Erasmus School of Economics

Thesis Msc Econometrics and Management Science

Is the leverage effect modelled correctly?

Author:
W.F. Rutgers, 431921

Supervisor:
Dr. R. Lange

Second assessor:
Prof. Dr. D.J.C. van Dijk

May 1, 2021

Abstract

This paper discusses the modelling of the leverage effect in financial returns. The leverage effect is a well-known feature in financial returns, but the cause remains unclear, and a clear explanation is absent. The general assumption is that volatility shocks depend on past return shocks; we argue that both the timing and causality of this assumption are wrong. Firstly, it makes more sense to model the leverage effect contemporaneously, as we do not expect the financial market to react with a delay. Second, it is well known that the price and return of an asset depend on its risk. Therefore, we model the return shock as a function of the volatility shock instead of the contrary. With an extensive simulation study, we show that, while return shocks do not seem to be a function of volatility shocks, there is also evidence against the traditional modelling of the leverage effect. Finally, we estimate various stochastic volatility models, and other types of state-space models, using both the new Bellman filter by Lange (2020) and the extended particle filter by Malik and Pitt (2011). We show that the Bellman filter performs almost as well as the particle filter while being much faster.

Keywords — volatility, stochastic volatility, leverage effect, state-space models, Bellman filter, particle filter

Contents

1 Introduction

Stochastic volatility models are widely used to describe the return and volatility process of financial returns. Stochastic volatility models allow independent innovations to both the returns and the log-volatility process, giving it an advantage over the famous ARCH models, where volatility is solely a function of past returns. First introduced in the seminal work of Taylor (1986), more sophisticated extensions to the stochastic volatility model are introduced, including the so called 'leverage effect'.

The leverage effect is widely defined as the asymmetric reaction of volatility to positive and negative returns. The timing of this effect is a subject of debate. For example, Jacquier et al. (2004) model this effect contemporaneously, while Harvey and Shephard (1996) suggest an inter-temporal leverage effect. This has lead to Yu (2005) and Catania (2020) incorporating both contemporaneous and inter-temporal leverage. Nonetheless, the inter-temporal specification is still the most used model in the literature. Using the inter-temporal specification seems counter intuitive, as financial markets these days are assumed to be highly efficient in the processing of information. One of the reasons that the inter-temporal model is often preferred is due to the characteristics of the contemporaneous model. Namely, financial returns turn out to be autocorrelated in the contemporaneous specification. This paper investigates whether it is correct to model the leverage effect inter-temporal or if this should be adjusted.

When modelling multiple leverage effects, Yu (2005) and Catania (2020) assume volatility shocks are partly explained by past and present return shocks. Catania (2020) shows that 70% to 90% of the volatility shocks of major index returns are explained by past and present return shocks. We argue that this does not seem reasonable and think that volatility shocks should contain mainly new information. To investigate this, we introduce a new model where volatility shocks are completely independent and return shocks are partly explained by present and future volatility shocks.

As stochastic volatility models consist of the observed financial returns and the non-observed volatility process, we can present them as state-space models. Still, the estimation of the stochastic volatility models is nonstandard. Due to the nonlinear relation between the volatility and observation, the regular Kalman filter cannot be used. As a solution, Harvey and Shephard (1996) suggested using log-squared returns and estimating the models with quasi maximum likelihood. Many simulated based methods also have been proposed. In this paper, we use one of these simulation methods, namely the extension of the particle filter by Malik and Pitt (2011). Furthermore, we use the new Bellman filter introduced by Lange (2020), a quick non-simulation-

based method, but rather based on dynamic programming, for filtering and estimating nonlinear and/or non-Gaussian state-space methods.

In our empirical application, we find that the leverage effect is more likely to be inter-temporal rather than contemporaneous. However, including both contemporaneous and inter-temporal leverage leads to the best volatility predictions. Comparing the empirical results to an extensive simulation study shows that the classic model specification is not entirely correct, and that the leverage structure of financial returns is different than what the stochastic volatility models assume. Suggesting that it would be better to model the leverage effect in a completely different way.

Furthermore, we find that the new Bellman filter by Lange (2020) can quickly and efficiently estimate both stochastic volatility models and other state-space models. Thus, the Bellman filter is an innovative way to model nonlinear/non-Gaussian state-space model without relying on the computationally intensive simulation-based methods.

The remainder of this paper is structured as follows. In Section 2, we discuss the stochastic volatility models that are present in literature and present our new way of modelling the leverage effect. In Section 3, we review three methods to filter and estimate state-space models. Section 4 presents the estimators of our sophisticated stochastic volatility models and an extensive simulation study on the performance of these models. In Section 5, we apply the methodology to empirical data, and Section 6 concludes.

2 Stochastic volatility models

The class of stochastic volatility models has its roots in mathematical finance and financial econometrics. Several stochastic volatility models originated from researching different issues. Clark (1973) modelled asset returns as a function of the random process of information arrival, which generated an asset return model with time-varying volatility. Tauchen and Pitts (1983) continued by using a mixture of distributions model of asset returns with temporal dependence on information arrival. Hull and White (1987) introduced a stochastic volatility model to price European options assuming continuous-time stochastic volatility models for the underlying asset. In the seminal work of Taylor (1986), the research in stochastic volatility models as used in this paper began. Taylor (1986) formulated a discrete-time stochastic volatility model as an alternative to Autoregressive Conditional Heteroskedasticity (ARCH) models. This section presents this stochastic volatility model, and shows how this model has led to many extensions incorporating the so called 'leverage effect'. Then, we introduce a new way of incorporating this leverage effect

into a stochastic volatility model.

2.1 Classic stochastic volatility model

Taylor (1986) introduced the stochastic volatility model as an alternative for the ARCH models. Discrete-time ARCH models are time series models that describe the volatility of the current return shock as a function of previous shocks. ARCH models originated in the seminal paper by Engle (1982) and were continued by Bollerslev (1986). In ARCH models, the process of volatility is fully explained by past and current financial returns, and thus both the financial returns and the volatility are subject to the same error term. The usefulness of ARCH models is that they do display conditional heteroskedasticity, something very much present in financial returns, and that the likelihood can easily be evaluated, such that ARCH models are straightforward to estimate. These reasons have led to the ARCH models being very popular and widely used. However, the major drawback of ARCH models is that the conditional volatility depends on past returns and does not itself incorporate any new information. A stochastic volatility model has this characteristic, as both the volatility and the returns follow their own stochastic process. Research has shown that this does lead to a better model specification. For example, Kim et al. (1998) provide evidence of better in-sample-fit of the basic stochastic volatility model relative to ARCH-type models. The standard stochastic volatility model is given by:

$$\begin{aligned} y_t &= \exp\left\{\frac{h_t}{2}\right\}\varepsilon_t, \quad \varepsilon_t \overset{iid}{\sim} \mathcal{N}(0,1), \\ h_t &= c + \phi h_{t-1} + \sigma_n \eta_t, \quad \eta_t \overset{iid}{\sim} \mathcal{N}(0,1). \end{aligned} \tag{1}$$

The time series y_t is the return of a financial asset and has standard Gaussian distributed shocks. As the expectation of the shock ε_t is zero, the expectation of the returns $E[y_t]$ is also zero. The conditional volatility of the financial return is given by $Var(y_t|h_t) = \sigma_{y_t}^2 = \exp(h_t)$. Thus, variable h_t is the log-volatility. Modelling the volatility process as log-volatility ensures that the volatility, and therefore the variance of y_t, is always positive. In the model, $\sigma_n > 0$ and $|\phi| < 1$ control for the variance and persistence of the volatility process, respectively. Furthermore, unrestricted parameter c controls for the level of the volatility process.[1] [2] The unconditional distribution of the log-volatility is Gaussian with mean $\frac{c}{1-\phi}$ and variance $\frac{\sigma_n^2}{1-\phi^2}$, $h_t \sim \mathcal{N}(\frac{c}{1-\phi}, \frac{\sigma_n^2}{1-\phi^2})$. [3] Inference and prediction of the model is nonstandard and was troublesome for some time, as the likelihood can not be evaluated analytically. Due to the nonlinearity of the model, the Kalman filter

[1] An alternative and interchangeable way of presenting the volatility model is using parameter β in the return equation controlling for the level of the log-volatility process.

[2] The definitions and restrictions for these parameters apply for all stochastic volatility models presented in this paper

[3] These are the unconditional mean and variance of a standard AR(1) model.

(Kalman (1960) could not be used. Taylor (1986) and Melino and Turnbull (1990) estimated the model using the method of moments (MM), to avoid the integration problems associated with evaluating the likelihood exactly. However, MM may be inefficient relative to a likelihood-based method (Pearson, 1936). Nelson (1988), Harvey et al. (1994), and Ruiz (1994) suggested using the Kalman filter after taking log-squares of the returns, after which the model can be estimated with quasi-maximum likelihood (QML). Since then, new methods have been introduced. For example, computationally intensive simulation-based methods using Bayesian analysis as in Jacquier et al. (2002) or particle filtering (Gordon et al., 1993). As these methods are computationally very intensive, the QML estimator via the Kalman filter is still a commonly used method. In this paper, we show that the new Bellman filter by Lange (2020) can quickly estimate the stochastic volatility model while not relying on any data transformations.

2.2 The financial leverage effect

The financial leverage effect is the asymmetric reaction of volatility to negative and positive financial returns. Specifically, it is often observed that volatility increases after negative financial returns. The leverage effect is an important and well-documented empirical feature in many financial time series. See, for example, Black (1976), Christie (1982), Nelson (1991), and Engle and Ng (1993). Hull and White (1987) show that option prices could be substantially biased when the leverage effect is not incorporated in a stochastic volatility model. A widely accepted explanation for the leverage effect is that bad news, which reduces the price (a negative return) and therefore increases the debt-to-equity (DE, also called financial leverage), makes the firm riskier and therefore increases future expected volatility. Nelson (1991) and Glosten et al. (1993) introduced this effect in the ARCH literature, by modelling the conditional variance as not only a function of the size, but also as a function of the sign of the previous return. Harvey and Shephard (1996) were the first to incorporate the leverage effect in a stochastic volatility model.

While many models have since incorporated the leverage effect, the exact interpretation remains a topic for debate. Figlewski and Wang (2000) find, among other anomalies, that there is no effect on volatility when a firm changes its debt-to-equity ratio by a change in outstanding debt or shares. Furthermore, Hasanhodzic and Lo (2011) show that the leverage effect is also present, and at least as strong, for firms that are fully financed by equity. Thus, while there is sometimes a link between financial leverage and the leverage effect, there is not enough proof to assume causality. Alternative economic interpretations have been suggested. French et al. (1987) and Campbell and Hentschel (1992) suggest that an anticipated increase in volatility requires a higher rate of return from an asset, which can only be produced by a fall in the asset price. They

suggest that the volatility affects the returns and not the other way around. Something that we theoretically agree on.

Despite the questionable causality of the leverage effect, it is agreed that there is indeed a link between (negative) returns and (increased) volatility, and the leverage effect is present in most sophisticated stochastic volatility models. Below, we discuss the different ways in which the leverage effect is incorporated in a stochastic volatility model and introduce a new way of doing this.

2.3 Stochastic volatility models with leverage

Harvey and Shephard (1996) were the first to model the leverage effect in a stochastic volatility model by assuming a correlated structure between the current return shock and future volatility shock:

$$\begin{pmatrix} \varepsilon_t \\ \eta_{t+1} \end{pmatrix} \sim \mathcal{N}\left\{ \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho_1 \\ \rho_1 & 1 \end{pmatrix} \right\}. \tag{2}$$

Here, we have an inter-temporal dependence between the return and volatility shocks. The shocks are multivariate normally distributed. With the well known conditional distribution of multivariate normally distributed variables, we can easily derive the conditional distribution of η_{t+1}, which is given by $\eta_{t+1}|\varepsilon_t \sim \mathcal{N}(\rho_1\varepsilon_t, 1-\rho_1^2)$. We can write this as $\eta_{t+1} = \rho_1\varepsilon_t + \sqrt{1-\rho_1^2}b_{t+1}$, where b_{t+1} is normally distributed with mean 0 and variance 1, $b_{t+1} \sim \mathcal{N}(0,1)$. Writing this in our model notation gives the inter-temporal leverage stochastic volatility model:

$$\begin{aligned} y_t &= \exp\left\{\frac{h_t}{2}\right\}\varepsilon_t, \quad \varepsilon_t \overset{iid}{\sim} \mathcal{N}(0,1), \\ h_t &= c + \phi h_{t-1} + \sigma_n \eta_t, \\ \eta_t &= \rho_1 \varepsilon_{t-1} + \sqrt{1-\rho_1^2}b_t, \quad b_t \overset{iid}{\sim} \mathcal{N}(0,1). \end{aligned} \tag{3}$$

As before, the log-volatility is (unconditionally) normally distributed with mean $\frac{c}{1-\phi}$ and variance $\frac{\sigma_n^2}{1-\phi^2}$, $h_t \sim \mathcal{N}(\frac{c}{1-\phi}, \frac{\sigma_n^2}{1-\phi^2})$. The expectation of tomorrow's return is given by:

$$\begin{aligned} E[y_{t+1}|\mathcal{I}_t] &= E\left[\exp\left\{\frac{h_{t+1}}{2}\right\}\varepsilon_{t+1}\middle|\mathcal{I}_t\right] = E\left[\exp\left\{\frac{c+\phi h_t + \sigma_n(\rho_1\varepsilon_t + \sqrt{1-\rho_1^2}b_{t+1})}{2}\right\}\varepsilon_{t+1}\middle|\mathcal{I}_t\right] \\ &= \exp\left\{\frac{c+\phi h_t}{2}\right\}\exp\left\{\frac{\sigma_n\rho_1\varepsilon_t}{2}\right\}E\left[\exp\left\{\frac{\sigma_n\sqrt{1-\rho_1^2}b_{t+1}}{2}\right\}\middle|\mathcal{I}_t\right]E\left[\varepsilon_{t+1}\middle|\mathcal{I}_t\right] = 0. \end{aligned} \tag{4}$$

Furthermore, it can easily be seen that the expected future volatility is affected by today's return:

$$E[h_{t+1}|y_t, h_t] = E\left[c + \phi h_t + \sigma_n \left(\rho_1 \varepsilon_t + \sqrt{1-\rho_1^2} b_{t+1}\right) \Big| y_t\right] = c + \phi h_t + \sigma_n \rho_1 E[\varepsilon_t | y_t, h_t] \tag{5}$$

However, as the log-volatility h_t is never actually observed, we take the expectation of future volatility conditional only on today's return:

$$\begin{aligned} E[h_{t+1}|y_t] &= c + \phi \frac{c}{1-\phi} + \sigma_n \rho_1 y_t E\left[\exp\left\{\frac{-h_t}{2}\right\}\right] \\ &= c + \phi \frac{c}{1-\phi} + \sigma_n \rho_1 y_t \exp\left\{\frac{c}{2(1-\phi)} + \frac{\sigma_n^2}{8(1-\phi^2)}\right\}. \end{aligned} \tag{6}$$

Here we use that we can write $\varepsilon_t = y_t \exp\left\{\frac{-h_t}{2}\right\}$ and that the unconditional distribution of the log-volatility is given by$h_t \sim \mathcal{N}(\frac{c}{1-\phi}, \frac{\sigma_n^2}{1-\phi^2})$. In this expression, the expectation of volatility is a linear function in y_t. If $\rho_1 < 0$, a negative return y_t leads to a higher expected volatility $E[h_{t+1}|y_t]$. If we define the leverage effect as an increase in volatility due to a negative return, the leverage effect is clearly present here as long as $\rho_1 < 0$. Harvey and Shephard (1996) estimate this model with QML using the Kalman filter, after transforming the returns and by conditioning on the signal of the return. Omori et al. (2007) estimate the model using Bayesian inference. These authors find, as many authors that followed them in using this model, a significant negative relation ρ_1 when applied to returns of stock indices.

Jacquier et al. (2004) were the first to assume a contemporaneous relation between the shocks. Instead of modelling an inter-temporal relation they assumed a correlation between the current return shock and the current volatility shock:

$$\begin{pmatrix} \varepsilon_t \\ \eta_t \end{pmatrix} \sim \mathcal{N}\left\{\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho_0 \\ \rho_0 & 1 \end{pmatrix}\right\}. \tag{7}$$

Again, the shocks are multivariate normally distributed. We write the conditional distribution of η_t as $\eta_t = \rho_0 \varepsilon_t + \sqrt{1-\rho_0^2}\xi_t$, where b_t is normally distributed with mean 0 and variance 1, $b_t \sim \mathcal{N}(0,1)$. Writing this in our model notation gives:

$$\begin{aligned} y_t &= \exp\left\{\frac{h_t}{2}\right\} \varepsilon_t, \quad \varepsilon_t \overset{iid}{\sim} \mathcal{N}(0,1), \\ h_t &= c + \phi h_{t-1} + \sigma_n \eta_t, \\ \eta_t &= \rho_0 \varepsilon_t + \sqrt{1-\rho_0^2} b_t, \quad b_t \overset{iid}{\sim} \mathcal{N}(0,1). \end{aligned} \tag{8}$$

The log-volatility is again (unconditionally) normally distributed with mean $\frac{c}{1-\phi}$ and variance $\frac{\sigma_n^2}{1-\phi^2}$, $h_t \sim \mathcal{N}(\frac{c}{1-\phi}, \frac{\sigma_n^2}{1-\phi^2})$. Yu (2005) shows that the expectation of tomorrow's return is given

by:

$$\begin{aligned} E[y_{t+1}|y_t, h_t] &= E\left[\exp\left\{\frac{h_{t+1}}{2}\right\}\varepsilon_{t+1}\middle|\mathcal{I}_t\right] \\ &= E\left[\exp\left\{\frac{c+\phi h_t+\sigma_n(\rho_0\varepsilon_{t+1}+\sqrt{1-\rho_0^2}b_{t+1})}{2}\right\}\varepsilon_{t+1}\middle|y_t, h_t\right] \\ &= \exp\left\{\frac{c+\phi h_t}{2}\right\}\exp\left\{\frac{\sigma_n^2}{8}\right\}\frac{1}{2}\sigma_n\rho_0, \end{aligned} \tag{9}$$

which is not equal to zero. Noting that the log-volatility is not observable, tomorrow's expected return conditional on today's expected return is equal to:

$$\begin{aligned} E[y_{t+1}|y_t] &= E[E[y_{t+1}|y_t, h_t]] = E\left[\frac{1}{2}\sigma_n\rho_0\exp\left\{\frac{c+\phi h_t}{2}\right\}\exp\left\{\frac{\sigma_n^2}{8}\right\}\right] \\ &= \frac{1}{2}\rho_0\sigma_n\exp\left\{\frac{2-\phi}{2-2\phi}c+\frac{2-\phi^2}{8-8\phi^2}\sigma_n^2\right\}, \end{aligned} \tag{10}$$

which is different from zero unless there is no contemporaneous leverage effect ($\rho_0 = 0$). Following the efficient market hypothesis, which states that a share price should reflect all information, it is argued that expected returns should be zero. This problematic feature in the contemporaneous leverage model was first argued by Harvey and Shephard (1996), who therefore modelled the leverage effect as inter-temporal. Also, this contemporaneous relation leads to autocorrelated returns (Catania, 2020) and a negative relation between returns and stochastic volatility as in Equation (5) can not be derived. Due to these reasons, the stochastic volatility model in (3) with inter-temporal leverage is more frequently used. However, this model is still used occasionally in literature, as Jacquier et al. (2004) and others find a significant correlation ρ_0 between the contemporaneous return shock ε_t and volatility shock η_t.

