatility. For the particle filter the amount of particles is $N = 1000$.

Table 9: Estimation and volatility forecasting results for the S&P 500 with the Bellman filter (top and middle panel) and the particle filter (bottom panel).

μ	c	φ	κ	ρ_{-1}	ρ_0	ρ_1	ρ_2	ρ_3	ρ_4	ρ_5	ρ_6	ρ_7	ρ_8	ρ_9	MSE y^2	MSE rv	LogL (*10^3)	BIC (*10^4)
0.0605	0.0057	0.9815	0.1956												2.0137	0.8515	-5.4466	1.0926
0.0574	-0.0002	0.9875	0.1839		-0.8209										1.5367	0.3237	-5.3686	1.0787
0.0385	-0.0036	0.9795	0.1811		-0.4960	-0.5306									1.5365	0.3251	-5.3190	1.0754
0.0358	-0.0029	0.9809	0.2463		-0.3576	-0.7141	0.3537								1.5286	0.3090	-5.3421	1.0742
0.0279	0.0003	0.9841	0.2523		-0.3712	-0.6791	-0.0420	0.4626							1.5314	0.3102	-5.3323	1.0731
0.0287	0.0001	0.9848	0.2534		-0.3739	-0.7446	0.0458	0.3414	0.1208						1.5317	0.3101	-5.3291	1.0733
0.0333	-0.0011	0.9870	0.2742		-0.3472	-0.6865	-0.0170	0.3920	-0.1696	0.3118					**1.5316**	**0.3067**	-5.3162	**1.0715**
0.0352	-0.0018	0.9876	0.2671		-0.3554	-0.7148	-0.0025	0.3785	-0.1577	0.2416	0.0920				1.5319	0.3076	-5.3135	1.0718
0.0371	-0.0019	0.9890	0.2720		-0.3599	-0.6803	-0.0182	0.3791	-0.1703	0.2478	-0.1050	0.2288			1.5324	0.3071	-5.3080	1.0716
0.0334	-0.0008	0.9888	0.2707		-0.3648	-0.6803	-0.0276	0.3821	-0.1644	0.2519	-0.0944	0.2126	0.0032		1.5330	0.3079	-5.3074	1.0723
0.0350	-0.0012	0.9893	0.2753		-0.3563	-0.6693	-0.0168	0.3719	-0.1721	0.2500	-0.1059	0.2263	-0.1090	0.1250	1.5330	0.3071	**−5.3043**	1.0725
0.0561	0.0002	0.9872	0.1934	0.0608	-0.8040										1.4618	0.2974	-5.3707	1.0791
0.0386	-0.0050	0.9798	0.1839	0.1881	-0.4657	-0.5077									**1.4342**	0.2780	-5.3516	1.0761
0.0426	-0.0059	0.9813	0.1925	0.1615	-0.4251	-0.5920	0.2031								1.4390	0.2712	-5.3431	1.0753
0.0398	-0.0042	0.9836	0.1880	0.1294	-0.4358	-0.5310	-0.1139	0.36878							1.4479	**0.2696**	-5.3354	1.0746
0.0396	-0.0039	0.9841	0.1859	0.1244	-0.4361	-0.5799	-0.0840	0.2577	0.1487						1.4467	0.2699	5.3332	1.0749
0.0343	-0.0076	0.9829	0.1701	-0.0390	-0.2009	-0.6272	-0.0904	0.4032	-0.2928	0.3947					1.4917	0.3094	-5.3171	1.0725
0.0358	-0.0100	0.9835	0.1363	-0.0725	-0.2007	-0.7595	-0.1254	0.3535	-0.0771	0.2504	0.1431				1.5138	0.3252	**−5.3077**	**1.0715**
0.0426	-0.0085	0.9909	0.1166	-0.0619	-0.3314	-0.6775	-0.1226	0.3429	-0.0693	0.2038	-0.0268	0.2852			1.6042	0.4040	-5.3091	1.0726
0.0506	-0.0097	0.9891	0.1400	-0.0547	-0.3548	-0.6540	-0.0539	0.4135	-0.0788	0.1782	-0.0785	0.2201	0.0388		1.5804	0.3820	-5.3106	1.0737
0.0450	-0.0070	0.9910	0.1318	-0.0315	-0.4132	-0.6250	-0.0612	0.3693	-0.0768	0.1581	-0.0846	0.2038	-0.0246	0.1891	1.5961	0.3985	-5.3099	1.0744
0.0559	-0.0030	0.9836	0.1745												1.5269	0.3239	-5.4633	1.0960
0.0573	-0.0018	0.9834	0.2021		-0.7840										1.4436	0.2881	-5.3533	1.0748
0.0408	-0.0040	0.9789	0.1874		-0.5967	-0.4347									**1.4342**	0.2909	-5.3391	1.0728
0.0437	-0.0031	0.9804	0.2265		-0.5600	-0.5506	0.2932								1.4404	0.2874	**−5.3345**	**1.0727**
0.0296	-0.0044	0.9775	0.1959		-0.4218	-0.6503	0.0431	0.02445							1.4370	0.2945	-5.3403	1.0747
0.0344	-0.0036	0.9838	0.2524		-0.3147	-0.0938	-0.7347	0.3058	0.2706						1.4471	0.2964	-5.3512	1.0777
0.0534	-0.0019	0.9902	0.2369		-0.1978	-0.0267	-0.1126	-0.3370	-0.0651	0.7161					1.4748	0.3154	-5.4220	1.0927
0.0277	-0.0049	0.9834	0.2083		-0.2009	-0.3043	-0.1355	0.0192	-0.6362	0.2711	0.3240				1.4419	0.3001	-5.3689	1.0829
0.0523	-0.0002	0.9900	0.2423		-0.6549	-0.2556	0.0233	0.1484	0.3256	-0.2608	0.0420	0.2916			1.4730	**0.2797**	-5.3478	1.0803
0.0387	-0.0014	0.9931	0.1839		-0.3407	-0.4583	-0.2208	0.0810	-0.0495	0.1138	0.2103	0.1352	0.4413		1.4570	0.2830	-5.3734	1.0863
0.0796	0.0001	0.9895	0.2071		-0.5350	0.1975	0.0197	0.1278	-0.0678	0.0011	-0.0512	0.2626	0.0746	0.0500	1.5227	0.3093	-5.4241	1.0973