2.4 Stochastic volatility models with multiple leverage effects

Due to the inconsistency with the efficient market hypothesis, the stochastic volatility model with contemporaneous leverage is less accepted in literature than the model with inter-temporal leverage. However, the inter-temporal model indicates that there is a delay in the response of log-volatility to information in the returns. Intuitively, it would make more sense if volatility was to react immediately to new information becoming available from a return shock. Research into both models led to significant relations ρ_0 and ρ_1 between the return shock and the volatility shocks, which lead to a combination of the two models. Yu (2005) was the first to combine the

contemporaneous and inter-temporal leverage models. We write this model in our notation as:

$$\begin{aligned} y_t &= \exp\left\{\frac{h_t}{2}\right\}\varepsilon_t, \quad \varepsilon_t \overset{iid}{\sim} \mathcal{N}(0,1), \\ h_t &= c + \phi h_{t-1} + \sigma_n \eta_t, \\ \eta_t &= \rho_0 \varepsilon_t + \rho_1 \varepsilon_{t-1} + \sqrt{1-\rho_0^2-\rho_1^2}\, b_t, \quad b_t \overset{iid}{\sim} \mathcal{N}(0,1). \end{aligned} \tag{11}$$

Yu (2005) estimated this model using Bayesian inference, finding negative values for both correlation parameters. To do so, he rewrites the model using the following nonlinear state-space form:

$$\begin{aligned} y_t &= \frac{\exp\left\{\frac{h_t}{2}\right\}}{\sigma_n(1+\rho_0\rho_1)}\left(\rho_0(h_t - c - \phi h_{t-1}) + \rho_1(h_{t+1} - c - \phi h_t)\right) + \exp\left\{\frac{h_t}{2}\right\}\sqrt{1-\frac{\rho_0^2+\rho_1^2}{1+\rho_0\rho_1}}\, z_t, \\ h_{t+1} &= \phi h_t + \rho_0\rho_1(h_t - \phi h_{t-1}) + \sigma_n\sqrt{1-\rho_0^2\rho_1^2}\, u_{t+1}. \end{aligned} \tag{12}$$

Catania (2020) notes that this representation is correct but cannot be used for a Markov chain Monte Carlo (MCMC) sampler, as he argues error terms are not uncorrelated but instead are MA(2) processes. We agree with this conclusion but add that this representation is not correct. Yu (2005) notes that the volatility shocks are correlated in this model, as $E(\eta_t\eta_{t+1}) = E((\rho_0\varepsilon_t + \rho_1\varepsilon_{t-1} + \sqrt{1-\rho_0^2-\rho_1^2}b_t)(\rho_0\varepsilon_{t+1} + \rho_1\varepsilon_t + \sqrt{1-\rho_0^2-\rho_1^2}b_t)) = \rho_0\rho_1$. We can write the multivariate distribution of $\varepsilon_t, \eta_t, \eta_{t+1}$ as:

$$\begin{pmatrix} \varepsilon_t \\ \eta_{t+1} \\ \eta_t \end{pmatrix} \sim \mathcal{N}\left\{\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho_1 & \rho_0 \\ \rho_1 & 1 & \rho_0\rho_1 \\ \rho_0 & \rho_0\rho_1 & 1 \end{pmatrix}\right\}. \tag{13}$$

Now, we first compute η_{t+1} conditional on η_t, which can be written as $\eta_{t+1}|\eta_t = \rho_0\rho_1\eta_t + \sqrt{1-\rho_0^2\rho_1^2}u_{t+1}$. Which leads to the same state equation as Yu (2005) in the model of Equation (12). The conditional distribution of ε_t on the state errors η_t and η_{t+1} is normal, $\varepsilon_t|\eta_{t+1}, \eta_t$,

with mean and variance given by:

$$\begin{aligned}
\mu_\varepsilon &= 0 + \begin{pmatrix}\rho_1 & \rho_0\end{pmatrix}\begin{pmatrix}1 & \rho_0\rho_1\\ \rho_0\rho_1 & 1\end{pmatrix}^{-1}\begin{pmatrix}\eta_{t+1}\\ \eta_t\end{pmatrix} = \frac{1}{1-\rho_0^2\rho_1^2}\begin{pmatrix}\rho_1 & \rho_0\end{pmatrix}\begin{pmatrix}1 & -\rho_0\rho_1\\ -\rho_0\rho_1 & 1\end{pmatrix}\begin{pmatrix}\eta_{t+1}\\ \eta_t\end{pmatrix}\\
&= \frac{1}{1-\rho_0^2\rho_1^2}((\rho_1-\rho_0^2\rho_1)\eta_{t+1} + (\rho_0-\rho_0\rho_1^2)\eta_t),\\
\Sigma_\varepsilon &= 1 - \begin{pmatrix}\rho_1 & \rho_0\end{pmatrix}\begin{pmatrix}1 & \rho_0\rho_1\\ \rho_0\rho_1 & 1\end{pmatrix}^{-1}\begin{pmatrix}\rho_1\\ \rho_0\end{pmatrix} = 1 - \frac{1}{1-\rho_0^2\rho_1^2}(\rho_1^2+\rho_0^2-2\rho_0^2\rho_1^2).
\end{aligned} \tag{14}$$

This is clearly different from the model of Yu (2005). Furthermore, we can show that the error terms in the model of Equation (12) are not uncorrelated. Namely, if we keep the definition of η_{t+1} the same, $\eta_{t+1} = \frac{1}{\sigma_n}(h_{t+1} - c - \phi h_t)$, we can write this as $\eta_{t+1} = \rho_0\rho_1\eta_t + \sigma_n\sqrt{1-\rho_0^2\rho_1^2}u_{t+1}$. Rewriting this with error term u_{t+1} on the left-hand side gives:

$$u_{t+1} = \frac{1}{\sigma_n\sqrt{1-\rho_0^2\rho_1^2}}(\eta_{t+1} - \rho_0\rho_1\eta_t). \tag{15}$$

Now, if we remember that all η_t have expectation zero, but are correlated with $E[\eta_{t+1}\eta_t] = \rho_0\rho_1$, we have correlation between u_t and u_{t+1} given by:

$$\begin{aligned}
E[u_{t+1}u_t] &= E\left[\frac{1}{\sigma_n\sqrt{1-\rho_0^2\rho_1^2}}(\eta_{t+1}-\rho_0\rho_1\eta_t)\frac{1}{\sigma_n\sqrt{1-\rho_0^2\rho_1^2}}(\eta_t - \rho_0\rho_1\eta_{t-1})\right]\\
&= \frac{1}{\sigma_n^2(1-\rho_0^2\rho_1^2)}E[(\eta_{t+1}-\rho_0\rho_1\eta_t)(\eta_t-\rho_0\rho_1\eta_{t-1})].
\end{aligned} \tag{16}$$

This is different from zero, as we have multiple terms of η_t and correlation between η_{t+1} and η_t as well as correlation between η_t and η_{t-1}. In a similar manner, it can be shown that $E[u_{t+1}, u_{t-1}]$ is correlated. Thus, the error terms of the state equation in the model of Yu (2005) are not correlated but, in fact, an MA(2) process. H Furthermore, taking the state equation of the Yu (2005) model, we can write

$$z_t = \frac{1}{\exp\left\{\frac{h_t}{2}\right\}\sqrt{1-\frac{\rho_0^2+\rho_1^2}{1+\rho_0\rho_1}}}\left(y_t - \frac{\exp\left\{\frac{h_t}{2}\right\}}{1+\rho_0\rho_1}(\rho_0\eta_t + \rho_1\eta_{t+1})\right), \tag{17}$$

where we use that $\eta_t = \frac{1}{\sigma_n}(h_t - c - h_{t-1})$ for easier notation. Now, using the same reasoning as for u_t, we can show that $E[z_t, z_{t-i}] \neq 0$ for $i = 1, 2$, and z_t is therefore also an MA(2) process. Furthermore, we can show in a similar way that the volatility shocks and return shocks are correlated, namely $E[u_t, z_{t-i}] \neq 0$ for $i = -1, 0, 1, 2, 3$. Thus, we agree with Catania (2020) that

the state-space representation used by Yu (2005) is not correct, which we have shown here. Also for our correct derivation of this model, as in Equation (14), we find similar correlation structures. Wen we estimate the model using our correct specification, we do find results close to that of Yu (2005). Namely, a significant leverage parameter ρ both contemporaneously and inter-temporal. However, as we have shown that this representation is not correct due to the correlated errors, we conclude that these results are not valid. We will discuss a valid representation of this stochastic volatility model later in this paper.

Catania (2020) extends the model of Yu (2005) by introducing a stochastic volatility model with a general leverage specification, where the leverage effect is both contemporaneous and inter-temporal, as that of Yu (2005). However, now the inter-temporal leverage can last for more than one time step. The amount of days the leverage effect exists is indicated by m, where $m = 0$ leads to the contemporaneous leverage effect model and $m \geq 0$ leads to a model containing both the contemporaneous leverage effect and m inter-temporal leverage effects at time steps $t+1, \ldots, t+m$. In our notation, we write this model as:

$$\begin{aligned} y_t &= \exp\left\{\frac{h_t}{2}\right\}\varepsilon_t, \quad \varepsilon_t \overset{iid}{\sim} \mathcal{N}(0,1), \\ h_t &= c + \phi h_{t-1} + \sigma_n \eta_t, \\ \eta_t &= \sum_{j=0}^{m} \rho_j \varepsilon_{t-j} + \sqrt{1 - \sum_{j=0}^{m} \rho_j^2}\, b_t, \quad b_t \overset{iid}{\sim} \mathcal{N}(0,1). \end{aligned} \tag{18}$$

The log-volatility state equation of this model is quite different, containing present and multiple past return shocks. While the unconditional mean of the log-volatility, $\frac{c}{1-\phi}$, is unchanged, the unconditional variance of the log-volatility is now given by $V[h_t] = \frac{\sigma_n^2}{1-\phi^2}\left[1 + 2\sum_{i=1}^{m}\phi^i \sum_{j=1}^{m}\rho_j\rho_{j-i}\right]$. Furthermore, the expectation of tomorrow's return is now given by:

$$E[y_{t+1}|y_t] = \frac{1}{2}\sigma_n\rho_0 \exp\left\{\frac{c}{2(1-\phi)} + \frac{\sigma_n^2}{8(1-\phi^2)}\left[1 + 2\sum_{i=1}^{m}\phi^i \sum_{j=1}^{m}\rho_j\rho_{j-i}\right]\right\}. \tag{19}$$

Which is, as in the model with contemporaneous leverage, not equal to zero if $\rho_0 \neq 0$. Catania (2020) implements the model using both a particle filter and quasi maximum likelihood via the Kalman filter. However, to use the Kalman filter, Catania (2020) adds the constraint $\rho_0 = 0$, such that there is only inter-temporal leverage. In the rest of this paper, we will use this model, and refer to it as the 'Catania model'. Nested in this model are the other stochastic volatility models. Setting the constraint that all ρ are zero leads to the classic volatility model of Taylor (1986). Using $m = 0$ leads to the contemporaneous leverage model of Jacquier et al. (2004), and

setting the constraint $\rho_0 = 0$ and $m = 1$ leads to the inter-temporal leverage model of Harvey and Shephard (1996).

We have seen that including contemporaneous leverage leads to a model that has the suboptimal characteristic of expected returns unequal to zero. However, we argue that this result might not be completely false. It is well known that the price of an asset is highly dependent on its risk. In fact, it is one of the most famous and researched topics in asset pricing. In the stochastic volatility models that we use, volatility is a measure of risk for the return. Thus, it does not seem unlikely that the expected return is unequal to zero. Therefore, in many of the stochastic volatility models used in this paper we include contemporaneous leverage, and compare these to stochastic volatility models without contemporaneous leverage.

2.5 New stochastic volatility model

In the Catania model, it is assumed that the volatility shock η_t is a function of current and past returns $\varepsilon_t, .., \varepsilon_t - m$. In his application to index returns, Catania argues that the best stochastic volatility models include around 5 leverage effects. Overall, the sum of squared leverage effects $\sum_{j=1}^{m} \rho_j^2$ for some major indices varies from 0.70 to 0.90. This would mean that around 70% to 90% of the shocks to volatility come from past returns. In other words, only 10% to 30% of the innovations of volatility is actually new information. This would mean that the stochastic volatility model approaches an ARCH model, where the volatility is only a function of past return shocks. This is not intuitive and seems to take away the usefulness of the stochastic volatility model. Therefore, we propose a new way of modelling the correlation between return shocks and volatility shocks. In this specification, volatility shocks are independent, while return shocks depend on current and/or future volatility shocks. This model is given by:

$$\begin{aligned} y_t &= \exp\left\{\frac{h_t}{2}\right\}\varepsilon_t, \\ h_t &= c + \phi h_{t-1} + \sigma_n \eta_t, \quad \eta_t \overset{iid}{\sim} \mathcal{N}(0,1), \\ \varepsilon_t &= \sum_{j=0}^{m} \rho_j \eta_{t+j} + \sqrt{1 - \sum_{j=0}^{m} \rho_j^2}\xi_t, \quad \xi_t \overset{iid}{\sim} \mathcal{N}(0,1). \end{aligned} \tag{20}$$

What differentiates this model from the previous is that now volatility shocks are completely independent while shocks to returns incorporate some effect of the volatility shocks. The unconditional distribution of the log-volatility is now the same as in the first stochastic volatility models, $h_t \sim \mathcal{N}(\frac{c}{1-\phi}, \frac{\sigma_n^2}{1-\phi^2})$. The amount of future volatility shocks that affect the return is left

variable for now, and is indicated by m. The expected future return is given by:

$$\begin{aligned} E[y_{t+1}|y_t, h_t] &= E\left[\exp\left\{\frac{h_{t+1}}{2}\right\}\varepsilon_{t+1}\middle|\mathcal{I}_t\right] \\ &= E\left[\exp\left\{\frac{c+\phi h_t+\sigma_n\eta_{t+1}}{2}\right\}\left(\sum_{j=0}^{m}\rho_j\eta_{t+j+1}+\sqrt{1-\sum_{j=0}^{m}\rho_j^2}\xi_{t+1}\right)\middle|y_t, h_t\right] \\ &= \exp\left\{\frac{c+\phi h_t}{2}\right\}E\left[\exp\left\{\frac{\sigma_n\eta_{t+1}}{2}\right\}(\rho_0\eta_{t+1})\middle|y_t, h_t\right] = XXX, \end{aligned} \tag{21}$$

which is again unequal to zero if $\rho_0 \neq 0$. However, as we argued before, this does not have to be completely unnatural, as long as we remember that risk is a very important driver for financial returns. One downside of this specification is that financial returns are now subject to the same shocks η_t. For example, when including contemporaneous leverage and one inter-temporal leverage effect, such that $\rho_0 \neq 0$ and $\rho_1 \neq 0$, we have correlation between returns given by:

$$\begin{aligned} E[y_{t+1}y_t] &= E\left[\exp\left\{\frac{h_{t+1}}{2}\right\}\varepsilon_{t+1}[\exp\left\{\frac{h_t}{2}\right\}\varepsilon_t\right] \\ &= E\Bigg[\exp\left\{\frac{h_{t+1}}{2}\right\}\left(\rho_0\eta_{t+1}+\rho_1\eta_{t+2}+\sqrt{1-\rho_0^2-\rho_1^2}\xi_t\right) \\ &\times\exp\left\{\frac{h_t}{2}\right\}\left(\rho_0\eta_t+\rho_1\eta_{t+1}+\sqrt{1-\rho_0^2-\rho_1^2}\xi_t\right)\Bigg] \\ &= E\left[\exp\left\{\frac{h_{t+1}}{2}\right\}(\rho_0\eta_{t+1})\left(\exp\left\{\frac{h_t}{2}\right\}(\rho_1\eta_{t+1}\right)\right] \\ &= \rho_0\rho_1 E[\exp\left\{\frac{h_{t+1}}{2}\right\}\exp\left\{\frac{h_t}{2}\right\}], \end{aligned} \tag{22}$$

which is clearly nonzero if $\rho_0 \neq 0$ and $\rho_0 \neq 1$. For leverage effects at other times, for example if $\rho_1 \neq 0$ and $\rho_2 \neq 1$, results are similar. Thus, if we have multiple leverage effects there is autocorrelation in the financial returns. As this is a characteristic that is not present in financial returns, this might be a problem when applying this stochastic volatility model. However, we think it might still be useful to apply this model to empirical data, as it might provide new insights on the timing of the leverage effect and understand the causality relation between return shocks and volatility shocks.

In this new stochastic volatility model, hereafter named the new SV model, we assume return shocks are a function of present and future volatility shocks. We assume that shocks to volatility are completely 'new' information, while shocks to returns are partly dependent on volatility shocks. We argue that this is a more intuitive way of modelling the leverage effect, as it is logical that return shocks depend on volatility shocks, since returns depend on risk. Whereas modelling

the volatility shocks dependent on past return shocks would indicate that risk is dependent on previous returns. While the 'financial leverage' theory was one way to argue why this is the case, we have discussed literature that proved that this does not correspond to empirical evidence about financial leverage.

Overall, we agree with the empirical findings of a correlation between volatility shocks and return shocks. However, we argue that modelling the leverage effect as an inter-temporal relation is not correct, as it would be expected for financial returns and risk to incorporate any new information directly into the price and the corresponding volatility process. Furthermore, we think that volatility shocks are not partly explained by return shocks, but that the causality is vice versa, namely that volatility shocks affect return shocks.

3 Filtering and estimation of state-space models

The stochastic volatility models described in the paper so far are state-space models. Models that have consist of a time series of one or more observed variables, in our case financial returns, for which we assume are subject to some underlying process, in our case, the log-volatility process. This section discusses and compares three methods to filter and estimate state-space models. After, in Section 4, we discuss and apply these estimation methods to our volatility models. We write a state-space model in the following notation:[4]

$$\boldsymbol{y}_t \sim p(\boldsymbol{y}_t|\boldsymbol{h}_t;\boldsymbol{\theta}), \quad \boldsymbol{h}_t \sim p(\boldsymbol{h}_t|\boldsymbol{h}_{t-1};\boldsymbol{\theta}), \quad \boldsymbol{h}_1 \sim p(\boldsymbol{h}_1;\boldsymbol{\theta}). \tag{23}$$

Here, for discrete times $t = 1, \ldots, T$, $\boldsymbol{y}_t$ are the observations and $\boldsymbol{h}_t$ are the unobserved states, which in the stochastic volatility models from section 2 are returns and volatilities, respectively. $\boldsymbol{\theta}$ denotes the model parameters, in our case $\boldsymbol{\theta} = (c, \sigma_n, \phi, \boldsymbol{\rho})$. The observation is drawn conditional on the unobserved states, while the states follow a first-order Markov process. We assume these probability densities are known.

To estimate state-space models, one can split the problem into two cases. First, the filtering problem, where one filters the unobserved states $\boldsymbol{h}_t$, given the observed $\boldsymbol{y}_t$ and assuming the model parameters $\boldsymbol{\theta}$ are known. Second, the estimation problem, where both the parameter $\boldsymbol{\theta}$ and unobserved states $\boldsymbol{h}_t$ have to be estimated. We use three different filtering and estimation methods, namely the Kalman filter, the particle filter, and the Bellman filter. These methods are presented in Section 3.1, Section 3.2, and Section 3.3 respectively. In Section 3.4 we perform

[4] State-space models can also incorporate exogenous factors $\boldsymbol{x}_t$ in the observation equation. For simplicity, this is left out.

simulation studies in order to investigate the filtering performance and estimation performance of the three methods.

3.1 Kalman filter

The Kalman filter is introduced in the seminal paper by Kalman (1960) and is widely used in literature. The Kalman filter can solve the filtering problem in closed form when the state-space model is linear and Gaussian. Thus, to apply the Kalman filter, we need the following model:

$$\begin{aligned} \boldsymbol{y}_t &= \boldsymbol{\alpha} + \boldsymbol{\beta}\boldsymbol{h}_t + \boldsymbol{\varepsilon}_t, & \boldsymbol{\varepsilon}_t &\sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}}) \\ \boldsymbol{h}_t &= \boldsymbol{c} + \boldsymbol{\phi}\boldsymbol{h}_{t-1} + \boldsymbol{\eta}_t, & \boldsymbol{\eta}_t &\sim \mathcal{N}(\mathbf{0}, \boldsymbol{Q}). \end{aligned} \tag{24}$$

We can see that the observations $\boldsymbol{y}_t$ and the unobserved states $\boldsymbol{h}_t$ are normally distributed conditional on $\boldsymbol{h}_t$ and $\boldsymbol{h}_{t-1}$, respectively. Due to these characteristics, the Kalman filter can analytically calculate the prediction and updating step for the state-space model. Thus, the Kalman filter gives the optimal solution for such a linear Gaussian model. Also, parameter estimation becomes relatively easy through the prediction error decomposition and can be done with maximum likelihood estimation. As the Kalman filter is extensively studied in literature and very well known, we refer to other literature for a more thorough description, for example (Welch and Bishop, 1995; Meinhold and Singpurwalla, 1983).

3.2 Particle filter

The particle filter, also called sampling importance resampling (SIR), is introduced in the seminal paper by Gordon et al. (1993). To use this algorithm for state-space filtering, the only requirements are that it is possible to simulate from the state density $\boldsymbol{h}_t \sim p(\boldsymbol{h}_t|h_{t-1};\boldsymbol{\theta})$ and that it is possible to evaluate the observation density $\boldsymbol{y}_t \sim p(\boldsymbol{y}_t|\boldsymbol{h}_t;\boldsymbol{\theta})$. This provided a breakthrough in the filtering of state-space models, whereas previous Markov chain Monte Carlo (MCMC) methods, for example, Kitagawa (1987), relied on evaluating the state density $\boldsymbol{h}_t \sim p(\boldsymbol{h}_t|h_{t-1};\boldsymbol{\theta})$ instead of only simulating from this density. [5]

[5] A new and innovative extension to the particle filter is to use Rao-Blackwellization. The idea of the Rao-Blackwellized particle filter (RBPF) (Akashi and Kumamoto, 1977; Doucet et al., 2001; Gordon et al., 2004), also called the Mixture Kalman filter (MKF) by Chen and Liu (2000), is to evaluate some of the filtering equations analytically. This requires that the observation equation is linear and Gaussian in part of the state variables. In this paper we consider only fully nonlinear/non-Gaussian observations, and thus do not use Rao-Blackwellized particle filters.