Note: MSE = mean squared error, LogL = log-likelihood and BIC = Bayesian information criterion. Data used is 100× the log returns of the S&P 500. Given are the parameter estimates based on the first 4013 data points. These parameters are used to predict the volatility and then used to calculate the MSE with the squared returns and realised volatility. For the particle filter the amount of particles is $N = 1000$.

8 Conclusion

In this paper, we investigate the different leverage effects present in financial returns and test which SV model can predict volatility the best. The log returns are modelled using several different stochastic volatility models with different leverage timings and causalities. These models range from a simple SV model without a leverage effect, the asymmetric models of Jacquier et al. (2004) and Harvey and Shephard (1996) who respectively model the leverage effect as contemporaneous and inter-temporal, to the model of Catania (2022) that includes a contemporaneous relation in addition to several inter-temporal lags. These models all model volatility shocks to be a function of past and/or current return shocks. We expand on this by arguing that there can also exist a relation between a future return shock and the current volatility shock and incorporate this into a SV model by extending the model of Catania (2022).

We estimate the parameters of the models with the newly developed Bellman filter of Lange (2020) and compare it to the estimates obtained with the CSIR method used by Catania (2022) originally. We perform an extensive simulation study to evaluate the accuracy of the parameter estimation and log-volatility prediction for different SV models. In the case of model misspecification the estimated leverage effects are not stable and can get estimated to have big values on the wrong timing if the true timing of the leverage effect is excluded from the model. We find that the Bellman filter does provide similar of more accurate estimates than the particle filter for the median of the returns and for the level and persistence of the log-volatility. Both methods underestimate the variance of the volatility shocks with a large margin. The Bellman filter does estimate the leverage parameters much more accurate than the particle filter in the case of both correctly specified and misspecified models. The particle filter leverage effect estimates are so poor that for log-volatility forecasting the model outperforms the correctly specified models when leverage effects are present in the data.

Finally, index return data of the S&P 500 and Nasdaq are used to estimate the parameters for ten different SV models and we evaluate what model provides the best volatility estimates, where the true volatility is proxied with squared returns and realised volatility. The Bellman filter and particle filter are used to estimate the basic SV model without leverage effects and the model of Catania (2022) with several leverage effects. The Bellman filter is also used to estimate the extended model with different leverage effects. Both indices do show a presence of the volatility feedback effect and several leverage effects.

The models selected using the BIC and from a volatility prediction point of view have no clear conclusion on the optimal amount of inter-temporal leverage effects to include in the model. For both data sets it is found that generally the extended model can forecast the volatility the most accurate when either squared returns or realised volatility is used as a proxy for the "true" volatility. While the relationship between the current volatility shock and the future return shock is found to be small it does increase volatility forecasts and is therefore worth incorporating into SV models that model financial log returns.

This research is still limited in the sense that it only investigates US indices. Estimation of the models on other indices, individual stock returns or exchange rate data is needed to confirm if the result found in this paper holds for different data. We also use a relatively small amount of particles to estimate the parameters with the particle filter to relieve the computational burden of the estimation process. It would be interesting to investigate if the same difference in forecasting performance is found when the amount of particles used is increased.

Future research may also include adding more than one extra leverage effect to the model. Including the relation between tomorrow's return shock en today's volatility shock increases the volatility predictions, perhaps incorporating a relation between the day after tomorrow's return shock and today's volatility shock will do the same.

References

Andersen, T. G., Bollerslev, T., Diebold, F. X., & Ebens, H. (2001). The distribution of realized stock return volatility. *Journal of Financial Economics*, *61*(1), 43–76.

Black, F. (1976). Studies of stock market volatility changes. *1976 Proceedings of the American Statistical Association Bisiness and Economic Statistics Section*.

Carr, P., & Wu, L. (2017). Leverage effect, volatility feedback, and self-exciting market disruptions. *Journal of Financial and Quantitative Analysis*, *52*(5), 2119–2156.

Catania, L. (2022). A Stochastic Volatility Model with a General Leverage specification. *Journal of Business & Economic Statistics*, *40*(2), 678-689.

Ghysels, E., Harvey, A. C., & Renault, E. (1996). Stochastic volatility. *Handbook of statistics*, *14*, 119–191.

Gordon, N. J., Salmond, D. J., & Smith, A. F. (1993). Novel approach to nonlinear/non-Gaussian Bayesian state estimation. *IEE proceedings F (radar and signal processing)*, *140*(2), 107–113.

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.

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, 03). A New Approach to Linear Filtering and Prediction Problems. *Journal of Basic Engineering*, *82*(1), 35-45.

Lange, R.-J. (2020). Bellman filtering for state-space models. *Tinbergen Institute Discussion Paper 2020-052/III*.

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

Nocedal, J., & Wright, S. J. (1999). *Numerical optimization.* Springer.

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

Teulings, C., Lange, R.-J., & van der Kroft, B. (2020, 04). Piek in volatiliteit laat zien dat groot deel daling aandelenkoersen tijdelijk is. *Economisch-statistische Berichten*, *105*, 153-155.

A Optimisation and updating step for the SV model of Catania (2020)

Since $\boldsymbol{a}_t$ and $\boldsymbol{a}_{t-1}$ share $m-1$ elements $\begin{bmatrix} \boldsymbol{a}_t \\ \boldsymbol{a}_{t-1} \end{bmatrix}$ is equal to $\begin{bmatrix} \boldsymbol{a}_t \\ h_{t-m-1} \end{bmatrix}$.
The optimisation and updating steps for the model of equation (7) will be

$$\begin{bmatrix} \boldsymbol{a}_t \\ h_{t-m-1} \end{bmatrix} \leftarrow \begin{bmatrix} \boldsymbol{a}_t \\ h_{t-m-1} \end{bmatrix} + \left[\begin{pmatrix} \boldsymbol{J}_t^{11} - \frac{1}{2}\frac{\mathrm{d}^2\ell(y_t|\boldsymbol{a})}{\mathrm{d}\boldsymbol{a}\mathrm{d}\boldsymbol{a}'} - \frac{1}{2}E\left(\frac{\mathrm{d}^2\ell(y_t|\boldsymbol{a})}{\mathrm{d}\boldsymbol{a}\mathrm{d}\boldsymbol{a}'}\right) & \mathbf{0}_{(m+1)\times 1} \\ \mathbf{0}_{1\times(m+1)} & 0 \end{pmatrix} + \begin{pmatrix} 0 & \mathbf{0}_{1\times(m+1)} \\ \mathbf{0}_{(m+1)\times 1} & \boldsymbol{I}_{t-1|t-1} \end{pmatrix} \right]^{-1} \left[\begin{pmatrix} \boldsymbol{J}_t^1 + \frac{\mathrm{d}\ell(y_t|\boldsymbol{a}_t)}{\mathrm{d}\boldsymbol{a}_t} \\ 0 \end{pmatrix} - \begin{pmatrix} 0 \\ \boldsymbol{I}_{t|t-1}(\boldsymbol{a}_{t-1} - \boldsymbol{a}_{t-1|t-1}) \end{pmatrix} \right], \tag{62}$$