3.2.1 Filtering

Particle filtering uses the principle of Bayesian updating, implying that the density of the state conditional on all available information is constructed from a prior and a likelihood. Suppose that at time t we have a set of random 'particles' $\boldsymbol{h}_t^1, \ldots, \boldsymbol{h}_t^N$ with corresponding discrete probability masses $\pi_t^1, \ldots, \pi_t^N$ drawn from the density $p(\boldsymbol{h}_t|\mathcal{I}_t;\boldsymbol{\theta})$. [6] Since we can simulate from the state density $\boldsymbol{h}_t \sim p(\boldsymbol{h}_t|h_{t-1};\boldsymbol{\theta})$ we can approximate the prediction density

$$p(\boldsymbol{h}_{t+1}|\mathcal{I}_t;\boldsymbol{\theta}) = \int p(\boldsymbol{h}_{t+1}|h_t;\boldsymbol{\theta})p(h_t|\mathcal{I}_t;\boldsymbol{\theta})\mathrm{d}\boldsymbol{h}_t. \tag{25}$$

Then at time step $t+1$ the observation y_{t+1} becomes available and we can update this prior via the Bayes rule

$$p(\boldsymbol{h}_{t+1}|\mathcal{I}_{t+1};\boldsymbol{\theta}) = \frac{p(y_{t+1}|\boldsymbol{h}_{t+1};\boldsymbol{\theta})p(\boldsymbol{h}_{t+1}|\mathcal{I}_t;\boldsymbol{\theta})}{p(\boldsymbol{y}_{t+1}|\mathcal{I}_{t+1};\boldsymbol{\theta})}. \tag{26}$$

Implementing this principle recursively forms the basis of the particle filter. By propagating and updating these particles, a sample is obtained which approximates the true filtering density:

$$p(\boldsymbol{h}_{t+1}|\mathcal{I}_{t+1};\boldsymbol{\theta}) \propto p(y_{t+1}|\boldsymbol{h}_{t+1};\boldsymbol{\theta}) \int p(\boldsymbol{h}_{t+1}|h_t;\boldsymbol{\theta})p(\boldsymbol{h}_t|\mathcal{I}_t;\boldsymbol{\theta})\mathrm{d}\boldsymbol{h}_t. \tag{27}$$

To sample from the density, we follow the bootstrap algorithm of Gordon et al. (1993), presented in Algorithm 1. By drawing N particles from the distribution $h_0 \sim p(h_0;\boldsymbol{\theta})$, which is generally the stationary distribution, we can initialize the particle filter.

This will yield an approximation of the desired posterior density $p(\boldsymbol{h}_{t+1}|\mathcal{I}_{t+1};\boldsymbol{\theta})$. Here, δ is the Dirac-delta function with mass zero. Step 3 is the sampling step, referred to as the weighted bootstrap step. Smith and Gelfand (1992) show that as $N \to \infty$ these discrete distributions go to their required density. Thus, the obtained samples in step 1 and step 3 will converge to the predictive density $p(\boldsymbol{h}_{t+1}|\mathcal{I}_t;\boldsymbol{\theta})$ and the filtered density $p(\boldsymbol{h}_{t+1}|\mathcal{I}_{t+1};\boldsymbol{\theta})$, respectively. Now, we can approximate the hidden state $\boldsymbol{h}_t$ by taking the average over the sampled states at that time step $\frac{1}{N}\sum_{i=1}^N h_t^i$.

3.2.2 Parameter estimation

The likelihood can be estimated using the prediction decomposition. Using the output from the algorithm, we can estimate the prediction density by taking the average of the unnormalized

[6] Note that in the rest of this paper $\mathcal{I}_t$ denotes the set off all available $y_{1:t}$ and $h_{1:t}$. Since the state variables h_t are not observed but only filtered, $\mathcal{I}_t$ does not denote the available information at time $t-1$ but rather the 'filtration' at time $t-1$.

Algorithm 1: SIR algorithm from Gordon et al. (1993)

Initialization:

Draw $\boldsymbol{h}_0^i \sim p(\boldsymbol{h}_0; \boldsymbol{\theta})$ for i = 1,...,N.

for $t = 0, \ldots, T-1$ **do**

Importance sampling step:

1. For $i = 1 : N$, sample $\tilde{\boldsymbol{h}}_{t+1}^i \sim p(\boldsymbol{h}_{t+1} | \boldsymbol{h}_t^i; \boldsymbol{\theta})$.
2. For $i = 1 : N$, calculate weights:

$$\omega_{t+1}^i = y_{t+1} \sim p(y_{t+1} | \boldsymbol{h}_{t+1}^i; \boldsymbol{\theta})$$

3. For $i = 1 : N$, calculate normalised weights:

$$\pi_{t+1}^i = \frac{\omega_{t+1}^i}{\sum_{j=1}^N \omega_{t+1}^j}.$$

Resampling (selection) step:

4. For $i = 1 : N$, sample $\boldsymbol{h}_{t+1}^i \sim \sum_{i=1}^N \pi_{t+1}^i \delta(\boldsymbol{h}_{t+1} - \tilde{\boldsymbol{h}}_{t+1}^i)$.

end

weights ω_{t+1}^i

$$\hat{p}(y_{t+1} | \mathcal{I}; \boldsymbol{\theta}) = \frac{1}{N} \sum_{i=1}^N p(y_{t+1} | \tilde{\boldsymbol{h}}_{t+1}^i; \boldsymbol{\theta}) = \frac{1}{N} \sum_{i=1}^N \omega_{t+1}^i. \tag{28}$$

Del Moral (2004) shows that this estimator is consistent and unbiased for $N \geq 2$. For parameter estimation we need to estimate the (log-)likelihood function, which is given by

$$\mathcal{L}(\boldsymbol{\theta}) = \log p(y_1, \ldots, y_T | \boldsymbol{\theta}) = \sum_{t=1}^T \log p(y_t | \boldsymbol{\theta}; \mathcal{I}_{t-1}). \tag{29}$$

Thus, using the estimate for the prediction density in Equation (28), the estimate for the log-likelihood is

$$\hat{\mathcal{L}}(\boldsymbol{\theta}) = \sum_{t=1}^T \log \hat{p}(y_t | \boldsymbol{\theta}; \mathcal{I}_{t-1}) = \sum_{t=1}^T \log \left(\frac{1}{N} \sum_{i=1}^N \omega_t^k \right). \tag{30}$$

This is the sum of the log of the mean of unnormalized weights over all time steps $t = 1, \ldots, T$, and can be estimated quickly with the already calculated weights.

Now, a problem that arises with the estimation of the log-likelihood of this particle filter is that the log-likelihood is not a continuous function of the parameters $\boldsymbol{\theta}$. If the particles $\boldsymbol{h}_t^i, t = 1, \ldots, N$ are slightly altered, the proposal samples h_{t+1}^i will suffer the same consequence. As the corresponding discrete probabilities also change, such that the resampled particles will not be close, and the likelihood will not be a continuous function of $\boldsymbol{\theta}$. This leads to implications for (gradient-based) maximization and computing standard errors, as is shown by Liu and West (2001) and Polson et al. (2008). Malik and Pitt (2011) present a filter, the continuous sampling importance resampling (CSIR) filter, which computes a smooth approximation of the likelihood. It is im-

portant to note that this method only works for models with one state variable. Hürzeler and Künsch (2001) propose an importance sampling method for multidimensional models. However, this method is more computationally intensive and only provides low variance estimates in the neighborhood of a suitably preselected parameter value (Kantas et al., 2015). In this paper we use the particle filter to estimate one-dimensional state-space models, and thus use the method of Malik and Pitt (2011).

Malik and Pitt (2011) introduce a new way of resampling in step 4 of Algorithm 1 form Gordon et al. (1993). From now on, we consider the one-dimensional state variable h_t. We start by sorting particles h_{t+1}^i in ascending order resulting in particles $h_{t+1}^1, h_{t+1}^2, \ldots, h_{t+1}^N$, such that $h_{t+1}^1 \leq h_{t+1}^2 \leq \cdots \leq h_{t+1}^N$. Malik and Pitt (2011) show that sampling $h_{t+1}^i \sim \sum_{i=1}^N \pi_{t+1}^i \delta(h_{t+1} - \tilde{h}_{t+1}^i)$ is equal to sampling from the empirical distribution function $\hat{F}_N(h_{t+1}) = \sum_{i=1}^N \pi_{t+1}^i I(h_{t+1} - h_{t+1}^i)$, where $I(q)$ is the indicator function taking value 1 if $q > 0$ and 0 otherwise. Sampling from this step function is what leads to a non-continuous likelihood function in $\boldsymbol{\theta}$. Malik and Pitt (2011) propose to approximate this distribution by a function which is constructed to be continuous. This function is given by:

$$\tilde{F}_N(h_{t+1}) = \sum_{i=0}^{N} \lambda_{t+1}^i G_i \left(\frac{h_{t+1} - h_{t+1}^i}{h_{t+1}^{i+1} - h_{t+1}^i} \right), \tag{31}$$

where $h_{t+1}^{N+1} = \infty$ and $h_{t+1}^0 = -\infty$. The new weights λ_{t+1} are given by:

$$\lambda_{t+1}^0 = \frac{\pi_{t+1}^1}{2}, \qquad \lambda_{t+1}^N = \frac{\pi_{t+1}^N}{2}, \qquad \lambda_{t+1}^i = \frac{\pi_{t+1}^{i+1} + \pi_{t+1}^i}{2}, \quad \text{for } i = 1, \ldots, N-1. \tag{32}$$

For function $G_i(\cdot), i = 1, \ldots, N$ we take the standard uniform distribution, such that $G_i(q) = q$ for $i = 1, \ldots, N$. Furthermore, at the end points we take $G_0(q) = I(q)$ and $G_N(q) = I(q)$. We now have a monotonically increasing function, which in fact is constructed from interpolating the empirical CDF $\hat{F}_N(h_{t+1})$. Malik and Pitt (2011) show that the distance between $\hat{F}_N(h_{t+1})$ and $\tilde{F}_N(h_{t+1})$ is of order $\frac{1}{N}$, such that as $N \to \infty$ we have $\tilde{F}_N(h_{t+1}) \to \hat{F}_N(h_{t+1})$.

For the above method we need a set of N uniform variables, $u_1 \leq \cdots \leq u_N$. To generate these uniforms, we use stratification. Stratification is used in particle filtering to reduce sample impoverishment (Kitagawa, 1996; Carpenter et al., 1999; Liu and Chen, 1998). We use the method proposed by Carpenter et al. (1999) to generate uniforms as:

$$u_j = \frac{(j-1) + u}{N}, \qquad \text{for } j = 1, \ldots, N, \qquad \text{where } u \sim UID(0, 1). \tag{33}$$

The new resampling step for the particle filter, to replace step 4 from Algorithm 1, is presented

Algorithm 2: Continuous resampling step from Malik and Pitt (2011)

Initialization:

Set $s = 0, j = 1$, and draw uniforms $u_1, \dots, u_N$ as in Equation (33). Initialise vector $r^1, \dots, r^N$ to store new regions.

Compute new regions:

for $i = 0, \dots, N$ **do**

1. $s = s + \lambda^i$
2. **while** $u_j \leq s$, $j \leq N$ **do**
   1. $r^j = i$.
   2. $u_j^* = (u_j - (s - \lambda^i))/\lambda^i$.
   3. j = j+1.

end

end

Assign new samples:

for $j = 0, \dots, N$ **do**

1. If $r^j = 0$, set $h_{t+1}^{j*} = h_{t+1}^1$
2. If $r^j = N$, set $h_{t+1}^{j*} = h_{t+1}^N$
3. Otherwise, set $h_{t+1}^{j*} = (h_{t+1}^{r^j+1} - h_{t+1}^{r^j}) \times u_j^* + h_{t+1}^{r^j}$

end

This produces samples $h_{t+1}^{j*}, j = 1, \dots, N$ from $\tilde{F}_N(h_{t+1})$.

in Algorithm 2. Here, we have suppressed subscript $t + 1$ for notational convenience.

Now, using the sampling importance resampling algorithm with this resampling schedule, we gain a smooth likelihood function in model parameters $\boldsymbol{\theta}$. Therefore, we can estimate the parameters with regular optimization techniques.

3.3 Bellman filter

The Bellman filter is a new filtering method introduced by Lange (2020). The Bellman filter is based on the mode, the mode being appealing due to its 'optimality property analogous to that of maximum likelihood estimates of fixed parameters in finite samples' (Durbin and Koopman, 2012, p. 252-3). The straightforward method for computing the mode is computationally unattainable because it involves re-estimating the entire sequence of states for each time step. Furthermore, the mode estimator does not address the parameter estimation problem. Lange (2020) employs the principle of dynamic programming from Bellman (1957) to circumvent these problems.

3.3.1 Filtering

The joint log-likelihood of the states and observations is written as $\ell(\boldsymbol{h}_{1:t}, \boldsymbol{y}_{1:t})$. As before, the observations $\boldsymbol{y}_{1:t}$ are known while the states $\boldsymbol{h}_{1:t}$ are unknown and to be estimated. For our state-space models, we can rewrite this joint log-likelihood using the 'probability chain rule' (Godsill

et al., 2004) as:

$$\ell(\boldsymbol{h}_{1:t}, \boldsymbol{y}_{1:t}) = \sum_{i=1}^{t} \ell(\boldsymbol{y}_i|\boldsymbol{h}_i) + \sum_{i=2}^{t} \ell(\boldsymbol{h}_i|\boldsymbol{h}_{i-1}) + \ell(\boldsymbol{h}_1). \tag{34}$$

By noting that the joint log-likelihood is a recursive relation for $2 \leq t \leq n$, we can rewrite this to:

$$\ell(\boldsymbol{h}_{1:t}, \boldsymbol{y}_{1:t}) = \ell(\boldsymbol{y}_t|\boldsymbol{h}_t) + \ell(\boldsymbol{h}_t|\boldsymbol{h}_{t-1}) + \ell(\boldsymbol{h}_{1:t-1}, \boldsymbol{y}_{1:t-1}). \tag{35}$$

Now, Lange (2020) defines the *value function* by maximizing $\ell(\boldsymbol{h}_{1:t}, \boldsymbol{y}_{1:t})$ with respect to all previous states $\boldsymbol{h}_{1:t-1}$. This value function is given by:

$$V_t(\boldsymbol{h}_t) := \max_{\boldsymbol{h}_{1:t-1} \in \mathbb{R}^{m \times (t-1)}} \ell(\boldsymbol{h}_{1:t}, \boldsymbol{y}_{1:t}) = \ell(\boldsymbol{y}_t|\boldsymbol{h}_t) + \max_{\boldsymbol{h}_{t-1} \in \mathbb{R}^m} \left\{ \ell(\boldsymbol{h}_t|\boldsymbol{h}_{t-1}) + V_{t-1}(\boldsymbol{h}_{t-1}) \right\}. \tag{36}$$

Now, when the state equation has linear and Gaussian dynamics, the Bellman filter can be exploited very usefully. The value function $V_{t-1}(\boldsymbol{h}_{t-1})$ can be approximated by a (multivariate) quadratic function:

$$V_{t-1}(\boldsymbol{h}_{t-1}) \approx -\frac{1}{2}(\boldsymbol{h}_{t-1} - \boldsymbol{h}_{t-1|t-1})' \boldsymbol{I}_{t-1|t-1} (\boldsymbol{h}_{t-1} - \boldsymbol{h}_{t-1|t-1}) + \text{constants}, \quad \boldsymbol{h}_{t-1} \in \mathbb{R}^m, \tag{37}$$

for an estimate of the state $\boldsymbol{h}_{t-1|t-1}$ and corresponding symmetric and positive definite precision matrix $\boldsymbol{I}_{t-1|t-1}$. As we are interested in maximizing the value function, constants in the equation can be ignored. As the state equation is linear and Gaussian, $\ell(\boldsymbol{h}_t|\boldsymbol{h}_{t-1})$ is a quadratic function in the state variables and can be written as:

$$\ell(\boldsymbol{h}_t|\boldsymbol{h}_{t-1}) \approx -\frac{1}{2}(\boldsymbol{h}_t - \boldsymbol{c} - \boldsymbol{\Phi}\boldsymbol{h}_{t-1})' \boldsymbol{Q}^{-1} (\boldsymbol{h}_t - \boldsymbol{c} - \boldsymbol{\Phi}\boldsymbol{h}_{t-1}) + \text{constants}, \quad \boldsymbol{h}_t, \boldsymbol{h}_{t-1} \in \mathbb{R}^m. \tag{38}$$

Lange (2020) shows that, by substituting the quadratic approximation in (37) and the state transition of (38) into Bellman's equation, we can write the value function as:

$$\begin{aligned} V_t(\boldsymbol{h}_t) = \ell(\boldsymbol{y}_t|\boldsymbol{h}_t) + \max_{\boldsymbol{h}_{t-1} \in \mathbb{R}^m} \Big\{ & -\frac{1}{2}(\boldsymbol{h}_t - \boldsymbol{c} - \boldsymbol{\Phi}\boldsymbol{h}_{t-1})' \boldsymbol{Q}^{-1} (\boldsymbol{h}_t - \boldsymbol{c} - \boldsymbol{\Phi}\boldsymbol{h}_{t-1}) \\ & - \frac{1}{2}(\boldsymbol{h}_{t-1} - \boldsymbol{h}_{t-1|t-1})' \boldsymbol{I}_{t-1|t-1} (\boldsymbol{h}_{t-1} - \boldsymbol{h}_{t-1|t-1}) \Big\} + \text{constants}, \quad \boldsymbol{h}_t \in \mathbb{R}^m, \end{aligned} \tag{39}$$

where we use an equality sign instead of an approximation. While the equation is not exact, for simplicity, we write this as equality while keeping in mind that the solution is in fact generally not exact. Now, since $\boldsymbol{h}_{t-1}$ appears quadratically on the right-hand side, we can compute the first-order condition of this equation and solve this for $\boldsymbol{h}_{t-1}$. The optimal solution depends linearly

on $\boldsymbol{h}_t$, substituting the optimal solution back into the value function gives:

$$V_t(\boldsymbol{h}_t) = \ell(\boldsymbol{y}_t|\boldsymbol{h}_t) + -\frac{1}{2}(\boldsymbol{h}_t - \boldsymbol{h}_{t|t-1})' \boldsymbol{I}_{t|t-1}^{-1}(\boldsymbol{h}_t - \boldsymbol{h}_{t|t-1}) + \text{constants}, \quad \boldsymbol{h}_t \in \mathbb{R}^m. \tag{40}$$

Here the predicted state $\boldsymbol{h}_{t|t-1}$ and the predicted precision matrix $\boldsymbol{I}_{t|t-1}$ are defined as:

$$\boldsymbol{h}_{t|t-1} := \boldsymbol{c} + \boldsymbol{\Phi}\boldsymbol{h}_{t-1|t-1} \tag{41}$$

$$\boldsymbol{I}_{t|t-1} := \boldsymbol{Q}^{-1} - \boldsymbol{Q}^{-1}\boldsymbol{\Phi}(\boldsymbol{I}_{t-1|t-1}\boldsymbol{\Phi}'\boldsymbol{Q}^{-1}\boldsymbol{\Phi})\boldsymbol{\Phi}'\boldsymbol{Q}^{-1}. \tag{42}$$

These equations form the prediction step of the Bellman filter of Lange (2020). For the filtering step, Lange (2020) show that the filtered state $\boldsymbol{h}_{t|t}$ and filtered precision matrix $\boldsymbol{I}_{t|t}$ can be found as:

$$\boldsymbol{h}_{t|t} = \max_{\boldsymbol{h}\in\mathbb{R}^m} V_t(\boldsymbol{h}), \qquad \boldsymbol{I}_{t|t} = -\frac{\mathrm{d}^2 V_t(\boldsymbol{h})}{\mathrm{d}\boldsymbol{h}\boldsymbol{h}'}\bigg|_{\boldsymbol{h}=\boldsymbol{h}_{t|t}}. \tag{43}$$

The gradient and negative Hessian of the value function can be approximated as:

$$\frac{\mathrm{d}V_t(\boldsymbol{h})}{\mathrm{d}\boldsymbol{h}} = \frac{\mathrm{d}\ell(\boldsymbol{y}_t|\boldsymbol{h})}{\mathrm{d}\boldsymbol{h}} - \boldsymbol{I}_{t|t-1}(\boldsymbol{h} - \boldsymbol{h}_{t|t-1}, \qquad \boldsymbol{h} \in \mathbb{R}^m, \tag{44}$$

$$-\frac{\mathrm{d}^2 V_t(\boldsymbol{h})}{\mathrm{d}\boldsymbol{h}\boldsymbol{h}'} = \boldsymbol{I}_{t|t-1} - \frac{\mathrm{d}^2\ell(\boldsymbol{y}_t|\boldsymbol{h})}{\mathrm{d}\boldsymbol{h}\boldsymbol{h}'}, \qquad \boldsymbol{h} \in \mathbb{R}^m. \tag{45}$$

Now, we can use Newton's optimization method, an iterative method, to maximize the value function $V_t(\boldsymbol{h})$ with respect to the state variable. Here, define the derivative of the likelihood with respect to the state, $\frac{\mathrm{d}^2 V_t(\boldsymbol{h})}{\mathrm{d}\boldsymbol{h}\boldsymbol{h}'}$, as the *score* quantity, and the second derivative $\frac{\mathrm{d}^2 V_t(\boldsymbol{h})}{\mathrm{d}\boldsymbol{h}\boldsymbol{h}'}$ as the *realized information*. The Newton optimization method (Nocedal and Wright, 2006) is online applicable if $\boldsymbol{I}_{t|t-1} - \frac{\mathrm{d}^2\ell(\boldsymbol{y}_t|\boldsymbol{h})}{\mathrm{d}\boldsymbol{h}\mathrm{d}\boldsymbol{h}'}$ is positive definite. When this is not the case, we can use the Fisher optimization step:

$$\boldsymbol{h}_{t|t}^{(i+1)} = \boldsymbol{h}_{t|t}^{(i)} + \left\{\boldsymbol{I}_{t|t-1} + E\left[-\frac{\mathrm{d}^2\ell(\boldsymbol{y}_t|\boldsymbol{h})}{\mathrm{d}\boldsymbol{h}\mathrm{d}\boldsymbol{h}'}\bigg|\boldsymbol{h}\right]\right\}\left\{\frac{\mathrm{d}\ell(\boldsymbol{y}_t|\boldsymbol{h})}{\mathrm{d}\boldsymbol{h}} - \boldsymbol{I}_{t|t-1}(\boldsymbol{h} - \boldsymbol{h}_{t|t-1})\right\}\Big|_{\boldsymbol{h}=\boldsymbol{h}_{t|t}^{(i)}}, \tag{46}$$

$$\boldsymbol{I}_{t|t} = \boldsymbol{I}_{t|t-1} + E\left[-\frac{\mathrm{d}^2\ell(\boldsymbol{y}_t|\boldsymbol{h})}{\mathrm{d}\boldsymbol{h}\mathrm{d}\boldsymbol{h}'}\bigg|\boldsymbol{h}\right]. \tag{47}$$

Here, we define $E\left[-\frac{\mathrm{d}^2\ell(\boldsymbol{y}_t|\boldsymbol{h})}{\mathrm{d}\boldsymbol{h}\mathrm{d}\boldsymbol{h}'}\bigg|\boldsymbol{h}\right]$ as the *expected information*. These derivations are the basis of the Bellman filter used in this paper. For the complete derivations, as well as a motivation, we encourage the reader to read Lange (2020). The algorithm of the Bellman filter is given in Algorithm 3.