with $\mathbf{0}_{a\times b}$ a matrix of zeros with dimension $a \times b$,

$$\boldsymbol{J}_t^1 = \frac{h_t - \mu_{h,t}}{\sigma_{h,t}^2} \begin{bmatrix} -1 \\ \varphi - \frac{\kappa}{2}\rho_1 \frac{y_{t-1}-\mu}{\exp(h_{t-1}/2)} \\ -\frac{\kappa}{2}\rho_2 \frac{y_{t-2}-\mu}{\exp(h_{t-2}/2)} \\ \vdots \\ -\frac{\kappa}{2}\rho_m \frac{y_{t-m}-\mu}{\exp(h_{t-m}/2)} \end{bmatrix} =: \frac{h_t - \mu_{h,t}}{\sigma_{h,t}^2}\boldsymbol{c}_t, \tag{63}$$

$$\boldsymbol{J}_t^{11} = \frac{-1}{\sigma_{h,t}^2}\boldsymbol{c}_t\boldsymbol{c}_t' + \frac{h_t - \mu_{h,t}}{\sigma_{h,t}^2}\frac{\kappa}{4}\operatorname{diag}\begin{bmatrix} 0 \\ \rho_1 \frac{y_{t-1}-\mu}{\exp(h_{t-1}/2)} \\ \rho_2 \frac{y_t-2-\mu}{\exp(h_{t-2}/2)} \\ \vdots \\ \rho_m \frac{y_{t-m}-\mu}{\exp(h_{t-m}/2)} \end{bmatrix},$$

$$\begin{aligned} \frac{\mathrm{d}\ell\left(y_t \mid \boldsymbol{a}_t, \mathcal{F}_{t-1}\right)}{\mathrm{d}\boldsymbol{a}_t} &= \frac{y_t - \mu_{y,t}}{\sigma_{y,t}^2}\left(\begin{bmatrix} (\mu_{y,t}-\mu)/2 \\ 0 \\ \vdots \\ 0 \end{bmatrix} + \frac{\rho_0 \exp(h_t/2)}{1 - \sum_{j=1}^m \rho_j^2}\begin{bmatrix} 1/\kappa \\ -\varphi/\kappa + \rho_1/2\frac{y_{t-1}-\mu}{\exp h_{t-1}/2} \\ \rho_2/2\frac{y_{t-2}-\mu}{\exp h_{t-2}/2} \\ \vdots \\ \rho_m/2\frac{y_{t-m}-\mu}{\exp h_{t-m}/2} \end{bmatrix}\right) + \left(\frac{(y_t-\mu_{y,t})^2}{\sigma_{y,t}^3} - \frac{1}{\sigma_{y,t}}\right)\begin{bmatrix} \sigma_{y,t}/2 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \\ &=: \frac{y_t - \mu_{y,t}}{\sigma_{y,t}^2}\left(\begin{bmatrix} (\mu_{y,t}-\mu)/2 \\ 0 \\ \vdots \\ 0 \end{bmatrix} + \boldsymbol{b}_t\right) + \left(\frac{(y_t-\mu_{y,t})^2}{\sigma_{y,t}^3} - \frac{1}{\sigma_{y,t}}\right)\begin{bmatrix} \sigma_{y,t}/2 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \end{aligned} \tag{64}$$

and

$$
\begin{aligned}
\frac{\mathrm{d}^2\ell\left(y_t \mid \boldsymbol{a}_t, \mathcal{F}_{t-1}\right)}{\mathrm{d}\boldsymbol{a}_t \mathrm{d}\boldsymbol{a}_t'} = & \frac{-1}{\sigma_{y,t}^2}\left(\begin{bmatrix} (\mu_{y,t}-\mu)/2 \\ 0 \\ \vdots \\ 0 \end{bmatrix} + \boldsymbol{b}_t\right)\left(\begin{bmatrix} (\mu_{y,t}-\mu)/2 \\ 0 \\ \vdots \\ 0 \end{bmatrix} + \boldsymbol{b}_t\right)' \\
& + \left(\frac{1}{\sigma_{y,t}^2} - \frac{3(y_t-\mu_{y,t})^2}{\sigma_{y,t}^4}\right)\begin{bmatrix} \sigma_{y,t}/2 \\ 0 \\ \vdots \\ 0 \end{bmatrix}\begin{bmatrix} \sigma_{y,t}/2 \\ 0 \\ \vdots \\ 0 \end{bmatrix}' \\
& - 2\frac{y_t-\mu_{y,t}}{\sigma_{y,t}^3}\left(\begin{bmatrix} (\mu_{y,t}-\mu)/2 \\ 0 \\ \vdots \\ 0 \end{bmatrix} + \boldsymbol{b}_t\right)\begin{bmatrix} \sigma_{y,t}/2 \\ 0 \\ \vdots \\ 0 \end{bmatrix}' - 2\frac{y_t-\mu_{y,t}}{\sigma_{y,t}^3}\begin{bmatrix} \sigma_{y,t}/2 \\ 0 \\ \vdots \\ 0 \end{bmatrix}\left(\begin{bmatrix} (\mu_{y,t}-\mu)/2 \\ 0 \\ \vdots \\ 0 \end{bmatrix} + \boldsymbol{b}_t\right)' \\
& + \frac{y_t-\mu_{y,t}}{\sigma_{y,t}^2}\times \\
& \left(\operatorname{diag}\begin{bmatrix} (\mu_{y,t}-\mu)/4 \\ 0 \\ \vdots \\ 0 \end{bmatrix} - \frac{1}{4}\frac{\rho_0 \exp\left(h_t/2\right)}{1-\sum_{j=1}^m \rho_j^2}\operatorname{diag}\begin{bmatrix} 0 \\ \rho_1 \frac{y_{t-1}-\mu}{\exp(h_{t-1}/2)} \\ \vdots \\ \rho_m \frac{y_{t-m}-\mu}{\exp(h_{t-m}/2)} \end{bmatrix} + \begin{bmatrix} 1/2 \\ 0 \\ \vdots \\ 0 \end{bmatrix}\boldsymbol{b}_t' + \boldsymbol{b}_t\begin{bmatrix} 1/2 \\ 0 \\ \vdots \\ 0 \end{bmatrix}'\right) \\
& + \left(\frac{(y_t-\mu_{y,t})^2}{\sigma_{y,t}^3} - \frac{1}{\sigma_{y,t}}\right)\operatorname{diag}\begin{bmatrix} \sigma_{y,t}/4 \\ 0 \\ \vdots \\ 0 \end{bmatrix}.
\end{aligned} \tag{65}
$$

For the updating of the information matrix $\boldsymbol{I}_{t|t}$ we will take the Schur complement of the bottom-right block of the negative Hessian with size $(m+1)\times(m+1)$ evaluated at $\boldsymbol{a}_{t|t}$

$$
\begin{pmatrix} \boldsymbol{J}_t^{11} - \frac{1}{2}\frac{\mathrm{d}^2\ell(y_t|\boldsymbol{a})}{\mathrm{d}\boldsymbol{a}\mathrm{d}\boldsymbol{a}'} - \frac{1}{2}\left(E\frac{\mathrm{d}^2\ell(y_t|\boldsymbol{a})}{\mathrm{d}\boldsymbol{a}\mathrm{d}\boldsymbol{a}'}\right) & \boldsymbol{0}_{(m+1)\times 1} \\ \boldsymbol{0}_{1\times(m+1)} & 0 \end{pmatrix} + \begin{pmatrix} 0 & \boldsymbol{0}_{1\times(m+1)} \\ \boldsymbol{0}_{(m+1)\times 1} & \boldsymbol{I}_{t-1|t-1} \end{pmatrix}, \tag{66}
$$

B Optimisation and updating step for the extended SV model

For the model in equation (10) the optimisation and updating steps will be different as now the first element of the state $\boldsymbol{a}_t$ contains the return shock ε_{t+1} and the first element of $\boldsymbol{a}_{t-1}$ is ε_t, the combined vector of the states is no longer $\begin{bmatrix} \boldsymbol{a}_t \\ h_{t-m-1} \end{bmatrix}$. With some creative rearranging where the first element of $\boldsymbol{a}_{t-1}$ is removed and placed at the end of $\boldsymbol{a}_{t-1}$ to get $\boldsymbol{a}^*_{t-1} = h_{t-1}, \ldots, h_{t-m-1}, \varepsilon_t$ they can be placed such that the combined vector will be $\begin{bmatrix} \boldsymbol{a}_t \\ h_{t-m-1} \\ \varepsilon_t \end{bmatrix}$. However $\boldsymbol{I}_{t-1|t-1}$ then also needs to be rearranged to match the new form of $\boldsymbol{a}^*_{t-1}$ by removing the first row and column and adding the transposed version of them at the end of the matrix to get $\boldsymbol{I}^*_{t-1|t-1}$. For the gradient this process is repeated for both $\boldsymbol{a}_{t-1|t-1}$ and $\boldsymbol{I}_{t|t-1}$ to obtain $\boldsymbol{a}^*_{t-1|t-1}$ and $\boldsymbol{I}^*_{t|t-1}$.