Algorithm 3: Bellman filter algorithm from Lange (2020) with Newton optimization

Set $\boldsymbol{h}_{0|0} = (1 - \boldsymbol{\Phi})^{-1}\boldsymbol{c}$ and $\text{vec}(\boldsymbol{I}_{0|0}^{-1}) = (1 - \boldsymbol{\Phi} \otimes \boldsymbol{\Phi})^{-1}\text{vec}(\boldsymbol{Q})$. Set $t = 1$.

for $t = 1, \ldots, T$ **do**

$\boldsymbol{h}_{t|t-1} = c + \phi \boldsymbol{h}_{t-1|t-1}$

$\boldsymbol{I}_{t|t-1} = \boldsymbol{Q}^{-1} - \boldsymbol{Q}^{-1}\boldsymbol{\Phi}(\boldsymbol{I}_{t-1|t-1} + \boldsymbol{\Phi}'\boldsymbol{Q}^{-1}\boldsymbol{\Phi})\boldsymbol{\Phi}'\boldsymbol{Q}^{-1}$

Set $\boldsymbol{h}_{t|t}^{0} = \boldsymbol{h}_{t|t-1}$. Set $i = 0$.

while $i \leq i_{max}$ ***and*** *not converged* **do**

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

end

$\boldsymbol{I}_{t|t} = \boldsymbol{I}_{t|t-1} - \frac{\mathrm{d}^2 \ell(\boldsymbol{y}_t|\boldsymbol{h})}{\mathrm{d}\boldsymbol{h}\mathrm{d}\boldsymbol{h}'} \Big|_{\boldsymbol{h}=\boldsymbol{h}_{t|t}^{(i)}}$

end

3.3.2 Parameter estimation

As before, the model parameters $\boldsymbol{\theta}$ are actually unknown and have to be estimated, while also estimating the states $\boldsymbol{h}_{1:t}$. Now, if the model has linear Gaussian state dynamics, Lange (2020) proposes decomposing the log-likelihood in terms 'fit' generated by the Bellman filter, penalized by the realized KullBack-Leibler (KL) (Kullback and Leibler, 1951) divergence between filtered and predicted states, such that we can use the output of the Bellman filter without the need of complementary techniques. Lange (2020) derives an approximation for the KL divergence, which leads to the following 'likelihood minus KL divergence' decomposition:

$$\hat{\boldsymbol{\theta}} = \arg\max_{\boldsymbol{\theta}} \sum_{t=1}^{T} \left\{ \ell(\boldsymbol{y}_t|\boldsymbol{h}_{t|t} + \frac{1}{2}\log\det(\boldsymbol{I}_{t|t}^{-1}\boldsymbol{I}_{t|t-1}) - \frac{1}{2}(\boldsymbol{h}_{t|t} - \boldsymbol{h}_{t|t-1})'\boldsymbol{I}_{t|t-1}(\boldsymbol{h}_{t|t} - \boldsymbol{h}_{t|t-1}) \right\}. \quad (48)$$

This can be maximized using regular optimization techniques also used for the Kalman filter and particle filter. As we are using the unconditional distribution of the states to initialize the model, the mode exists at time $t = 1$. [7]

3.4 Simulation studies

This section investigates the filtering and estimating performances of the three methods mentioned above on models with various distributions. First, we investigate the performance of the three methods on a linear Gaussian model: the Random Walk plus Noise model, also called the Gaussian local level model. This is a model that the Kalman filter can exactly evaluate, thus gives optimal results. Lange (2020) shows that the Bellman filter is able to give the exact same result with just

[7] When the unconditional distribution is not available, Lange (2020) proposes to start the maximization from $t_0 \geq 0$, where t_0 is large enough to ensure the mode exists.

Table 1: Performance of filtering and estimating Gaussian Local Level models

Panel A: Mean absolute errors of one-step-ahead predictions of state variable				
	True parameters		Estimated parameters	
Method	MAE	Relative MAE	MAE	Relative MAE
Kalman filter	0.2195	1.0000	0.2199	1.0000
Particle filter	0.2199	1.0019	0.2204	1.0020
Bellman filter	0.2195	1.0000	0.2199	1.0000

Panel B: Mean and mean absolute errors of estimated model parameters				
Method	c	ϕ	σ_n	σ_y
Kalman filter	-0.001	0.978	0.150	0.448
	(0.002)	(0.004)	(0.006)	(0.005)
Particle filter	-0.001	0.978	0.150	0.448
	(0.002)	(0.004)	(0.006)	(0.005)
Bellman filter	-0.001	0.978	0.150	0.448
	(0.002)	(0.004)	(0.006)	(0.005)

NOTE: We simulate 100 time series of length $T = 5,000$ for the Gaussian local level model. We use parameter values $c = 0, \phi = 0.98, \sigma_n = 0.15$. Then, we use the true values of the parameters to predict the state variable one-step-ahead and compute the mean absolute error (MAE) as the distance between the one-step-ahead prediction and the true value, over the last 2,500 observations. MAEs are presented in Panel A. After, we use the first $2,500$ observations to estimate the model and compute MAEs of the estimated parameters in Panel B. Finally, we use the estimated parameters to predict the one-step-ahead state variable for the last $2,500$ observations and compute the corresponding MAEs, which are presented in Panel A.

one iteration, we expect it to have the same performance. The local level model is given by:

$$\begin{aligned} y_t &= \mu_t + \varepsilon_t, \qquad \varepsilon_t \sim \mathcal{N}(0, \sigma_y^2), \\ \mu_t &= c + \phi\mu_{t-1} + \eta_t, \qquad \eta_t \sim \mathcal{N}(0, \sigma_\eta^2). \end{aligned} \tag{49}$$

In Table 1, we present the results of this simulation. We find the Kalman filter and the Bellman filter to give the same results, as was expected. Furthermore, the particle filter performs almost exactly equal to these methods, showing that it can compute with the Kalman filter even for the optimal linear Gaussian model. However, the estimation time of the particle filter is significantly larger.

3.4.1 Design

To further investigate the performances of the Kalman filter, particle filter, and Bellman filter, we follow the Monte Carlo study from Lange (2020). We conduct a Monte Carlo study for ten data-generating processes (DGPs), nine of them from Koopman et al. (2016) and the final from Lange (2020), all with linear Gaussian state dynamics and observation densities as in Table 2. The simulation study is equal to that of Lange (2020). However, we compare the performance of the new Bellman filter to that of the particle filter, whereas Lange (2020) compares it to the

numerically accelerated importance sampling (NAIS) method of Koopman et al. (2016).

For each of the GDPs, we simulate 100 time series of length 5000.[8] For the first nine, the parameter values are obtained from Koopman et al. (2016), and for the last DGP the parameter values are from Lange (2020). The state-transition parameters are $c = 0, \phi = 0.98, \sigma_n = 0.15$ for eight of the models. For the dependence models we use parameters $c = 0.2, \phi = 0.98, \sigma_n = 0.1$. Furthermore, Student's t-distributions have degrees of freedom $\nu = 10$ except for the local-level t model, which has degrees of freedom $\nu = 3$. The shape parameters have value $k = 4$ for the negative binomial distribution, $k = 1.5$ for the Gamma distribution, $k = 1.2$ for the Weibull distribution, and $\sigma_y = 0.45$ for the local-level t model. The goal of the simulation study is twofold.

First, we investigate only the filtering performance of the methods. We use the true parameter values and produce one-step-ahead predictions of the quantities of interest, which come from the link functions in Table 2 for timesteps $t = 2501, \ldots, 5000$. Then, we compute the mean absolute error (MAE) between the predicted quantity and its true simulated value. As we have 2500 predictions for 100 time series, this method results in an average error of 250,000 predictions.

Second, we investigate the performance of the methods when estimating the parameters is necessary. We use the first 2500 observations to estimate the model parameters. We then compute the means and MAEs of the estimated parameters from the true values. For each parameter, we have one estimation for each DGP, such that we take the mean and mean absolute error of 100 estimations. Then, like before, we produce one-step-ahead predictions of the quantities of interest, but now using the estimated parameters. After which, we again compute MAEs of 250,000 predictions. The three methods are used as follows:

1. The Kalman filter (see Section 3.1) can only be applied for linear models. To use the Kalman filter for the stochastic volatility models, we transform the observations by taking squares and logarithms to obtain a linear state-space model, as is often done in literature (Ruiz, 1994; Harvey et al., 1994). The resulting observation equations are non-Gaussian but can be estimated using quasi maximum likelihood (QML) with the Kalman filter. As the local-level t model is linear but non-Gaussian, it can be estimated directly through QML.

2. The algorithm to use the Bellman filter from Lange (2020) (see Section 3.3) is presented in Algorithm 3. We initialize the model using the unconditional distribution. For the dependence models and the local-level t model the realized information quantity might be negative, in which case we use the Fisher updating step to ensure $\boldsymbol{I}_{t|t} \geq \boldsymbol{I}_{t|t}$. Following Lange (2020), we stop the

[8] We deviate slightly from the simulations in Lange (2020) and Koopman et al. (2016) as we only use 100 times series. This is due to the large estimation time of the particle filter

Table 2: Overview of data-generating processes in simulation studies

DGP Model type	Distribution	**Link function**	**Density** $p(\boldsymbol{y}_t \mid h_t)$
Count	Poisson	$\lambda_t = \exp(h_t)$	$\lambda_t^{y_t} \exp(-\lambda_t)/y_t!$
Count	Negative binomial	$\lambda_t = \exp(h_t)$	$\frac{\Gamma(k+y_t)}{\Gamma(k)\Gamma(y_t+1)} \left(\frac{k}{k+\lambda_t}\right)^k \left(\frac{\lambda_t}{k+\lambda_t}\right)^{y_t}$
Intensity	Exponential	$\lambda_t = \exp(h_t)$	$\lambda_t \exp(-\lambda_t y_t)$
Duration	Gamma	$\beta_t = \exp(h_t)$	$\frac{1}{\Gamma(k)\beta_t^k} y_t^{k-1} \exp(-y_t/\beta_t)$
Duration	Weibull	$\beta_t = \exp(h_t)$	$\frac{k}{\beta_t} \left(\frac{y_t}{\beta_t}\right)^{k-1} \exp(-(y_t/\beta_t)^k)$
Volatility	Gaussian	$\sigma_t^2 = \exp(h_t)$	$\frac{1}{\sqrt{2\pi\sigma_t}} \exp(-y_t^2/(2\sigma_t^2))$
Volatility	Student's t	$\sigma_t^2 = \exp(h_t)$	$\frac{\Gamma(\frac{\nu+1}{2})}{\sqrt{(\nu-2)\pi}\Gamma(\frac{\nu}{2})\sigma_t} \left(1 + \frac{y_t^2}{(\nu-2)\sigma_t^2}\right)^{-\frac{\nu+1}{2}}$
Dependence	Gaussian	$\rho_t = \frac{1-\exp(-h_t)}{1+\exp(-h_t)}$	$\frac{\frac{1}{2\pi\sqrt{1-\rho_t^2}} \exp\left\{-\frac{y_{1t}^2+y_{2t}^2-2\rho_t y_{1t} y_{2t}}{2(1-\rho_t^2)}\right\}}{\prod_{i=1}^2 \frac{1}{2\pi} \exp(-y_{it}^2/2)}$
Dependence	Student's t	$\rho_t = \frac{1-\exp(-h_t)}{1+\exp(-h_t)}$	$\frac{\Gamma(\frac{\nu+2}{2})\Gamma(\frac{\nu}{2})}{\Gamma(\frac{\nu+1}{2})^2} \frac{\frac{1}{\sqrt{1-\rho_t^2}} \left(1+\frac{y_{1t}^2+y_{2t}^2-2\rho_t y_{1t} y_{2t}}{\nu(1-\rho_t^2)}\right)^{\frac{\nu+2}{2}}}{\prod_{i=1}^2 (1+y_{it}^2/\nu)^{-\frac{\nu+1}{2}}}$
Local level	Student's t	$\mu_t = h_t$	$\frac{\Gamma(\frac{\nu+1}{2})}{\sqrt{(\nu-2)\pi}\Gamma(\frac{\nu}{2})\sigma_y} \left(1 + \frac{(y_t-\mu_t)^2}{(\nu-2)\sigma_y^2}\right)^{-\frac{\nu+1}{2}}$

NOTE: This table contains the distributions and link functions of then data-generating processes (DGPs) used in the simulation study of Lange (2020). The first nine DGPs come from Koopman et al. (2016). For each DGP, we simulate 100 state-space time series of length 5,000 with state variable h_t from a linear Gaussian state equation. The link function show how the state variable is transformed when used to the observation equation. Finally, we to use the Bellman filter we compute scores and information quantities for each of the GDPs. For these quantities we refer to Table 3 of Lange (2020).

optimization step when either the convergence criterion $|h_{t|t}^i - h_{t|t}^{i-1}| < 0.0001$ or the maximum number of iterations $i_{\max} = 40$ is reached. For the score and information quantities of the DGP processes, we refer to Table 3 of Lange (2020). To obtain predictions for the quantities of interest, Bellman predicted states $h_{t|t-1}$ are transformed using the link functions of Table 2. To have an exact transformation of the predictions using the link (monotone) functions, the predictions should be based on the median. However, the Bellman filter is based on the mode. Following Lange (2020), we ignore this for simplicity.

3. We use the continuous sampling importance resampling (CSIR) algoritm from Malik and Pitt (2011) as the particle filter (see Section 3.2). We use number of particles $N = 1,000$.[9] At each time step, we compute one-step-ahead predictions of the quantities of interest by first transforming all predicted particles with the link function, and then taking the median.

[9]We found that using larger N did not noticeably improve performance, and using $N = 1000$ gave consistent results.

Table 3: Mean absolute errors (MAEs) and relative mean absolute erros of one-step-ahead predictions in simulation studies

		Bellman filter		**Particle filter**		**Kalman filter**	
Type	Distribution	True	Estimated	True	Estimated	True	Estimated
Count	Poisson	0.3616	0.9977	0.9947	0.9984	n/a	n/a
Count	Neg. Bin.	0.3838	1.0055	0.9989	1.0058	n/a	n/a
Intensity	Exponential	0.4029	1.0000	1.0011	0.9931	n/a	n/a
Duration	Gamma	0.3652	1.0039	1.0050	0.9947	n/a	n/a
Duration	Weibull	0.3714	0.9988	0.9996	0.9931	n/a	n/a
Volatility	Gaussian	0.1877	0.9990	0.9992	0.9921	1.1714	1.1627
Volatility	Student's t	0.1958	1.0100	1.0074	0.9973	1.1392	1.1282
Dependence	Gaussian	0.1198	1.0267	1.0040	0.9906	n/a	n/a
Dependence	Student's t	0.1225	1.0192	0.9958	0.9898	n/a	n/a
Local level	Student's t	0.2009	1.0150	0.9984	0.9958	1.0746	1.0715

NOTE: We simulate 100 time series of length 5,000 for the 10 data-generating processes shown in Table 2 with linear Gaussian state dynamics. We first evaluate the filtering performance by computing one-step-ahead predictions of the quantity of interest, given by the link function in Table 2, using the true model parameters. We compute the mean absolute errors (MAEs) by taking the distance of the predicted quantity of interest from its true simulated value over the last 2,500 observations. Results for each filtering method are presented in the columns 'True'. Then we evaluate the estimation performance of the methods by using the first 2,500 observations to estimate the parameter values. We then use these estimated parameters to make one-step-ahead predictions over the last 2,500 observations and again compute MAEs of the quantities of interest. Relative MAEs of the estimation methods are shown in the columns 'Estimated'.

3.4.2 Results

Table 3 shows the MAEs of one-step-ahead predictions when using true parameters and estimated parameters. To compute MAEs, we have used the median of the predicted particles in the particle filter. The median being optimal for the MAE loss function. The Bellman filter prediction is based on the mode, and therefore not optimal.

Using the true parameter values, we find the particle filter and Bellman filter to perform very closely. The Bellman filter performs slightly better, with an MAE of 0.1% to 1% lower than the particle filter, for four DGPs. The particle filter has lower MAEs for six DGPs, all of them with a difference less than 0.5% of the Bellman filter. Finally, the Kalman filter performs much worse than the other methods, from a 7% higher MAE for the Local level t to a 17% higher MAE for the Gaussian stochastic volatility model. As the models are non-Linear and/or non-Gaussian and estimated by quasi maximum likelihood, this was to be expected.

Most of the time, the true parameters of the model are unknown and have to be estimated beforehand. The results in Table 3 show that the particle filter performs slightly better when the parameters have to be estimated. For all DGPs, except the Negative Binomial, does the particle filter perform around 0.5% to 1% better than the Bellman filter. For the Negative Binomial DGP, the Bellman filter performs around 0.5% better. Again, the Kalman filter cannot compete with

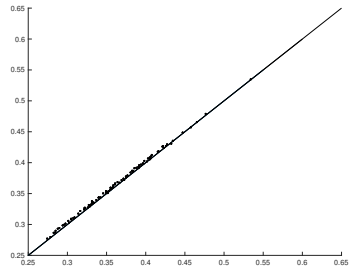
Table 4: Mean and mean absolute errors (MAEs) of estimated parameters in simulation studies

		Bellman filter				**Particle filter**			
Type	Distribution	c	ϕ	σ_n	k/ν	c	ϕ	σ_n	k/ν
Count	Poisson	-0.008	0.977	0.153	n/a	-0.000	0.978	0.151	n/a
		(0.008)	(0.005)	(0.012)	n/a	(0.002)	(0.005)	(0.012)	n/a
Count	Neg. Bin.	-0.004	0.980	0.148	4.221	-0.000	0.978	0.151	4.056
		(0.004)	(0.005)	(0.012)	(0.474)	(0.002)	(0.005)	(0.012)	(0.314)
Intensity	Exponential	-0.008	0.976	0.160	n/a	-0.000	0.978	0.153	n/a
		(0.008)	(0.006)	(0.015)	n/a	(0.002)	(0.005)	(0.011)	n/a
Duration	Gamma	0.007	0.977	0.156	1.504	-0.000	0.978	0.151	1.498
		(0.007)	(0.005)	(0.011)	(0.033)	(0.003)	(0.005)	(0.010)	(0.032)
Duration	Weibull	0.009	0.976	0.160	1.206	-0.000	0.978	0.152	1.198
		(0.008)	(0.006)	(0.011)	(0.019)	(0.002)	(0.004)	(0.011)	(0.017)
Volatility	Gaussian	0.007	0.975	0.163	n/a	0.000	0.978	0.151	n/a
		(0.007)	(0.007)	(0.020)	n/a	(0.002)	(0.005)	(0.015)	n/a
Volatility	Student's t	0.004	0.975	0.161	14.398	0.000	0.977	0.153	10.342
		(0.005)	(0.007)	(0.022)	(5.008)	(0.003)	(0.006)	(0.017)	(0.492)
Dependence	Gaussian	0.016	0.983	0.064	n/a	0.025	0.975	0.105	n/a
		(0.008)	(0.007)	(0.037)	n/a	(0.009)	(0.009)	(0.019)	n/a
Dependence	Student's t	0.013	0.988	0.007	13.139	0.029	0.970	0.121	16.328
		(0.008)	(0.009)	(0.023)	(3.237)	(0.012)	(0.013)	(0.032)	(6.328)
Local level*	Student's t	-0.000	0.982	0.121	3.913	-0.000	0.977	0.150	3.015
		(0.002)	(0.004)	(0.029)	(0.914)	(0.003)	(0.005)	(0.006)	(0.169)

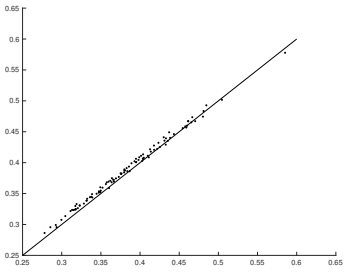
NOTE: We simulate 100 time series of length 5,000 for the 10 data-generating processes shown in Table 2 with linear Gaussian state dynamics. We use the first 2,500 observations to estimate the parameter values according to the methods presented in Sections 3.1, 3.2, and 3.3. We then compute the mean of the 100 estimated parameters and compute mean absolute errors (MAEs) using the distance between the estimated parameters and their true value. The mean of the estimations is presented first with the MAEs in brackets below. *For the Local level t, we also estimate parameter σ_y. Both models estimate this parameter accurately and with similar errors as the estimation of σ_n.

the two other methods, with MAEs of 7% to 16% higher. Looking at the estimated parameters over the same estimation period of 2500 observations, we find both the Bellman filter and particle filter to perform very well. The level of hidden state c, which is 0 in the DGPs, is estimated very well for almost every DGP. Only for the two dependence models do the Bellman filter and particle filter estimate the parameter a little too high. Furthermore, state transition parameter ϕ and state volatility σ_n are estimated very accurate. The degrees of freedom ν are somewhat harder to estimate, with the Bellman filter predicting a too high degrees of freedom for the stochastic volatility Student's t model, while the particle filter does the same for the dependence Student's t model. We conclude that both models very accurate, with the particle filter being slightly more precise.