In the original model we need to use the degenerate extension of the Bellman filter as the elements of $\boldsymbol{a}_t$ apart from h_t are deterministic functions of the past state $\boldsymbol{a}_{t-1}$ but this is not the case now for ε_t. Adjusting the optimisation step for this will result in

$$\begin{bmatrix} \boldsymbol{a}_t \\ h_{t-m-1} \\ \varepsilon_t \end{bmatrix} \leftarrow \begin{bmatrix} \boldsymbol{a}_t \\ h_{t-m-1} \\ \varepsilon_t \end{bmatrix} + \left[\begin{pmatrix} \boldsymbol{J}^{11}_t - \frac{1}{2}\frac{\mathrm{d}^2\ell(y_t|\boldsymbol{a})}{\mathrm{d}\boldsymbol{a}\mathrm{d}\boldsymbol{a}'} - \frac{1}{2}E\left(\frac{\mathrm{d}^2\ell(y_t|\boldsymbol{a})}{\mathrm{d}\boldsymbol{a}\mathrm{d}\boldsymbol{a}'}\right) & \mathbf{0}_{(m+2)\times 1} & \boldsymbol{J}^{12}_t \\ \mathbf{0}'_{(m+2)\times 1} & 0 & 0 \\ \boldsymbol{J}^{21}_t & 0 & J^{22}_t \end{pmatrix} + \begin{pmatrix} \mathbf{0}_{2\times 2} & \mathbf{0}'_{(m+2)\times 2} \\ \mathbf{0}_{(m+2)\times 2} & \boldsymbol{I}_{t-1|t-1} \end{pmatrix} \right]^{-1}$$
$$\left[\begin{pmatrix} \boldsymbol{J}^1_t + \frac{\mathrm{d}\ell(y_t|\boldsymbol{a}_t)}{\mathrm{d}\boldsymbol{a}_t} \\ 0 \\ J^2_t \end{pmatrix} - \begin{pmatrix} 0 \\ 0 \\ \boldsymbol{I}^*_{t|t-1}(\boldsymbol{a}^*_{t-1} - \boldsymbol{a}^*_{t-1|t-1}) \end{pmatrix} \right], \tag{67}$$

with $\mathbf{0}_{a\times b}$ a matrix of zeros with dimensions $a \times b$.

The first and second derivatives of the state transition density with respect to the state $\boldsymbol{a}_t$ will now be

$$\boldsymbol{J}_t^1 = \frac{h_t - \mu_{h,t}^*}{\sigma_{h,t}^{2*}} \begin{bmatrix} 0 \\ -1 \\ \varphi - \frac{\kappa}{2}\rho_1 \frac{y_{t-1}-\mu}{\exp(h_{t-1}/2)} \\ -\frac{\kappa}{2}\rho_2 \frac{y_{t-2}-\mu}{\exp(h_{t-2}/2)} \\ \vdots \\ -\frac{\kappa}{2}\rho_m \frac{y_{t-m}-\mu}{\exp(h_{t-m}/2)} \end{bmatrix} + \frac{\varepsilon_{t+1} - \mu_{\varepsilon,t+1}}{\sigma_{\varepsilon,t+1}^2} \left(\frac{\rho_{-1}}{1 - \sum_{j=0}^m \rho_j^2} \begin{bmatrix} 0 \\ 1/\kappa \\ -\varphi/\kappa + \frac{\rho_1}{2} \frac{y_{t-1}-\mu}{\exp(h_{t-1}/2)} \\ \frac{\rho_2}{2} \frac{y_{t-2}-\mu}{\exp(h_{t-2}/2)} \\ \vdots \\ \frac{\rho_m}{2} \frac{y_{t-m}-\mu}{\exp(h_{t-m}/2)} \end{bmatrix} + \begin{bmatrix} -1 \\ 0 \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \right)$$

$$=: \frac{h_t - \mu_{h,t}^*}{\sigma_{h,t}^{2*}} \boldsymbol{c}_t^* + \frac{\varepsilon_{t+1} - \mu_{\varepsilon,t+1}}{\sigma_{\varepsilon,t+1}^2} \boldsymbol{d}_t^*,$$

$$J_t^2 = \frac{h_t - \mu_{h,t}^*}{\sigma_{h,t}^{2*}} \kappa \frac{\rho_0}{1 - \sum_{j=1}^m \rho_j^2} - \frac{\varepsilon_{t+1} - \mu_{\varepsilon,t+1}}{\sigma_{\varepsilon,t+1}^2} \frac{\rho_{-1}\rho_0}{1 - \sum_{j=0}^m \rho_j^2},$$

$$\boldsymbol{J}_t^{11} = \frac{-1}{\sigma_{h,t}^{2*}} \boldsymbol{c}_t^* \boldsymbol{c}_t^{*\prime} + \frac{h_t - \mu_{h,t}^*}{\sigma_{h,t}^{2*}} \frac{\kappa}{4} \operatorname{diag} \begin{bmatrix} 0 \\ 0 \\ \rho_1 \frac{y_{t-1}-\mu}{\exp(h_{t-1}/2)} \\ \vdots \\ \rho_m \frac{y_{t-m}-\mu}{\exp(h_{t-m}/2)} \end{bmatrix} +$$

$$\frac{-1}{\sigma_{\varepsilon,t+1}^2} \boldsymbol{d}_t^* \boldsymbol{d}_t^{*\prime} + \frac{\varepsilon_{t+1} - \mu_{\varepsilon,t+1}}{\sigma_{\varepsilon,t+1}^2} \frac{\rho_{-1}}{1 - \sum_{j=1}^m \rho_j^2} \frac{1}{4} \operatorname{diag} \begin{bmatrix} 0 \\ 0 \\ \rho_1 \frac{y_{t-1}-\mu}{\exp(h_{t-1}/2)} \\ \vdots \\ \rho_m \frac{y_{t-m}-\mu}{\exp(h_{t-m}/2)} \end{bmatrix},$$

$$\boldsymbol{J}_t^{12} = \frac{-1}{\sigma_{h,t}^{2*}} \kappa \frac{\rho_0}{1 - \sum_{j=1}^m \rho_j^2} \boldsymbol{c}_t^* + \frac{-1}{\sigma_{\varepsilon,t+1}^{2*}} \frac{\rho_{-1}\rho_0}{1 - \sum_{j=0}^m \rho_j^2} \boldsymbol{d}_t^*,$$

$$\boldsymbol{J}_t^{21} = \frac{-1}{\sigma_{h,t}^{2*}} \kappa \frac{\rho_0}{1 - \sum_{j=1}^m \rho_j^2} \boldsymbol{c}_t^{*\prime} + \frac{-1}{\sigma_{\varepsilon,t+1}^{2*}} \frac{\rho_{-1}\rho_0}{1 - \sum_{j=0}^m \rho_j^2} \boldsymbol{d}_t^{*\prime},$$

$$J_t^{22} = \frac{-1}{\sigma_{h,t}^{2*}} \kappa^2 \frac{\rho_0^2}{(1 - \sum_{j=1}^m \rho_j^2)^2} + \frac{-1}{\sigma_{\varepsilon,t+1}^{2*}} \frac{(\rho_{-1}\rho_0)^2}{(1 - \sum_{j=0}^m \rho_j^2)^2}, \tag{68}$$

and the derivatives of $p(y_t \mid \boldsymbol{a}_t, \mathcal{F}_{t-1})$ with respect to $\boldsymbol{a}_t$ are

$$\begin{aligned}
\frac{\mathrm{d}\ell\left(y_t \mid \boldsymbol{a}_t, \mathcal{F}_{t-1}\right)}{\mathrm{d}\boldsymbol{a}_t} =& \frac{y_t-\mu^*_{y,t}}{\sigma^{2*}_{y,t}} \left(\begin{bmatrix} 0 \\ (\mu^*_{y,t}-\mu)/2 \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} + \frac{\rho_0 \exp(h_t/2)}{1-\sum_{j=1}^m \rho_j^2 - \rho_{-1}^2} \begin{bmatrix} -\rho_{-1} \\ 1/\kappa \\ -\varphi/\kappa + \rho_1/2 \frac{y_{t-1}-\mu}{\exp(h_{t-1}/2)} \\ \rho_2/2 \frac{y_{t-2}-\mu}{\exp(h_{t-2}/2)} \\ \vdots \\ \rho_m/2 \frac{y_{t-m}-\mu}{\exp(h_{t-m}/2)} \end{bmatrix} \right) + \left(\frac{(y_t-\mu^*_{y,t})^2}{\sigma^{3*}_{y,t}} - \frac{1}{\sigma^*_{y,t}} \right) \begin{bmatrix} 0 \\ \sigma^*_{y,t}/2 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \\
=:& \frac{y_t-\mu^*_{y,t}}{\sigma^{2*}_{y,t}} \left(\begin{bmatrix} 0 \\ (\mu^*_{y,t}-\mu)/2 \\ 0 \\ \vdots \\ 0 \end{bmatrix} + \boldsymbol{b}^*_t \right) + \left(\frac{(y_t-\mu^*_{y,t})^2}{\sigma^{3*}_{y,t}} - \frac{1}{\sigma^*_{y,t}} \right) \begin{bmatrix} 0 \\ \sigma^*_{y,t}/2 \\ 0 \\ \vdots \\ 0 \end{bmatrix},
\end{aligned} \tag{69}$$