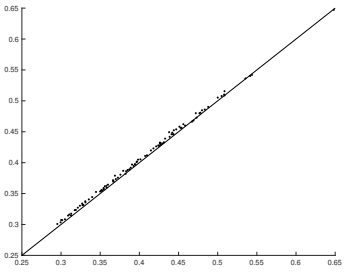
Figure 1: Scatter plots of 100 out-of-sample MAEs for each DGP after estimating the model parameters, together with a 45-degree line. Bellman filter errors are presented horizontally, particle filter errors are presented vertically.



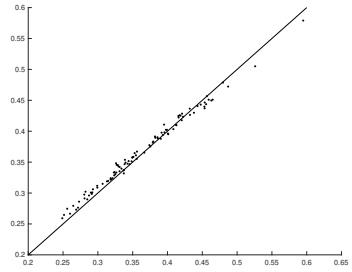
(a) Poisson



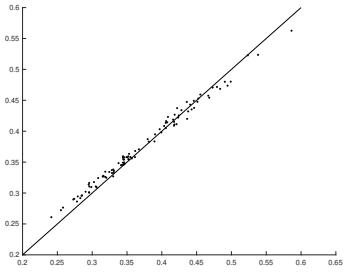
(b) Negative Binomial



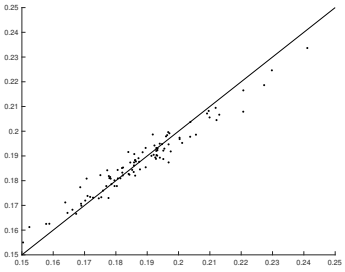
(c) Exponential



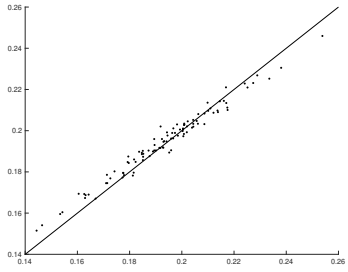
(d) Gamma



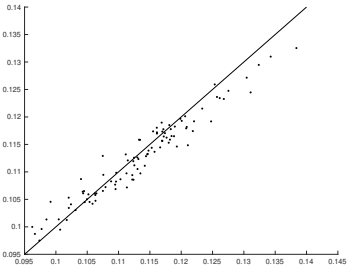
(e) Weibull



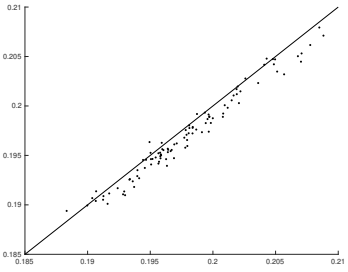
(f) Volatility Gaussian



(g) Volatility Student's t



(h) Dependence Gaussian



(i) Local level Student's t

NOTE: We simulate 100 time series of length 5,000 for the 10 data-generating processes (DGPs) shown in Table 2 with linear Gaussian state dynamics. We use the first 2,500 observations to estimate the parameters after which use these estimations to make one-step-ahead predictions of the quantities of interest over the last 2,500 observations. See the note of Table 3 for more detail. These plots show the mean absolute errors (MAEs) for each of the 100 time series of the 10 DGPs. The MAEs of the Bellman filter are shown on the horizontal axis and the MAEs of the particle filter on the vertical axis. A 45-degree line is also added for reference. Dots below the 45-degree line indicate a lower MAE for the particle filter than for the Bellman filter, and above the 45-degree line vice versa.

Table 5: Average computation times in seconds

Type	Distribution	Bellman filter		Particle filter	
		Filtering	Estimation	Filtering	Estimation
Count	Poisson	0.002	0.110	1.037	27.305
Count	Neg. Bin.	0.003	0.365	1.304	80.047
Intensity	Exponential	0.002	0.150	0.806	20.028
Duration	Gamma	0.003	0.302	1.211	66.808
Duration	Weibull	0.007	0.443	1.545	82.501
Volatility	Gaussian	0.004	0.227	0.976	23.737
Volatility	Student's t	0.005	0.511	1.274	53.718
Dependence	Gaussian	0.007	0.849	0.960	29.376
Dependence	Student's t	0.007	1.464	1.319	131.429
Local level	Student's t	0.005	1.342	0.968	127.018

NOTE: We simulate 100 time series of length 5,000 for the 10 data-generating processes shown in Table 2 with linear Gaussian state dynamics. This table shows the time the Bellman filter and particle filter need to filter the time series of length 5,000 under the columns 'Filtering'. Furthermore, this table shows the time needed to estimate the parameters for the time series of length 2,500. The optimization is for both methods is done in MATLAB using function fminunc with identical settings. For the particle filter, we use number of particles N = 1000. Not all simulations are run on the same computer. Thus, times across different DGPs are not comparable. Simulations for the same DGP are done on the same computer, such that estimation and filtering times within a DGP are comparable.

Figure 1 shows the MAEs per simulated time series, such that for each DGP, we have 100 errors. The mean errors of the Bellman filter is shown on the horizontal axis and the mean errors of the particle filter on the vertical axis. We find that the errors concentrate around the 45-degree line, indicating that, from a predictive point of view, the simulated time series is of much more importance than the estimation method used. Looking closer, we find that the errors with a lower value generally lie more on the left of the 45-degree line, while larger errors generally lie on the right of the 45-degree line. This indicates that the Bellman filter performs slightly better when average errors are lower, while the particle filter performs slightly better when average errors are higher.

So far, we have seen the Bellman filter and particle filter perform very comparably, with the particle filter performing slightly better in estimation. What differentiates the Bellman filter and particle filter is the computational time necessary for the methods. Table 5 present the average computation time of filtering the complete time series of length 5000 and the average computation time of parameter estimation on the estimation samples of length 2500. We find the Bellman filter to be far superior from the computational point of view, with maximum filtering times of less than 0.01 seconds. The particle filter, on the other hand, needs around one full second for filtering. Furthermore, the estimation of one time series takes on average between 0.1 and 1.5 seconds for the Bellman filter, depending on the DGP. Average estimating times for the particle filter range from 20 seconds for the exponential DGP to 131 seconds for the dependence Student's

t DGP. Overall, the Bellman filter is faster in both filtering and estimation by more than a factor 100 for most of the DGPs. Concluding, while the methods are close when filtering and estimating, with a slight preference for the particle filter, the Bellman filter is computationally far superior to the particle filter.

4 Estimating stochastic volatility models

In this section, we present the estimation methods for the stochastic volatility models presented in Section 2. The models that we estimate are the Catania model from Equation (18) and the new stochastic volatility model from Equation (20). The classical stochastic volatility model from Equation (1) (no leverage), the contemporaneous stochastic volatility model from Equation (8) ($m = 0$), and the inter-temporal stochastic volatility model from Equation (3) ($m = 1$, no contemporaneous leverage) are nested in these models.

4.1 Estimating the Catania model with the particle filter

We follow Catania (2020) in estimating the Catania model with the particle filter. To do so, we write the model in the following nonlinear state-space representation:

$$\begin{aligned} y_t &= \exp\left\{\frac{h_t}{2}\right\}\left[\frac{\rho_0}{\sigma_n(1-\sum_{j=1}^m \rho_j^2)}\left(h_t - c - \phi h_{t-1} - \sigma_n \sum_{j=1}^m \rho_j y_{t-j} \exp\left\{-\frac{h_{t-j}}{2}\right\}\right)\right] \\ &\quad + \exp\left\{\frac{h_t}{2}\right\}\sqrt{1 - \frac{\rho_0^2}{1-\sum_{j=1}^m \rho_j^2}}u_t, \\ h_t &= c + \phi h_{t-1} + \sigma_n \sum_{j=1}^m \rho_j y_{t-j} \exp\left\{-\frac{h_{t-j}}{2}\right\} + \sigma_n \sqrt{\sum_{j=1}^m \rho_j^2} w_t, \end{aligned} \tag{50}$$

where u_t and w_t are standard normally distributed, $u_t \sim \mathcal{N}(0,1), w_t \sim \mathcal{N}(0,1)$, and are uncorrelated $E[u_t w_t] = 0$. The derivation of this state-space representation is given in Appendix A. This representation with $m = 1$ and $\rho_0 \neq 0$ is a correct state-space representation of the two-leverage model by Yu (2005), for which we earlier showed the errors were not independent and uncorrelated. The conditional distribution of $y_t|h_t, \mathcal{I}_{t-1}$ is normal with the following mean and

variance:

$$
\begin{aligned}
&y_t|h_t, \mathcal{I}_{t-1} \sim \mathcal{N}(\mu_{yt}, \sigma_{yt}^2), \\
&\quad \mu_{yt} = \exp\left\{\frac{h_t}{2}\right\}\left[\frac{\rho_0}{\sigma_n(1-\sum_{j=1}^m \rho_j^2)}\left(h_t - c - \phi h_{t-1} - \sigma_n \sum_{j=1}^m \rho_j y_{t-j} \exp\left\{-\frac{h_{t-j}}{2}\right\}\right)\right] \\
&\quad \sigma_{yt}^2 = \exp\{h_t\}\left(1 - \frac{\rho_0^2}{\sum_{j=1}^m \rho_j^2}\right),
\end{aligned}
\tag{51}
$$

and the conditional distribution of $h_t|\mathcal{I}_{t-1}$ is given by:

$$
\begin{aligned}
&h_t|\mathcal{I}_{t-1} \sim \mathcal{N}(\mu_{ht}, \sigma_{ht}^2) \\
&\quad \mu_{ht} = c + \phi h_{t-1} + \sigma_n \sum_{j=1}^m \rho_j y_{t-j} \exp\left\{-\frac{h_{t-j}}{2}\right\}, \\
&\quad \sigma_{ht}^2 = \sigma_n^2\left(\sum_{j=1}^m \rho_j^2\right).
\end{aligned}
\tag{52}
$$

This model can be estimated using the particle filter from Malik and Pitt (2011) presented in Section 3.2. The likelihood can be obtained with the prediction error decomposition:

$$
\mathcal{L}(\boldsymbol{\theta}) = \prod_{t=1}^T p(y_t|\mathcal{I}_{t-1}; \boldsymbol{\theta}) = \prod_{t=1}^T p(y_t|h_t, \mathcal{I}_{t-1}; \boldsymbol{\theta}) p(h_t|\mathcal{I}_{t-1}; \boldsymbol{\theta}) dh_t. \tag{53}
$$

Which we approximate using the particle filter by:

$$
p(y_t|\mathcal{I}_{t-1}; \boldsymbol{\theta}) \approx \frac{1}{N}\sum_{i=1}^N p(y_t|h_t^i, \mathcal{I}_{t-1}), \tag{54}
$$

where h_t^i, $i = 1, \ldots, N$ are draws from the predictive distribution $p(h_t|\mathcal{I}_{t-1})$ in the particle filter. [10]

4.2 Estimating the new stochastic volatility model with the Bellman filter

The new stochastic volatility model from Equation (20) is estimated with the Bellman filter from Lange (2020). In this model, the state equation for h_t is linear and Gaussian, and therefore the model can easily be estimated by the Bellman filter. [11] We write the model in the following

[10] For the first steps in the particle filter, past observations $y_0, y_{-1}, \ldots, y_{-m+1}$ are needed. We solve this by moving the start of the sample by m steps, and draw the first m log volatilities from the unconditional distribution. Catania (2020) noted that different sampling schemes have no effect on the behavior of the algorithm.

[11] This model could also be estimated using the particle filter. However, the model is a multivariate state-space model, for which estimation with the particle filter cannot be done with the particle filtering method described above, but would require additional more intensive and less accurate techniques.

nonlinear state-space representation:

$$\begin{aligned} y_t &= \exp\left\{\frac{h_t}{2}\right\}\left(\sum_{j=0}^{m}\rho_j\eta_{t+j} + \sqrt{1-\sum_{j=0}^{m}\rho_j^2}\xi_t\right), \qquad \xi_t \sim \mathcal{N}(0,1) \\ \boldsymbol{\alpha}_t &= \boldsymbol{c} + \boldsymbol{\Phi}\boldsymbol{\alpha}_{t-1} + \boldsymbol{\xi}_t, \qquad \boldsymbol{\Xi}_t \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{Q}), \end{aligned} \tag{55}$$

where we have the following definitions:

$$\boldsymbol{\alpha}_t = \begin{pmatrix} h_t \\ \eta_{t+m} \\ \eta_{t+m-1} \\ \dots \\ \eta_{t+1} \\ \eta_t \end{pmatrix}, \quad \boldsymbol{c} = \begin{pmatrix} c \\ 0 \\ 0 \\ \dots \\ 0 \\ 0 \end{pmatrix}, \quad \boldsymbol{\Phi} = \begin{pmatrix} \phi & 0 & 0 & \dots & \sigma_n & 0 \\ 0 & 0 & 0 & \dots & 0 & 0 \\ 0 & 1 & 0 & \dots & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & 0 & 0 \\ 0 & 0 & 0 & \dots & 1 & 0 \end{pmatrix},$$

$$\boldsymbol{\xi} = \begin{pmatrix} 0 \\ \xi \\ 0 \\ \dots \\ 0 \\ 0 \end{pmatrix}, \quad \boldsymbol{Q} = \begin{pmatrix} 0 & 0 & 0 & \dots & 0 & 0 \\ 0 & 1 & 0 & \dots & 0 & 0 \\ 0 & 0 & 0 & \dots & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & 0 & 0 \\ 0 & 0 & 0 & \dots & 0 & 0 \end{pmatrix}.$$

The conditional distribution of $y_t|\boldsymbol{\alpha}_t, \mathcal{I}_{t-1}$ is nonlinear and Gaussian, given by:

$$\begin{aligned} y_t|\boldsymbol{\alpha}_t, \mathcal{I}_{t-1} &\sim \mathcal{N}(\mu_{yt}, \sigma_{yt}^2), \\ \mu_{yt} &= \exp\left\{\frac{h_t}{2}\right\}\left(\sum_{j=0}^{m}\rho_j\eta_{t+j}\right) \\ \sigma_{yt}^2 &= \exp\{h_t\}\left(1-\sum_{j=0}^{m}\rho_j^2\right). \end{aligned} \tag{56}$$

The conditional distribution of the hidden states $\boldsymbol{\alpha}_t|\mathcal{I}_{t-1}$ is linear and Gaussian, as is necessary to implement Algorithm 3 from Lange (2020). However, we have to adjust Algorithm 3 slightly to implement this state-space model. Due to the unusual state-space representation with singular variance matrix $\boldsymbol{Q}$, the prediction step for the information matrix has to be adjusted to $\boldsymbol{I}_{t|t-1} = (\boldsymbol{\Phi}\boldsymbol{I}_{t-1|t-1}^{-1}\boldsymbol{\Phi}' + \boldsymbol{Q})^{-1}$.[12] Apart from these adjustments, the Bellman filter can be used as before. From the conditional distribution of y_t in (57) we can derive the conditional probability density

[12] See Lange (2020) for more details.

of y_t:

$$p(y_t|\boldsymbol{\alpha}_t) = \frac{1}{\sqrt{2\pi\sigma_{yt}^2}} \exp\left\{\frac{1}{2\sigma_{yt}^2} - (y_t - \mu_{yt})^2\right\}. \tag{57}$$

This leads the following score, realized information and expected information quantities:

$$\begin{aligned}\frac{\mathrm{d}\ell(y_t|\boldsymbol{\alpha}_t)}{\mathrm{d}\boldsymbol{\alpha}_t} &= \left[\frac{\mathrm{d}\ell(y_t|\boldsymbol{\alpha}_t)}{\mathrm{d}h_t}, \frac{\mathrm{d}\ell(y_t|\boldsymbol{\alpha}_t)}{\mathrm{d}\eta_{t+m}}, \ldots, \frac{\mathrm{d}\ell(y_t|\boldsymbol{\alpha}_t)}{\mathrm{d}\eta_t}\right] \\ &= \left[-\frac{1}{2} + \frac{1}{2\sigma_{yt}^2} y_t(y_t - \mu_{yt}), \frac{\rho_m}{\sqrt{1-\sum_{j=0}^m \rho_j^2}} \frac{y-\mu_{yt}}{\sigma_{yt}}, \ldots, \frac{\rho_0}{\sqrt{1-\sum_{j=0}^m \rho_j^2}} \frac{y-\mu_{yt}}{\sigma_{yt}}\right]',\end{aligned} \tag{58}$$

$$\begin{aligned}-\frac{\mathrm{d}^2\ell(y_t|\boldsymbol{\alpha}_t)}{\mathrm{d}\boldsymbol{\alpha}_t \mathrm{d}\boldsymbol{\alpha}_t'} &= \begin{bmatrix} -\frac{\mathrm{d}^2\ell(y_t|\boldsymbol{\alpha}_t)}{\mathrm{d}h_t^2} & -\frac{\mathrm{d}^2\ell(y_t|\boldsymbol{\alpha}_t)}{\mathrm{d}h_t \mathrm{d}\eta_{t+m}} & \cdots & -\frac{\mathrm{d}^2\ell(y_t|\boldsymbol{\alpha}_t)}{\mathrm{d}h_t \mathrm{d}\eta_t} \\ -\frac{\mathrm{d}^2\ell(y_t|\boldsymbol{\alpha}_t)}{\mathrm{d}h_t \mathrm{d}\eta_{t+m}} & -\frac{\mathrm{d}^2\ell(y_t|\boldsymbol{\alpha}_t)}{\mathrm{d}\eta_{t+m}^2} & \cdots & -\frac{\mathrm{d}^2\ell(y_t|\boldsymbol{\alpha}_t)}{\mathrm{d}\eta_{t+m} \mathrm{d}\eta_t} \\ \cdots & \cdots & \cdots & \cdots \\ -\frac{\mathrm{d}^2\ell(y_t|\boldsymbol{\alpha}_t)}{\mathrm{d}h_t \mathrm{d}\eta_t} & -\frac{\mathrm{d}^2\ell(y_t|\boldsymbol{\alpha}_t)}{\mathrm{d}\eta_{t+m} \mathrm{d}\eta_t} & \cdots & -\frac{\mathrm{d}^2\ell(y_t|\boldsymbol{\alpha}_t)}{\mathrm{d}\eta_t^2} \end{bmatrix} \\ &= \begin{bmatrix} -\frac{1}{4}\frac{\mu_{yt}y_t}{\sigma_{yt}^2} + \frac{1}{2}\frac{y_t^2}{\sigma_{yt}^2} & \frac{1}{2}\frac{\rho_m}{\sqrt{1-\sum_{j=0}^m \rho_j^2}} \frac{y_t}{\sigma_{yt}} & \cdots & \frac{1}{2}\frac{\rho_0}{\sqrt{1-\sum_{j=0}^m \rho_j^2}} \frac{y_t}{\sigma_{yt}} \\ \frac{1}{2}\frac{\rho_m}{\sqrt{1-\sum_{j=0}^m \rho_j^2}} \frac{y_t}{\sigma_{yt}} & \frac{\rho_m^2}{1-\sum_{j=0}^m \rho_j^2} & \cdots & \frac{\rho_0\rho_m}{\sqrt{1-\sum_{j=0}^m \rho_j^2}} \\ \cdots & \cdots & \cdots & \cdots \\ \frac{1}{2}\frac{\rho_0}{\sqrt{1-\sum_{j=0}^m \rho_j^2}} \frac{y_t}{\sigma_{yt}} & \frac{\rho_0\rho_m}{1-\sum_{j=0}^m \rho_j^2} & \cdots & \frac{\rho_0^2}{1-\sum_{j=0}^m \rho_j^2} \end{bmatrix}.\end{aligned} \tag{59}$$

$$\begin{aligned}E\left[-\frac{\mathrm{d}^2\ell(y_t|\boldsymbol{\alpha}_t)}{\mathrm{d}\boldsymbol{\alpha}_t \mathrm{d}\boldsymbol{\alpha}_t'}\middle|\boldsymbol{\alpha}_t\right] &= E\left[\begin{bmatrix} -\frac{\mathrm{d}^2\ell(y_t|\boldsymbol{\alpha}_t)}{\mathrm{d}h_t^2} & -\frac{\mathrm{d}^2\ell(y_t|\boldsymbol{\alpha}_t)}{\mathrm{d}h_t \mathrm{d}\eta_{t+m}} & \cdots & -\frac{\mathrm{d}^2\ell(y_t|\boldsymbol{\alpha}_t)}{\mathrm{d}h_t \mathrm{d}\eta_t} \\ -\frac{\mathrm{d}^2\ell(y_t|\boldsymbol{\alpha}_t)}{\mathrm{d}h_t \mathrm{d}\eta_{t+m}} & -\frac{\mathrm{d}^2\ell(y_t|\boldsymbol{\alpha}_t)}{\mathrm{d}\eta_{t+m}^2} & \cdots & -\frac{\mathrm{d}^2\ell(y_t|\boldsymbol{\alpha}_t)}{\mathrm{d}\eta_{t+m} \mathrm{d}\eta_t} \\ \cdots & \cdots & \cdots & \cdots \\ -\frac{\mathrm{d}^2\ell(y_t|\boldsymbol{\alpha}_t)}{\mathrm{d}h_t \mathrm{d}\eta_t} & -\frac{\mathrm{d}^2\ell(y_t|\boldsymbol{\alpha}_t)}{\mathrm{d}\eta_{t+m} \mathrm{d}\eta_t} & \cdots & -\frac{\mathrm{d}^2\ell(y_t|\boldsymbol{\alpha}_t)}{\mathrm{d}\eta_t^2} \end{bmatrix}\middle|\boldsymbol{\alpha}_t\right] \\ &= \begin{bmatrix} \frac{1}{4}\frac{\mu_{yt}^2}{\sigma_{yt}^2} + \frac{1}{2} & \frac{1}{2}\frac{\rho_m}{\sqrt{1-\sum_{j=0}^m \rho_j^2}} \frac{\mu_{yt}}{\sigma_{yt}} & \cdots & \frac{1}{2}\frac{\rho_0}{\sqrt{1-\sum_{j=0}^m \rho_j^2}} \frac{\mu_{yt}}{\sigma_{yt}} \\ \frac{1}{2}\frac{\rho_m}{\sqrt{1-\sum_{j=0}^m \rho_j^2}} \frac{\mu_{yt}}{\sigma_{yt}} & \frac{\rho_m^2}{1-\sum_{j=0}^m \rho_j^2} & \cdots & \frac{\rho_0\rho_m}{1-\sum_{j=0}^m \rho_j^2} \\ \cdots & \cdots & \cdots & \cdots \\ \frac{1}{2}\frac{\rho_0}{\sqrt{1-\sum_{j=0}^m \rho_j^2}} \frac{\mu_{yt}}{\sigma_{yt}} & \frac{\rho_0\rho_m}{1-\sum_{j=0}^m \rho_j^2} & \cdots & \frac{\rho_0^2}{1-\sum_{j=0}^m \rho_j^2} \end{bmatrix}.\end{aligned} \tag{60}$$

For a full derivation see Appendix B. With the quantities and adjustments mentioned above, we can straightforwardly implement the Bellman filter of Lange (2020) from Algorithm 3.[13]

4.3 Simulation results

We conduct an extensive simulation study to investigate how the different stochastic volatility models are estimated when data is simulated from different DGPs. In all simulations, we set $c = 0, \phi = 0.975, \sigma_n = 0.15$, but differ the values the leverage effect $\boldsymbol{\rho}$. Furthermore, for each different DGP, we simulate 20 time series of length $T = 5000$.[14] Estimation is done for the Catania model as in Section 4.1 and the new SV model as in Section 4.2. On all simulated series, we estimate six versions of the Catania model and six versions of the new stochastic volatility model. Firstly, we estimate the models without any leverage effect. Furthermore, we estimate the models when only allowing contemporaneous leverage ($\rho_0 \neq 0$) and when only allowing inter-temporal leverage ($\rho_1 \neq 0$). We also estimate the models when allowing for two leverage effects, once when having both contemporaneous leverage and inter-temporal leverage ($\rho_0 \neq 0, \rho_1 \neq 0$), and once when allowing two inter-temporal leverage effects, ($\rho_0 \neq 0, \rho_1 \neq 0$). Finally, we estimate both models when allowing for three leverage effects, both contemporaneous and two inter-temporal leverage effects ($\rho_0 \neq 0, \rho_1 \neq 0, \rho_2 \neq 0$). For the particle filter, we use $N = 5000$. Results indicated that this was necessary, as for lesser values the particle filter did not give optimal results. We report the average and errors of the estimated model parameters. Furthermore, similar to Section 3.4, we report the out-of-sample estimation errors (on the extend of the time series with a further $T = 1000$) of our quantity of interest, the volatility given by $\exp\left\{\frac{h_{t+1}}{2}\right\}$, after estimating the parameters.

We first simulate data from the standard stochastic volatility model of Taylor (1986), the contemporaneous leverage stochastic volatility model of Jacquier et al. (2004), and the inter-temporal stochastic volatility model of Harvey and Shephard (1996). For the leverage parameters ρ_0 and ρ_1, we use the true parameter value $\rho_j = -0.7$, a value similar to that found in empirical applications. Also, for both the Catania model and the new SV model, we consider models without the contemporaneous leverage effect, $\rho_0 = 0$, and models with the contemporaneous leverage effect.

Table 6 shows the results of this simulation. In Panel A we have simulated data according to the classic stochastic volatility model without leverage. We find the models to perform very similar. The out-of-sample volatility predictions are very similar for all models. Furthermore, the

[13] We use the same stopping criterion and maximum iterations as in Section 3.4.

[14] Due to the immense computational time required for the particle filter, which also increased for the stochastic volatility models with leverage, we simulate a limited amount of time series.

estimations of c, ϕ, and σ_n are pretty accurate, with the Catania model that uses particle filter a little better. The models estimate the leverage effects ρ quite close to zero on average. However, looking at the root mean squared errors (RMSEs) of the leverage parameters, we see that they are not equal to zero but around 0.1. For the models with one leverage effect, models with only $\rho_0 \neq 0$ or $\rho_1 \neq 0$, RMSEs of the leverage parameters are less than 0.1. This is also the case for the new SV models with two leverage effects. The Catania models with two leverage effects have RMSEs around 0.10 to 0.15, indicating that the values are on average estimated zero, but single estimates could be wrong by over 0.10. Finally, for the three leverage effects model, we find the values again lying close to zero, but with even larger RMSEs. Further research into the results shows that the models often have offsetting parameters, for example $\rho_0 = -0.3$ and $\rho_2 = 0.3$. This is especially the case for the new SV model with three leverage effects, as can be seen by the high RMSEs for the leverage parameters. Thus, while in this case there is no leverage effect, the more comprehensive models are not very good in fully finding the lack of leverage effects.

Furthermore, Panel B and Panel C show estimations when the contemporaneous leverage SV model is the DGP and when the inter-temporal leverage SV model is the DGP, respectively. We find that this time, the estimation of σ_n is somewhat less accurate, but still quite good. More interesting is the estimation of the leverage effects. When estimating a model with just one leverage effect, it will show a very large value despite the true timing of the leverage effect. For example, while the DGP in Panel B has a contemporaneous leverage effect of $\rho_0 = -0.7$, the models in the third row estimate an average inter-temporal leverage effect of $\rho_1 = -0.718$ and $\rho_1 = -0.740$ for the new SV model and Catania model respectively. This is an important result, as this explains why researchers as Harvey and Shephard (1996) find a large value for ρ_1 and researchers as Jacquier et al. (2004) find a large value for ρ_0. When there is only one leverage effect in the DGP, and the new SV model is estimated with two leverage effects, it is pretty good at identifying the leverage effect, as can be seen from the fourth rows of Panel B and Panel C. For example, in Panel B, the new SV model estimates an average ρ_0 of -0.802 and an average ρ_1 of -0.006, very close to the true values. The Catania model is performs worse for identifying the correct leverage effect, with average parameters of around -0.4 and -0.3 when estimated with only contemporaneous or only inter-temporal leverage. A further look into these estimations shows that the two leverage parameters are almost always very close, and we do not see any switching behaviour as we saw for the new SV model with three leverage effects. Finally, both models perform a bit worse than before when estimating three leverage effects. We find the new SV model often estimates a large ρ_2, while the true value is zero. The Catania model estimates a leverage effect for all three parameters, while the DGPs only contain one leverage effect. Finally,

Table 6: Simulation results where basic stochastic volatility models are the data-generating processes

New SV model							**Catania model**						
c	ϕ	σ_n	ρ_0	ρ_1	ρ_2	$\hat{\sigma}_t$	c	ϕ	σ_n	ρ_0	ρ_1	ρ_2	$\hat{\sigma}_t$
Panel A: DGP is the classic stochastic volatility model without leverage													
0.005	0.971	0.163	-	-	-	0.242	-0.001	0.975	0.149	-	-	-	0.239
0.006	0.006	0.023	-	-	-	-	0.002	0.005	0.019	-	-	-	-
0.005	0.971	0.163	0.009	-	-	0.242	-0.001	0.974	0.150	-0.011	-	-	0.239
0.006	0.006	0.023	0.078	-	-	-	0.003	0.004	0.016	0.105	-	-	-
0.005	0.972	0.163	-	0.009	-	0.242	-0.001	0.974	0.149	-	0.009	-	0.239
0.006	0.006	0.023	-	0.070	-	-	0.002	0.005	0.017	-	0.076	-	-
0.005	0.972	0.163	0.015	-0.006	-	0.242	-0.001	0.974	0.151	-0.002	0.010	-	0.239
0.006	0.006	0.023	0.086	0.068	-	-	0.002	0.005	0.017	0.125	0.104	-	-
0.005	0.972	0.163	-	0.035	-0.031	0.242	-0.001	0.974	0.152	-	0.059	-0.057	0.240
0.006	0.006	0.023	-	0.087	0.081	-	0.002	0.004	0.016	-	0.162	0.141	-
0.005	0.972	0.163	0.126	0.028	-0.101	0.242	-0.001	0.974	0.153	-0.002	0.052	-0.048	0.240
0.004	0.006	0.027	0.307	0.121	0.266	-	0.002	0.005	0.015	0.133	0.178	0.131	-
Panel B: DGP is the contemporaneous leverage stochastic volatility model with $\rho_0 = -0.7$													
0.004	0.971	0.147	-	-	-	0.197	-0.001	0.978	0.134	-	-	-	0.197
0.004	0.008	0.013	-	-	-	-	0.002	0.007	0.023	-	-	-	-
-0.001	0.976	0.128	-0.803	-	-	0.150	-0.005	0.967	0.154	-0.590	-	-	0.166
0.001	0.003	0.023	0.105	-	-	-	0.007	0.020	0.048	0.164	-	-	-
-0.004	0.974	0.124	-	-0.718	-	0.151	-0.005	0.975	0.120	-	-0.740	-	0.151
0.004	0.003	0.028	-	0.718	-	-	0.005	0.003	0.031	-	0.740	-	-
-0.001	0.976	0.128	-0.802	-0.006	-	0.150	-0.006	0.979	0.132	-0.463	-0.295	-	0.172
0.001	0.003	0.023	0.104	0.011	-	-	0.006	0.009	0.046	0.466	0.350	-	-
-0.004	0.973	0.126	-	-0.492	-0.226	0.154	-0.005	0.975	0.121	-	-0.657	-0.095	0.151
0.004	0.004	0.026	-	0.596	0.385	-	0.005	0.004	0.031	-	0.678	0.259	-
-0.002	0.978	0.145	-0.706	-0.089	-0.169	0.165	-0.005	0.964	0.123	-0.397	-0.256	-0.317	0.176
0.002	0.007	0.016	0.226	0.126	0.366	-	0.006	0.021	0.036	0.314	0.339	0.362	-
Panel C: DGP is the inter-temporal leverage stochastic volatility model with $\rho_1 = -0.7$													
0.003	0.972	0.142	-	-	-	0.205	-0.001	0.977	0.135	-	-	-	0.205
0.004	0.007	0.015	-	-	-	-	0.002	0.005	0.022	-	-	-	-
0.004	0.980	0.112	-0.691	-	-	0.160	0.001	0.977	0.123	-0.579	-	-	0.173
0.004	0.005	0.039	0.693	-	-	-	0.002	0.010	0.034	0.590	-	-	-
0.001	0.973	0.133	-	-0.782	-	0.145	0.000	0.976	0.126	-	-0.834	-	0.151
0.002	0.003	0.019	-	0.087	-	-	0.002	0.004	0.002	-	0.156	-	-
0.000	0.976	0.127	-0.006	-0.797	-	0.136	0.001	0.982	0.111	-0.374	-0.461	-	0.153
0.001	0.003	0.024	0.011	0.099	-	-	0.003	0.011	0.047	0.404	0.293	-	-
0.000	0.976	0.128	-	-0.666	-0.131	0.138	-0.001	0.976	0.124	-	-0.740	-0.094	0.136
0.001	0.003	0.024	-	0.302	0.314	-	0.002	0.003	0.030	-	0.142	0.252	-
0.002	0.978	0.164	-0.215	-0.200	-0.799	0.160	0.003	0.976	0.116	-0.297	-0.502	-0.329	0.180
0.002	0.003	0.018	0.215	0.500	0.799	-	0.008	0.009	0.040	0.347	0.247	0.344	-

NOTE: We simulate 20 stochastic volatility time series of length 5,000. In all simulations we use parameter values $c = 0$, $\phi = 0.975$, and $\sigma_n = 0.15$. In Panel A we simulate the classic stochastic volatility model without leverage. In Panel B, we simulate the stochastic volatility model with contemporaneous leverage using value $\rho_0 = -0.7$. In Panel C, we simulate the stochastic volatility model with inter-temporal leverage using $\rho_1 = -0.7$. We then estimate the volatility models of Section 4.1 and Section 4.2 while varying the amount of leverage effects within these models between zero and three. This table shows the average of the estimated parameters with the root mean squared error (RMSE) below. Furthermore, we compute one-step-ahead volatility predictions $\hat{\sigma}_t$ on an extend of the time series of length 1,000 and report the RMSE.

we also find the volatility estimates to worsen when adding the third leverage effect, while two leverage effects give the best volatility predictions.

Next, we simulate data using the Catania model as the DGP, while using three different sets of leverage effects. First, we simulate time series from the Catania model with $\rho_0 = -0.7$ and $\rho_1 = -0.4$. Then, we simulate with values $\rho_1 = -0.7$ and $\rho_2 = -0.4$ and finally with values $\rho_0 = -0.7$, $\rho_1 = -0.4$, and $\rho_2 = -0.2$. Results are presented in Table 7. We note that the new SV model has very high estimations for a single leverage effect in every model type. When estimated with only ρ_0 or with only ρ_1, an average parameter of around -0.9 is estimated. Also when estimated with two leverage effects, the new SV model assigns all weight to a single ρ. This is probably a result of the lack of autocorrelation in the simulated data. Whereas the new SV model with multiple leverages assumes correlation between the data. When estimated with three leverage effects, the new SV model estimates on average a high ρ_2 with lower values of ρ_0 and ρ_1, even when the DGP contains only leverage effects for ρ_0 and ρ_1. The Catania model predicts the leverage effects somewhat less extreme. When only estimated with contemporaneous leverage, the estimation for ρ_0 is on the lower side of the true value, whereas the new SV model estimated ρ_0 larger than its true value. When the Catania model corresponding to its DGP is estimated, results are on average pretty good, but have a high RMSE, especially in the model with three leverage effects. These results are in line with those of Catania (2020). However, when the Catania model is estimated on a different DGP, the results are not always as accurate. For example, in Panel B we have $\rho_1 = -0.7$ and $\rho_2 = -0.4$. When the Catania model is estimated with only contemporaneous and one inter-temporal leverage effect, is estimates average values of $\rho_0 = -0.426$ and $\rho_1 = -0.483$. In this case, the model cannot identify the true leverage effect. Also when estimating the Catania model with three leverage effects, it is not able to identify the two leverage effects of the DGP. Finally, when the Catania model with three leverage effects is estimated on its corresponding DGP, it is again not able to correctly identify the right sizes of the true leverage effects, but gives average estimations of around -0.4 for all three parameters. Again, these results are quite in line with Catania (2020), who shows that when using $T = 50,000$ or $T = 100,000$ the leverage effects are identified better. However, time series of such length are empirically not available. Even if such a time series would be available, it is infeasible to assume the underlying model would remain constant of such a long time period.

Lastly, we simulate data using the new SV model as the DGP, with true parameter values for ρ as in the simulations of the Catania model. Results are presented in Table 8. We find that the new SV model performs quite well when data is simulated from its own DGP. When only one leverage effect is estimated, average parameters are again very high. For models with two leverage

Table 7: Simulation results where the Catania model is the data-generating process

New SV model							**Catania model**						
c	ϕ	σ_n	ρ_0	ρ_1	ρ_2	$\hat{\sigma}_t$	c	ϕ	σ_n	ρ_0	ρ_1	ρ_2	$\hat{\sigma}_t$
Panel A: DGP is the Catania model with $\rho_0 = -0.7$ *and* $\rho_1 = -0.4$													
0.008	0.972	0.188	-	-	-	0.251	0.000	0.975	0.171	-	-	-	0.249
0.008	0.007	0.041	-	-	-	-	0.002	0.006	0.025	-	-	-	-
0.002	0.979	0.156	-0.942	-	-	0.129	-0.001	0.968	0.166	-0.569	-	-	0.191
0.002	0.004	0.008	0.242	-	-	-	0.006	0.010	0.023	0.168	-	-	-
-0.008	0.975	0.162	-	-0.937	-	0.126	-0.008	0.975	0.162	-	-0.942	-	0.126
0.008	0.002	0.013	-	0.537	-	-	0.008	0.002	0.013	-	0.543	-	-
0.002	0.979	0.157	-0.942	-0.007	-	0.129	-0.004	0.971	0.130	-0.549	-0.387	-	0.158
0.002	0.004	0.008	0.242	0.394	-	-	0.005	0.006	0.026	0.163	0.083	-	-
-0.008	0.975	0.162	-	-0.937	-0.001	0.126	-0.008	0.975	0.158	-	-0.912	-0.076	0.127
0.008	0.002	0.013	-	0.537	0.002	-	0.008	0.002	0.021	-	0.520	0.080	-
-0.002	0.973	0.217	-0.218	-0.207	-0.794	0.171	-0.004	0.973	0.146	-0.445	-0.292	-0.346	0.196
0.002	0.004	0.068	0.482	0.207	0.794	-	0.006	0.012	0.041	0.306	0.257	0.344	-
Panel B: DGP is the Catania model with $\rho_1 = -0.7$ *and* $\rho_2 = -0.4$													
0.008	0.971	0.188	-	-	-	0.265	-0.001	0.974	0.172	-	-	-	0.263
0.008	0.008	0.041	-	-	-	-	0.002	0.006	0.026	-	-	-	-
0.007	0.983	0.135	-0.889	-	-	0.141	0.003	0.983	0.132	-0.579	-	-	0.209
0.007	0.008	0.015	0.889	-	-	-	0.006	0.010	0.030	0.591	-	-	-
-0.001	0.977	0.164	-	-0.950	-	0.109	-0.001	0.977	0.159	-	-0.964	-	0.113
0.001	0.002	0.015	-	0.250	-	-	0.001	0.002	0.018	-	0.263	-	-
-0.001	0.977	0.164	0.001	-0.950	-	0.109	0.001	0.975	0.115	-0.426	-0.483	-	0.171
0.001	0.002	0.015	0.009	0.250	-	-	0.002	0.009	0.049	0.430	0.284	-	-
-0.001	0.976	0.167	-	-0.480	-0.474	0.116	-0.001	0.976	0.131	-	-0.720	-0.516	0.105
0.001	0.002	0.018	-	0.372	0.370	-	0.001	0.002	0.020	-	0.042	0.155	-
0.003	0.974	0.218	-0.190	-0.175	-0.822	0.136	0.005	0.978	0.138	-0.389	-0.450	-0.332	0.170
0.004	0.002	0.070	0.194	0.478	0.423	-	0.006	0.010	0.055	0.408	0.278	0.111	-
Panel C: DGP is the Catania model with $\rho_0 = -0.7$ *and* $\rho_1 = -0.4$ *and* $\rho_2 = -0.2$													
0.011	0.972	0.215	-	-	-	0.286	-0.001	0.975	0.197	-	-	-	0.284
0.011	0.007	0.067	-	-	-	-	0.002	0.005	0.049	-	-	-	-
0.005	0.980	0.176	-0.961	-	-	0.130	0.003	0.970	0.189	-0.598	-	-	0.204
0.005	0.005	0.026	0.260	-	-	-	0.005	0.007	0.047	0.112	-	-	-
-0.009	0.976	0.187	-	-0.966	-	0.124	-0.010	0.976	0.186	-	-0.969	-	0.124
0.009	0.001	0.037	-	0.566	-	-	0.010	0.001	0.037	-	0.269	-	-
0.005	0.980	0.177	-0.961	-0.008	-	0.130	-0.004	0.974	0.130	-0.562	-0.501	-	0.175
0.005	0.005	0.028	0.261	0.561	-	-	0.006	0.003	0.030	0.178	0.191	-	-
-0.009	0.975	0.189	-	-0.648	-0.318	0.136	-0.010	0.975	0.161	-	-0.901	-0.251	0.123
0.009	0.002	0.039	-	0.510	0.463	-	0.010	0.002	0.018	-	0.508	0.200	-
-0.002	0.973	0.252	-0.221	-0.208	-0.786	0.179	-0.006	0.973	0.124	-0.475	-0.371	-0.451	0.163
0.002	0.004	0.104	0.480	0.193	0.586	-	0.007	0.003	0.036	0.237	0.142	0.269	-

NOTE: We simulate 20 stochastic volatility time series of length 5,000. In all simulations we use parameter values $c = 0$, $\phi = 0.975$, and $\sigma_n = 0.15$. In Panel A we simulate data from the Catania model (Equation 18) using parameters $\rho_0 = -0.7$ and $\rho_1 = -0.4$. In Panel Bw e simulate data from the Catania model using values $\rho_1 = -0.7$ and $\rho_2 = -0.4$. In Panel C we simulate data from the Catania model using values $\rho_0 = -0.7$, $\rho_1 = -0.4$, and $\rho_2 = -0.2$. We then estimate the volatility models of Section 4.1 and Section 4.2 while varying the amount of leverage effects within these models between zero and three. This table shows the average of the estimated parameters with the root mean squared error (RMSE) below. Furthermore, we compute one-step-ahead volatility predictions $\hat{\sigma}_t$ on an extend of the time series of length 1,000 and report the RMSE.