and

$$\begin{aligned}
\frac{\mathrm{d}^2\ell\left(y_t \mid \boldsymbol{a}, \mathcal{F}_{t-1}\right)}{\mathrm{d}\boldsymbol{a}\mathrm{d}\boldsymbol{a}'} =& \frac{-1}{\sigma^{2*}_{y,t}} \left(\begin{bmatrix} 0 \\ (\mu^*_{y,t}-\mu)/2 \\ 0 \\ \vdots \\ 0 \end{bmatrix} + \boldsymbol{b}^*_t \right) \left(\begin{bmatrix} 0 \\ (\mu^*_{y,t}-\mu)/2 \\ 0 \\ \vdots \\ 0 \end{bmatrix} + \boldsymbol{b}^*_t \right)' + \left(\frac{1}{\sigma^{2*}_{y,t}} - \frac{3(y_t-\mu^*_{y,t})^2}{\sigma^{4*}_{y,t}} \right) \begin{bmatrix} 0 \\ \sigma^*_{y,t}/2 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \begin{bmatrix} 0 \\ \sigma^*_{y,t}/2 \\ 0 \\ \vdots \\ 0 \end{bmatrix}' \\
& - 2\frac{y_t-\mu^*_{y,t}}{\sigma^{3*}_{y,t}} \left(\begin{bmatrix} 0 \\ (\mu^*_{y,t}-\mu)/2 \\ 0 \\ \vdots \\ 0 \end{bmatrix} + \boldsymbol{b}^*_t \right) \begin{bmatrix} 0 \\ \sigma^*_{y,t}/2 \\ 0 \\ \vdots \\ 0 \end{bmatrix}' - 2\frac{y_t-\mu^*_{y,t}}{\sigma^{3*}_{y,t}} \begin{bmatrix} 0 \\ \sigma^*_{y,t}/2 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \left(\begin{bmatrix} 0 \\ (\mu^*_{y,t}-\mu)/2 \\ 0 \\ \vdots \\ 0 \end{bmatrix} + \boldsymbol{b}^*_t \right)' \\
& + \frac{y_t-\mu^*_{y,t}}{\sigma^{2*}_{y,t}} \times \\
& \left(\operatorname{diag} \begin{bmatrix} 0 \\ \left(\mu^*_{y,t}-\mu\right)/4 \\ 0 \\ \vdots \\ 0 \end{bmatrix} - \frac{1}{4} \frac{\rho_0 \exp\left(h_t/2\right)}{1-\sum_{j=1}^k \rho_j^2} \operatorname{diag} \begin{bmatrix} 0 \\ 0 \\ \rho_1 \frac{y_{t-1}-\mu}{\exp(h_{t-1}/2)} \\ \vdots \\ \rho_m \frac{y_{t-m}-\mu}{\exp(h_{t-m}/2)} \end{bmatrix} + \begin{bmatrix} 0 \\ 1/2 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \boldsymbol{b}^{*\prime}_t + \boldsymbol{b}^*_t \begin{bmatrix} 0 \\ 1/2 \\ 0 \\ \vdots \\ 0 \end{bmatrix}' \right) \\
& + \left(\frac{(y_t-\mu^*_{y,t})^2}{\sigma^{3*}_{y,t}} - \frac{1}{\sigma^*_{y,t}} \right) \operatorname{diag} \begin{bmatrix} 0 \\ \sigma^*_{y,t}/4 \\ 0 \\ \vdots \\ 0 \end{bmatrix}.
\end{aligned} \tag{70}$$

For the updating of the information matrix $\boldsymbol{I}_{t|t}$ we will take the Schur complement of the bottom-right block of the negative Hessian with size $m \times m$

$$\begin{pmatrix} \boldsymbol{J}_t^{11} - \frac{1}{2}\frac{\mathrm{d}^2\ell(y_t|\boldsymbol{a})}{\mathrm{d}\boldsymbol{a}\mathrm{d}\boldsymbol{a}'} - \frac{1}{2}E\left(\frac{\mathrm{d}^2\ell(y_t|\boldsymbol{a})}{\mathrm{d}\boldsymbol{a}\mathrm{d}\boldsymbol{a}'}\right) & \mathbf{0}_{m+2\times 1} & \boldsymbol{J}_t^{12} \\ \mathbf{0}'_{m+2\times 1} & 0 & 0 \\ \boldsymbol{J}_t^{21} & 0 & J_t^{22} \end{pmatrix} + \begin{pmatrix} \mathbf{0}_{2\times 2} & \mathbf{0}'_{m+2\times 2} \\ \mathbf{0}_{m+2\times 2} & \boldsymbol{I}^*_{t-1|t-1} \end{pmatrix} \tag{71}$$