Table 8: Simulation results where the new SV model is the data-generating process

New SV model							**Catania model**						
c	ϕ	σ_n	ρ_0	ρ_1	ρ_2	$\hat{\sigma}_t$	c	ϕ	σ_n	ρ_0	ρ_1	ρ_2	$\hat{\sigma}_t$
Panel A: DGP is the new SV model with $\rho_0 = -0.7$ *and* $\rho_1 = -0.4$													
-0.001	0.961	0.206	-	-	-	0.286	-0.009	0.965	0.186	-	-	-	0.267
0.005	0.017	0.060	-	-	-	-	0.010	0.012	0.040	-	-	-	-
-0.005	0.974	0.127	-0.875	-	-	0.191	-0.010	0.971	0.159	-0.625	-	-	0.222
0.005	0.002	0.025	0.179	-	-	-	0.011	0.015	0.034	0.187	-	-	-
-0.011	0.970	0.135	-	-0.876	-	0.182	-0.012	0.971	0.135	-	-0.897	-	0.177
0.012	0.007	0.019	-	0.481	-	-	0.012	0.006	0.018	-	0.501	-	-
-0.007	0.974	0.155	-0.790	-0.460	-	0.181	-0.012	0.968	0.158	-0.453	-0.430	-	0.205
0.007	0.002	0.011	0.100	0.067	-	-	0.013	0.010	0.025	0.257	0.072	-	-
-0.015	0.961	0.157	-	-0.442	-0.572	0.221	-0.010	0.973	0.221	-	-0.885	0.372	0.179
0.019	0.040	0.076	-	0.044	0.573	-	0.010	0.003	0.078	-	0.486	0.376	-
-0.011	0.972	0.165	-0.314	-0.552	-0.413	0.181	-0.014	0.964	0.142	-0.352	-0.277	-0.309	0.206
0.012	0.007	0.023	0.362	0.176	0.349	-	0.015	0.018	0.044	0.358	0.161	0.349	-
Panel B: DGP is the new SV model with $\rho_1 = -0.7$ *and* $\rho_2 = -0.4$													
-0.002	0.957	0.213	-	-	-	0.293	-0.010	0.963	0.191	-	-	-	0.276
0.004	0.021	0.069	-	-	-	-	0.011	0.015	0.047	-	-	-	-
0.001	0.981	0.112	-0.797	-	-	0.189	-0.004	0.981	0.125	-0.562	-	-	0.221
0.001	0.006	0.039	0.798	-	-	-	0.006	0.010	0.037	0.571	-	-	-
0.005	0.003	0.020	-	-0.877	-	0.173	-0.005	0.975	0.130	-	-0.891	-	0.168
0.012	0.007	0.019	-	0.170	-	-	0.005	0.003	0.024	-	0.195	-	-
-0.003	0.977	0.145	-0.429	-0.834	-	0.172	-0.004	0.981	0.117	-0.246	-0.524	-	0.197
0.003	0.003	0.011	0.430	0.137	-	-	0.005	0.010	0.037	0.306	0.282	-	-
-0.008	0.972	0.145	-	-0.421	-0.584	0.208	-0.006	0.975	0.108	-	-0.692	-0.372	0.169
0.008	0.004	0.010	-	0.279	0.185	-	0.006	0.006	0.043	-	0.149	0.166	-
-0.005	0.974	0.172	-0.274	-0.432	-0.618	0.169	-0.007	0.972	0.119	-0.015	-0.545	-0.450	0.174
0.006	0.005	0.025	0.278	0.280	0.274	-	0.007	0.008	0.035	0.200	0.223	0.130	-
Panel C: DGP is the new SV model with $\rho_0 = -0.7$ *and* $\rho_1 = -0.4$ *and* $\rho_2 = -0.2$													
-0.031	0.949	0.252	-	-	-	0.361	-0.039	0.955	0.226	-	-	-	0.337
0.033	0.029	0.105	-	-	-	-	0.040	0.023	0.078	-	-	-	-
-0.017	0.975	0.122	-0.747	-	-	0.298	-0.013	0.983	0.134	-0.598	-	-	0.285
0.017	0.002	0.030	0.065	-	-	-	0.017	0.015	0.030	0.130	-	-	-
-0.026	0.968	0.136	-	-0.762	-	0.290	-0.027	0.969	0.132	-	-0.780	-	0.283
0.026	0.007	0.019	-	0.366	-	-	0.027	0.006	0.021	-	0.383	-	-
-0.020	0.977	0.121	-0.647	-0.598	-	0.315	-0.025	0.971	0.115	-0.364	-0.415	-	0.295
0.020	0.006	0.034	0.063	0.202	-	-	0.028	0.015	0.039	0.342	0.084	-	-
-0.029	0.970	0.123	-	-0.476	-0.556	0.334	-0.025	0.971	0.157	-	-0.836	0.183	0.283
0.029	0.005	0.028	-	0.076	0.356	-	0.025	0.005	0.022	-	0.437	0.394	-
-0.027	0.971	0.155	-0.421	-0.524	-0.512	0.313	-0.023	0.972	0.099	-0.340	-0.332	-0.203	0.294
0.028	0.004	0.009	0.282	0.143	0.358	-	0.024	0.009	0.058	0.397	0.129	0.126	-

NOTE: We simulate 20 stochastic volatility time series of length 5,000. In all simulations we use parameter values $c = 0$, $\phi = 0.975$, and $\sigma_n = 0.15$. In Panel A we simulate data from the new SV model (Equation 20) using parameters $\rho_0 = -0.7$ and $\rho_1 = -0.4$. In Panel B we simulate data from the new SV model using values $\rho_1 = -0.7$ and $\rho_2 = -0.4$. In Panel C simulate data from the new SV model using values $\rho_0 = -0.7$, $\rho_1 = -0.4$, and $\rho_2 = -0.2$. We then estimate the volatility models of Section 4.1 and Section 4.2 while varying the amount of leverage effects within these models between zero and three. This table shows the average of the estimated parameters with the root mean squared error (RMSE) below. Furthermore, we compute one-step-ahead volatility predictions $\hat{\sigma}_t$ on an extend of the time series of length 1,000 and report the RMSE.

effects, both are estimated quite accurately. When estimated with three leverage effects, there is again a too high average estimation for the last parameter ρ_2. The Catania model estimates similar parameters for ρ_0 and ρ_1 regardless of the DGP, while when estimated with three leverage effects again estimates all three leverage effects around -0.4. As the stochastic volatility process does not depend on past/current returns and is a fully independent process, we find that the volatility is harder to predict, with higher RMSEs for the volatility predictions $\hat{\sigma}_t$.

Noticeable is that, in all simulation studies above, adding contemporaneous leverage $\rho_0 \neq 0$ did not improve the volatility predictions for the Catania model. Furthermore, the three leverage models often performed worse when estimating the volatility for both the new stochastic volatility model as the Catania model. For the Catania model, we know from Equation (52) that the conditional distribution of the volatility $h_t|\mathcal{I}_{t-1}$ does not depend on ρ_0, which explains why the volatility predictions did not improve.

5 Empirical application

In the previous section, we have seen that the models can estimate the parameters somewhat accurately, but that it very much depends on the underlying DGP and the choice of leverage effects in the model. In this section, we apply the models to empirical data.

5.1 Data

We use daily logarithmic returns for two major financial indices: the Standard & Poor's 500 (SP500) and the Financial Times Stock Exchange 100 Index (FTSE). These returns are obtained from Yahoo Finance. [15] The data is obtained over the period 1 January 1990 to 31 December 2019, resulting in 7558 observations for the SP500 and 7606 observations for the FTSE.

Furthermore, we use foreign exchange (FX) data for two exchange rates: The Japanese yen (Yen) to the United States dollar (USD) and the United States dollar to the United Kingdom pound sterling (Pound). The exchange rates are obtained from the Federal Reserve Bank Reports via the Wharton Research Data Services (WRDS). We construct daily logarithmic returns over the period 1 January 1990 to 31 December 2018, resulting in 7285 observations for both the (YEN/USD) series and the (Pound/USD) series.

Finally, we use daily logarithmic returns on two stocks that have been part of the SP500 for longer than the sample period, namely Apple Inc.(AAPL) and The Boeing Company (BA). These companies are active in different industries: Technology and Industrials (Aerospace & Defense)

[15] Available from https://finance.yahoo.com.

respectively. Returns are obtained from Yahoo Finance over the period 1 January 1990 to 31 December 2019. Resulting in 7557 returns for both the AAPL series and the BA series.

5.2 Results

We have estimated the same models as in the simulation studies in Section 4.3 on the two major indices. Results are presented in Table 9. About the standard parameters c, ϕ, and σ_n the models are generally in agreement. With c fluctuating between -0.19 and -0.30 for the SP500 series and between -0.11 and -0.16 for the FTSE series. The log-volatility transition parameter ϕ is close to one, ranging between 0.968 and 0.986 for all models. Finally, the volatility of the log-volatility state equation, σ_n, is somewhat higher for the SP500 than for the FTSE, with values around 0.20 and 0.15 respectively.

Looking at the leverage parameters for the SP500, both the new SV model and the Catania model estimate both the contemporaneous leverage parameter and the inter-temporal leverage parameter at -0.7 when it is the only parameter in the model. As we have seen in the simulations of Section 4.3, no conclusions can be derived from this result, as these parameters are estimated high regardless of the correct timing of the leverage effect. Looking at the estimations of the new SV model with two leverage effects, in rows four and five, we find that models the parameter ρ_1 is the only estimated leverage effect, with values of -0.686 and -0.702. In the simulation studies we have seen that the new SV model estimates only one leverage effect different from zero, if the data is generated from a different GDP. Furthermore, in Panel B and C of Table 6 we have seen that the new SV model correctly estimates the leverage effect at the right time, when there is only one leverage effect in the DGP but two are estimated. Finally, the new SV model with three leverage effects estimates parameters $\rho_0 = -0.591, \rho_1 = -0.177$, and $\rho_2 = 0.362$. This is an interesting result, as the model gave similar results in Panel A and B of Table 6, where no leverage effect or a contemporaneous leverage effect was present. When data was simulated with inter-temporal leverage, as in Panel C of Table 6, and when data was simulated from the Catania model as DGP, as in Table 7, the three leverage new SV model always estimated ρ_2 very large and negative.

The two aforementioned results are somewhat contrary. Comparing the results of row four and five of the SP500 estimations with the simulations, one would say the correct timing of the leverage is at ρ_1. However, comparing the result of row six with the simulations, where the three leverage new SV model is estimated, would suggest that neither the inter-temporal leverage model nor the Catania models are the DGP. When the Catania model is estimated with ρ_0 and ρ_1 we obtain values -0.781 and -0.382 respectively, suggesting a leverage effect at both times, but more

significant contemporaneously. When the model is subsequently estimated with three leverage effects, we obtain values $\rho_0 = -0.113$ and $\rho_1 = -0.747$, which contradict the results from the two leverage model. Furthermore, when the Catania model is estimated with ρ_2, we obtain estimates around 0.2, suggesting a small reverted leverage effect. However, as we seen in the simulations that the estimated parameters can still be somewhat off for such a sample size, we cannot draw any large conclusion from that parameter.

Results for the FTSE index are similar, except this time there is even a more pronounced inter-temporal leverage effect. Again, both the new SV models and the Catania models with a single leverage effect have values around -0.7, regardless of the timing. Furthermore, the new SV model specification with two leverage effects assigns all weight on the first inter-temporal parameter ρ_1. The Catania models with two leverage effects assign a much larger weight to ρ_1 as well, in contrast to the SP500 series. Finally, the new SV model with three leverage effects shows again very unclear results, with a positive parameter ρ_1 and a strongly negative ρ_0 and slightly negative ρ_2. In the simulations we have seen that while this model does not identify the leverage effects correctly, it does assign negative values to all parameters when data is simulated from the classic stochastic volatility with leverage or the Catania model. As this does not happen here, it seems like the underlying process is not defined as clearly as the Catania model or the classic leverage models assume.

5.3 Forecasting performance

The above results suggest that, if we would construct a stochastic volatility model with only one leverage specification, it would be best to use the inter-temporal leverage parameter ρ_1, as that is the parameter which the new SV model estimated to be nonzero. While this is an interesting result, we are also interested in whether this specification then also leads to the best volatility predictions. To investigate this, we perform an out-of-sample analysis. We split the index returns series in an out-of-sample period of the last 1500 returns, and estimate the model parameters on the preceding in-sample sample period. With both models, we make predictions of the variance (squared volatility) $var[y_t|\mathcal{I}_{t-1}] = \hat{\sigma_t}^2$. For the Catania model, we make predictions of the volatility using the predicted mean of the volatility $\hat{\sigma_t}^2 = E[\sigma_t^2|\mathcal{I}_{t-1}] = E[\exp\{h_t\}|\mathcal{I}_{t-1}]$, which can be computed with the predicted state particles, as in Section 3.4. For the new SV model, we compute the Bellman predicted state of $h_{t|t-1}$ and transform it into predicted volatility by $\hat{\sigma}_t^2 = \exp\{h_{t|t-1}\}$. Following Patton (2011), we use the mean squared error (MSE) loss function to compare different variance predictions. As Patton (2011) noted, MSE induces a perfect ranking

Table 9: Estimated parameters from the Catania model and the new SV model on stock index returns

New SV model						**Catania model**					
c	ϕ	σ_n	ρ_0	ρ_1	ρ_2	c	ϕ	σ_n	ρ_0	ρ_1	ρ_2
Panel A: Standard & Poor's 500 (SP500)											
-0.194	0.979	0.203	-	-	-	-0.199	0.982	0.179	-	-	-
0.026	0.003	0.011	-	-	-	0.018	0.002	0.008	-	-	-
-0.191	0.981	0.190	-0.722	-	-	-0.196	0.981	0.193	-0.727	-	-
0.022	0.002	0.010	0.027	-	-	0.020	0.007	0.007	0.023	-	-
-0.295	0.968	0.243	-	-0.692	-	-0.270	0.971	0.228	-	-0.722	-
0.033	0.004	0.014	-	0.026	-	0.023	0.003	0.017	-	0.019	-
-0.297	0.968	0.245	-0.014	-0.686	-	-0.244	0.982	0.220	-0.781	-0.382	-
0.033	0.004	0.014	0.018	0.028	-	0.011	0.011	0.007	0.008	0.011	-
-0.294	0.968	0.243	-	-0.702	0.020	-0.249	0.973	0.276	-	-0.802	0.241
0.033	0.003	0.014	-	0.026	0.012	0.027	0.003	0.020	-	0.019	0.050
-0.286	0.968	0.226	-0.591	-0.177	0.362	-0.225	0.976	0.219	-0.113	-0.747	0.190
0.050	0.005	0.018	0.014	0.010	0.013	0.010	0.006	0.006	0.104	0.088	0.149
Panel B: Financial Times Stock Exchange 100 Index (FTSE)											
-0.130	0.986	0.145	-	-	-	-0.119	0.987	0.132	-	-	-
0.027	0.003	0.012	-	-	-	0.014	0.002	0.004	-	-	-
-0.111	0.987	0.138	-0.703	-	-	-0.132	0.985	0.140	-0.688	-	-
0.018	0.002	0.009	0.039	-	-	0.021	0.003	0.010	0.040	-	-
-0.151	0.983	0.150	-	-0.695	-	-0.143	0.985	0.145	-	-0.715	-
0.023	0.002	0.010	-	0.037	-	0.026	0.003	0.009	-	0.038	-
-0.152	0.983	0.151	-0.005	-0.692	-	-0.152	0.969	0.177	-0.257	-0.512	-
0.025	0.003	0.011	0.020	0.040	-	0.014	0.003	0.013	0.039	0.009	-
-0.151	0.983	0.151	-	-0.699	0.008	-0.143	0.985	0.153	-	-0.745	0.079
0.022	0.002	0.010	-	0.037	0.014	0.053	0.006	0.032	-	0.058	0.109
-0.149	0.984	0.162	-0.608	0.233	-0.137	-0.159	0.983	0.147	-0.231	-0.428	0.199
0.399	0.001	0.615	0.097	0.101	0.098	0.025	0.003	0.003	0.010	0.011	0.008

NOTE: This table shows parameter estimations of the stochastic volatility models from Section 4.1 and Section 4.2 on returns of two major indices. We vary the leverage effects in the stochastic volatility models between zero and three. The estimated parameters are shown in this table, with below their estimated standard errors using MATLAB's fminunc function.

of competing volatility forecasts, and is given by

$$\text{MSE}_t = (\tilde{\sigma_t}^2 - \hat{\sigma_t}^2)^2, \tag{61}$$

where $\hat{\sigma_t}^2$ is the predicted variance and $\tilde{\sigma_t}^2$ is a proxy for the variance. As proxy for variance, we use squared returns y_t^2. While squared returns are generally noisy, they are valid volatility proxies (Patton, 2011).

Table 10 reports the average MSE computed over the out of sample period of length $T = 1500$. We make forecasts for the same 12 models that we used before, now numbered model (1) to (6), with the same ordering as before. Model (1) indicates the no leverage specification, and model (6) is the model specification with three leverage effects. We set the no leverage specification for

Table 10: Relative mean square error (MSE) average losses for volatility predictions of the stochastic volatility models on index returns

	New SV model						**Catania model**					
Index	(1)	(2)	(3)	(4)	(5)	(6)	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Standard & Poor's 500 (SP500)												
MSE	1.000	0.944	0.933	0.944	0.932	0.930	1.004	0.922	0.928	0.928	0.954	0.921
Panel B: Financial Times Stock Exchange 100 Index (FTSE)												
MSE	1.000	0.960	0.961	0.943	0.961	0.964	1.000	0.993	0.956	0.931	0.961	0.960

NOTE: This table shows mean square error (MSE) average losses for volatility predictions on returns of two major indices. We use the MSE loss function from Patton (2011). As volatility proxy we use squared returns. We make out-of-sample volatility predictions for the returns on the last 1,500 observations, after estimating the stochastic volatility models from Section 4.1 and Section 4.2 on the previous observations. We vary the leverage effects in the stochastic volatility models between zero and three. Models (1) to (6) follow the same order as in the simulation studies of Section 4.3, but are now presented as columns. We use the first model, the new SV model (1), as benchmark and set its relative MSE to 1.

the new SV model as the benchmark with relative MSE of 1. As expected, the corresponding Catania model with no leverage performs almost equally, given that the only differences between the models is the estimation method, namely the Bellman filter and particle filter respectively.

For the SP500, we find the Catania model with three leverage effects to give the best results, almost 10% better than the benchmark model. Looking only at the new SV model results, the three leverage specification gives the best forecasts as well. Furthermore, the other models including contemporaneous leverage effects, (2) and (4), give worse results than the models with only inter-temporal leverage effects. For the Catania model we find the contrary, with models including contemporaneous leverage giving better forecasts than models without contemporaneous leverage.

Results for the FTSE index returns are different. Model (4) with contemporaneous and inter-temporal leverage gives best predictions for both the new SV model and the Catania model. Furthermore, predictions are now closer to the benchmark, with the best model giving less than 6% improvement relative to the stochastic volatility model without leverage.

While there is still not a clear best specification, including a leverage effect does always lead to an improvement in volatility predictions from the benchmark model, with prediction errors decreasing with at least 3% and at most almost 8%. In every case, the best forecasts are from a model that includes contemporaneous leverage. Thus, including contemporaneous leverage in a stochastic volatility certainly is useful from a forecasting point of view. As this was clearly not the case in the simulation studies of Section 4.3, it seems that there is a different contemporaneous relation in the financial returns than in the data-generating processes of the simulation studies.

Table 11: Estimated parameters from the Catania model and the new SV model on stock index returns

	New SV model						**Catania model**					
Index	(1)	(2)	(3)	(4)	(5)	(6)	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Standard & Poor's 500 (SP500)												
MSE	1.000	0.946	0.934	0.946	0.934	0.932	1.004	0.923	0.931	0.942	0.951	0.929
Panel B: Financial Times Stock Exchange 100 Index (FTSE)												
MSE	1.000	0.964	0.964	0.951	0.964	0.975	1.004	0.969	0.961	0.943	0.964	0.965

NOTE: This table shows mean square error (MSE) average losses for volatility predictions on returns of two major indices. We use the MSE loss function from Patton (2011). As volatility proxy we use the realized volatility from the Oxford-Man Realized Library. We make out-of-sample volatility predictions for the returns on the last 1,500 observations, after estimating the stochastic volatility models from Section 4.1 and Section 4.2 on the previous observations. We vary the leverage effects in the stochastic volatility models between zero and three. Models (1) to (6) follow the same order as in the simulation studies of Section 4.3, but are now presented as columns. We use the first model, the new SV model (1), as benchmark and set its relative MSE to 1.

Which suggests that none of the specifications used in this paper match the exact correlation structure of returns and volatility that is present in financial returns.

5.4 Robustness check using realized volatility

While the results of the above forecasting analysis is certainly useful, it might be affected by the choice of volatility proxy. Like we mentioned before, squared returns are generally very noisy. Therefore, we also compute analogous results when using the 5 minute realized volatility (RV) as volatility proxy. 5 minute realized volatility is obtained from the Oxford-Man Institute's realized library. [16]

Results are presented in Table 11. As expected, we find very similar results as before. Including leverage effects improves predictions from the no-leverage benchmark with the same magnitude of 4% to 7%. For the SP500, we find the Catania model with only contemporaneous leverage to give best results. Second best are the models with three leverage effects. Again, the Catania models where contemporaneous leverage is included provide better predictions, while the new SV models with contemporaneous leverage perform worse than the models without contemporaneous leverage. For the FTSE we find exactly the same results as before, with the fourth model, including contemporaneous and one inter-temporal leverage effect, providing best predictions for both the new SV model as the Catania model.

[16] Available from https://realized.oxford-man.ox.ac.uk

5.5 Robustness check using foreign exchange data

As we discussed in Section 2.2, empirical evidence finds that the leverage effect is present even when a firm does not have any debt, and that changes in debt do not increase volatility. Still, the leverage effect is very much related to the 'financial leverage', namely the debt/equity ratio of a firm. If the leverage effect does indeed come from financial leverage, despite the evidence mentioned earlier, it would be expected that foreign exchanges (FX) do not show a leverage effect, as foreign exchanges have no deb/equity structure. Meyer and Yu (2000) estimate the inter-temporal leverage model from Harvey and Shephard (1996) on the Pound/US dollar exchange rate series and do find a significant negative parameter ρ_1, but conclude that the effect is not as strong as found in index and stock returns. Furthermore, Catania (2020) finds a small 'inverted' leverage effect for the Canadian dollar (CAD) to US dollar (USD) series, but notes that the uncertainty is very large.

We estimate all 12 stochastic volatility models on the exchange rate series USD/Pound and Yen/USD. Complete estimation results are presented in Appendix C. Interestingly, we mostly find positive leverage effects for the USD/Pound series and negative leverage effects for the Yen/USD series. Both series include the US dollar, but the first as numerator and the second as denominator. It seems that the price of the US dollar influences the underlying volatility. If the price of the US dollar goes down, series USD/Pound has a positive return while Yen/USD has a negative return, which both lead to an increase in volatility. Furthermore, the leverage effect seems to take place contemporaneously, as the models with both contemporaneous leverage ρ_0 and inter-temporal leverage ρ_1 assign the largest (absolute) value to ρ_0.

We find the new SV model with three leverage effects to give almost off-setting ρ_0 and ρ_2, which we have seen from the simulation studies happen when there is no leverage effect in the DGP. The Catania model with three leverage effects assigns small values to all three parameters. However, from the simulation studies in Section 4.3, as well as the simulations in Catania (2020), we have seen that the model is cannot fully identify the leverage effects for time series of this length. However, the fact that both positive and negative values are found indicates that there is not a strong negative relation, as the simulation studies have shown that all parameters would be estimated negatively when this is the case. In fact, it seems that there is a small leverage effect when the price of the US dollar goes down, and that this effect somewhat corrected after two days.

Table 12: Estimated parameters from the Catania model and the new SV model on stock returns

New SV model						Catania model					
c	ϕ	σ_n	ρ_0	ρ_1	ρ_2	c	ϕ	σ_n	ρ_0	ρ_1	ρ_2
Panel A: Apple Inc. (AAPL)											
-0.488	0.933	0.378	-	-	-	-0.386	0.950	0.321	-	-	-
0.069	0.009	0.027	-	-	-	0.037	0.005	0.012	-	-	-
-0.461	0.937	0.364	-0.120	-	-	-0.342	0.955	0.292	-0.148	-	-
0.067	0.009	0.027	0.033	-	-	0.024	0.001	0.002	0.010	-	-
-0.491	0.933	0.378	-	-0.267	-	-0.400	0.948	0.324	-	-0.290	-
0.070	0.009	0.026	-	0.032	-	0.028	0.004	0.009	-	0.028	-
-0.490	0.933	0.378	-0.012	-0.263	-	-0.412	0.946	0.331	0.015	-0.293	-
0.061	0.008	0.024	0.029	0.031	-	0.001	0.002	0.004	0.017	0.023	-
-0.492	0.933	0.378	-	-0.255	-0.016	-0.387	0.949	0.320	-	-0.298	0.005
0.061	0.008	0.024	-	0.044	0.039	0.017	0.002	0.008	-	0.055	0.057
-0.439	0.952	0.311	-0.502	-0.225	0.411	-0.395	0.948	0.323	0.022	-0.307	0.003
0.045	0.006	0.023	0.014	0.014	0.012	0.004	0.001	0.007	0.021	0.005	0.010
Panel B: The Boeing Company (BA)											
-0.568	0.930	0.327	-	-	-	-0.424	0.949	0.262	-	-	-
0.095	0.011	0.031	-	-	-	0.071	0.008	0.024		-	-
-0.512	0.936	0.305	-0.127	-	-	-0.415	-0.950	0.260	-0.140	-	-
0.080	0.010	0.025	0.037	-	-	0.018	0.003	0.005	0.012	-	-
-0.535	0.934	0.314	-	-0.257	-	-0.410	0.950	0.257	-	-0.288	-
0.080	0.010	0.026	-	0.033	-	0.017	0.002	0.006	-	0.015	-
-0.532	0.934	0.313	-0.024	-0.247	-	-0.421	0.949	0.259	-0.004	-0.289	-
0.077	0.009	0.024	0.032	0.040	-	0.016	0.001	0.020	0.035	0.031	-
-0.568	0.930	0.324	-	-0.114	-0.161	-0.417	0.950	0.254	-	-0.156	-0.165
0.087	0.010	0.026	-	0.070	0.074	0.015	0.003	0.005	-	0.085	0.087
-0.530	0.961	0.237	-0.532	-0.196	0.424	-0.425	0.949	0.256	0.001	-0.159	-0.163
0.073	0.004	0.014	0.011	0.008	0.011	0.007	0.001	0.007	0.005	0.059	0.108

NOTE: This table shows parameter estimations of the stochastic volatility models from Section 4.1 and Section 4.2 on returns of two stocks part of the Standard & Poor's 500 (SP500). We vary the leverage effects in the stochastic volatility models between zero and three. The estimated parameters are shown in this table, with below their estimated standard errors using MATLAB's fminunc function.

5.6 Application to stock returns

In this section, we estimate the stochastic volatility models on returns of two major stocks part of the SP500: Apple Inc. (AAPL) and The Boeing Company (BA). Results are presented in Table 12. The first thing we notice is that the transition parameter ϕ is lower for stocks returns, while c is much more negative. Combining these, we find that the unconditional level of the log-volatility, given by $\frac{c}{1-\phi}$, does not differ significantly between stocks and indices. However, the volatility of the log-volatility, σ_n, is much higher for stocks than for indices, with values around 0.3 for both the AAPL returns and BA returns compared to 0.2 and 0.15 for the SP500 and FTSE, respectively. Thus, shocks to log-volatility are higher for stock returns, but die out more quickly. As returns of stocks are more extreme than index returns, this is as expected.

The leverage parameter estimations are quite surprising. When only estimated contempo-

raneous leverage ρ_0, the coefficient of around -0.13 is twice as small as when estimating only contemporaneous leverage ρ_1. In our simulation studies, this only happened when there was no contemporaneous leverage present. Furthermore, when estimating model (4), with both contemporaneous and one inter-temporal leverage effect, both the new SV model and Catania model estimate only a significant negative value for ρ_1. Furthermore, when estimating model (5) for BA with both ρ_1 and ρ_2, both the Catania model and the new SV model estimates equal values of the ρ_1 and ρ_2. Comparing this to the simulation results suggests that the BA returns have a small leverage effect at both inter-temporal times. For the AAPl returns, all results suggest that there is only a leverage effect at ρ_1.

6 Conclusion

In this paper, we review the different ways in which stochastic volatility models incorporate the leverage effect. The classical stochastic volatility model of Taylor (1986) is extended to incorporate contemporaneous leverage by Jacquier et al. (2004) as well as inter-temporal leverage by Harvey and Shephard (1996). Then, Yu (2005) and Catania (2020) extended these models by including both contemporaneous leverage and inter-temporal leverage. In these models, the leverage effect is modelled as the effect of past/present return shocks on the volatility shock. Catania shows that 70% to 90% of the shock to volatility is explained by past return shocks, meaning that the stochastic volatility models approach ARCH models in which past shocks to returns fully explain volatility. We argue that this assumption is wrong, as volatility shocks should contain only new information. Also, we argue that it is more likely that shocks to returns are affected by shocks to volatility, such that volatility shocks contain new information while shocks to returns are affected by these shocks to volatility. Furthermore, we disagree with the inter-temporal modelling of the leverage effect. It is unlikely that information is processed with such a delay, while we assume that markets are very efficient and incorporate all information immediately.

In an attempt to model the leverage effect correctly, we introduce a new model where current and future volatility shocks partly explain return shocks, instead of vice versa. While this model leads to autocorrelation when including multiple leverage effects, which is a characteristic not present in financial returns, we show that it could help investigate the leverage effect. Estimating this model on empirical data leads to only an inter-temporal leverage effect. This result of only one leverage effect is most likely to prevent autocorrelation. While we conclude that this new model is not an improvement from the other stochastic volatility models due to the resulting autocorrelation, it indicates that the leverage effect in financial returns is more likely to be inter-temporal than contemporaneous.

In an extensive simulation study, we show that this new stochastic volatility gives clear negative parameter values for all three leverage effects when data is simulated from the Catania model. As we do not find similar estimation results in the empirical application, we are not convinced that empirical data does contain such a leverage structure. Furthermore, we show that the Catania model cannot correctly distinguish the leverage effect when estimated on time series of representative length. Also, in the simulation study, we found that incorporating three leverage effects does not lead to an improvement in volatility predictions, but in fact, performs worse. In the empirical application, we find that incorporating three leverage effects does lead to improved predictions. Another indicator that the empirical data does not follow a process such as is assumed in the stochastic volatility models discussed in this paper.

Our results suggest that the leverage structure of financial returns is not equal to that which is assumed in stochastic volatility models. However, our new model is not an improvement. Thus, we encourage future research to find a model that better fits financial returns and the correlation structure between return and volatility shocks.

In our application to stock returns, we find that the leverage effect is much smaller for stocks than for indices. We argue that this result adds to the already abundant evidence that the leverage effect is not due to the financial leverage structure of assets but comes from a different relation between returns and volatility.

Literature has shown that the leverage effect is often not present in foreign exchanges. We find a small leverage effect in the Japanese Yen to United States dollar foreign exchange returns, while we find an inverted leverage effect of similar size in the United States dollar to United Kingdom pound sterling. This suggests that the price of the US dollar is highly correlated with the volatility of the exchanges. Namely, there is a positive relation between the price of the US dollar and the volatility of foreign exchanges.

Finally, we reconstruct the simulation studies of Lange (2020) to investigate the performance of the new Bellman filter by Lange (2020) with that of the particle filtering method by Malik and Pitt (2011). We show that the methods perform similarly when filtering state-space models. When parameters also have to be estimated, the particle filter performs slightly better. However, the Bellman filter is computationally far superior. Therefore, we conclude that the Bellman filter is preferably when applied to large datasets or repetitive tasks. The particle filter extension of Malik and Pitt (2011) can only be applied to one-dimensional state-space models. We expect the Bellman filter to be an even greater computational improvement from the particle filter when applied to multivariate state-space models. However, we leave this to future research to validate.

References

Akashi, H. and Kumamoto, H. (1977). Random sampling approach to state estimation in switching environments. *Automatica*, 13(4):429–434.

Bellman, R. (1957). A Markovian decision process. *Journal of Mathematics and Mechanics*, 6(5):679–684.

Black, F. (1976). Studies of stock market volatility changes. *Meeting of the Business and Economics. Statistics Section, American Statistical Association, Washington, DC.*

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

Campbell, J. Y. and Hentschel, L. (1992). No news is good news: An asymmetric model of changing volatility in stock returns. *Journal of Financial Economics*, 31(3):281–318.

Carpenter, J., Clifford, P., and Fearnhead, P. (1999). Improved particle filter for nonlinear problems. *IEE Proceedings-Radar, Sonar and Navigation*, 146(1):2–7.

Catania, L. (2020). A stochastic volatility model with a general leverage specification. *Journal of Business & Economic Statistics*, pages 1–23.

Chen, R. and Liu, J. S. (2000). Mixture Kalman filters. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 62(3):493–508.

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

Clark, P. K. (1973). A subordinated stochastic process model with finite variance for speculative prices. *Econometrica: Journal of the Econometric Society*, pages 135–155.

Del Moral, P. (2004). Feynman-Kac formulae. In *Feynman-Kac Formulae*, pages 47–93. Springer.

Doucet, A., De Freitas, N., and Gordon, N. (2001). An introduction to sequential Monte Carlo methods. In *Sequential Monte Carlo methods in practice*, pages 3–14. Springer.

Durbin, J. and Koopman, S. J. (2012). *Time series analysis by state space methods*. Oxford University Press.

Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica: Journal of the Econometric Society*, pages 987–1007.

Engle, R. F. and Ng, V. K. (1993). Measuring and testing the impact of news on volatility. *The Journal of Finance*, 48(5):1749–1778.

Figlewski, S. and Wang, X. (2000). Is the "leverage effect" a leverage effect?

French, K. R., Schwert, G. W., and Stambaugh, R. F. (1987). Expected stock returns and volatility. *Journal of Financial Economics*, 19(1):3–29.

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

Godsill, S. J., Doucet, A., and West, M. (2004). Monte Carlo smoothing for nonlinear time series. *Journal of the American Statistical Association*, 99(465):156–168.

Gordon, N., Ristic, B., and Arulampalam, S. (2004). Beyond the Kalman filter: Particle filters for tracking applications. *Artech House, London*, 830(5):1–4.

Gordon, N. J., Salmond, D. J., and Smith, A. F. M. (1993). Novel approach to nonlinear/non-Gaussian Bayesian state estimation. *IEE proceedings F (Radar and Signal Processing)*, 140(2):107–113.

Harvey, A., Ruiz, E., and Shephard, N. (1994). Multivariate stochastic variance models. *The Review of Economic Studies*, 61(2):247–264.

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

Hasanhodzic, J. and Lo, A. W. (2011). Black's leverage effect is not due to leverage. *Available at SSRN 1762363*.

Hull, J. and White, A. (1987). The pricing of options on assets with stochastic volatilities. *The Journal of Finance*, 42(2):281–300.

Hürzeler, M. and Künsch, H. R. (2001). Approximating and maximising the likelihood for a general state-space model. In *Sequential Monte Carlo methods in Practice*, pages 159–175. Springer.

Jacquier, E., Polson, N. G., and Rossi, P. E. (2002). Bayesian analysis of stochastic volatility models. *Journal of Business & Economic Statistics*, 20(1):69–87.

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

Kalman, R. E. (1960). A new approach to linear filtering and prediction problems. *American Society of Mechanical Engineers: Journal of Basic Engineering*, 82(1):3545.

Kantas, N., Doucet, A., Singh, S. S., Maciejowski, J., Chopin, N., et al. (2015). On particle methods for parameter estimation in state-space models. *Statistical Science*, 30(3):328–351.

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

Kitagawa, G. (1987). Non-Gaussian state—space modeling of nonstationary time series. *Journal of the American Statistical Association*, 82(400):1032–1041.

Kitagawa, G. (1996). Monte Carlo filter and smoother for non-Gaussian nonlinear state space models. *Journal of Computational and Graphical Statistics*, 5(1):1–25.

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

Kullback, S. and Leibler, R. A. (1951). On information and sufficiency. *The Annals of Mathematical Statistics*, 22(1):79–86.

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

Liu, J. and West, M. (2001). Combined parameter and state estimation in simulation-based filtering. In *Sequential Monte Carlo methods in practice*, pages 197–223. Springer.

Liu, J. S. and Chen, R. (1998). Sequential Monte Carlo methods for dynamic systems. *Journal of the American Statistical Association*, 93(443):1032–1044.

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

Meinhold, R. J. and Singpurwalla, N. D. (1983). Understanding the Kalman filter. *The American Statistician*, 37(2):123–127.

Melino, A. and Turnbull, S. M. (1990). Pricing foreign currency options with stochastic volatility. *Journal of Econometrics*, 45(1-2):239–265.

Meyer, R. and Yu, J. (2000). BUGS for a Bayesian analysis of stochastic volatility models. *The Econometrics Journal*, 3(2):198–215.

Nelson, D. B. (1988). *The time series behavior of stock market volatility and returns*. PhD thesis, Massachusetts Institute of Technology.

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

Nocedal, J. and Wright, S. (2006). *Numerical optimization*. Springer Science & Business Media.

Omori, Y., Chib, S., Shephard, N., and Nakajima, J. (2007). Stochastic volatility with leverage: Fast and efficient likelihood inference. *Journal of Econometrics*, 140(2):425–449.

Patton, A. J. (2011). Volatility forecast comparison using imperfect volatility proxies. *Journal of Econometrics*, 160(1):246–256.

Pearson, K. (1936). Method of moments and method of maximum likelihood. *Biometrika*, 28(1/2):34–59.

Polson, N. G., Stroud, J. R., and Müller, P. (2008). Practical filtering with sequential parameter learning. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 70(2):413–428.

Ruiz, E. (1994). Quasi-maximum likelihood estimation of stochastic volatility models. *Journal of Econometrics*, 63(1):289–306.

Smith, A. F. and Gelfand, A. E. (1992). Bayesian statistics without tears: a sampling–resampling perspective. *The American Statistician*, 46(2):84–88.

Tauchen, G. E. and Pitts, M. (1983). The price variability-volume relationship on speculative markets. *Econometrica: Journal of the Econometric Society*, pages 485–505.

Taylor, S. J. (1986). *Modelling financial time series*. Great Britain: Wiley.

Welch, G. and Bishop, G. (1995). An introduction to the Kalman filter.

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

Appendix A Catania model state-space representation

We have the following stochastic volatility model from Catania (2020):

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

We know that the observation errors ε_t are i.i.d. distributed. Thus, we can write the multivariate distribution of the state error η_t and the observation errors $\varepsilon_t, \ldots, \varepsilon_{t-m}$ as:

$$\begin{pmatrix} \varepsilon_t \\ \eta_t \\ \varepsilon_{t-1} \\ \ldots \\ \varepsilon_{t-m} \end{pmatrix} \sim \mathcal{N}\left\{ \begin{pmatrix} 0 \\ 0 \\ 0 \\ \ldots \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho_0 & 0 & \ldots & 0 \\ \rho_0 & 1 & \rho_1 & \ldots & \rho_m \\ 0 & \rho_1 & 1 & \ldots & 0 \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ 0 & \rho_m & 0 & \ldots & 1 \end{pmatrix} \right\}. \tag{A.2}$$

We first compute the conditional distribution of η_t given $\varepsilon_{t-1}, \ldots, \varepsilon_{t-m}$, which is normally distributed, $\eta_t | \varepsilon_{t-1}, \ldots, \varepsilon_{t-m} \sim \mathcal{N}(\mu_\eta, \Sigma_\eta)$, with mean μ_η and variance Σ_η given by:

$$\begin{aligned} \mu_\eta &= 0 + \begin{pmatrix} \rho_1 & \ldots & \rho_m \end{pmatrix} \begin{pmatrix} 1 & \ldots & 0 \\ \ldots & 1 & \ldots \\ 0 & \ldots & 1 \end{pmatrix}^{-1} \begin{pmatrix} \varepsilon_{t-1} \\ \ldots \\ \varepsilon_{t-m} \end{pmatrix} = \sum_{j=1}^{m} \rho_j \varepsilon_{t-j}, \\ \Sigma_\eta &= 1 - \begin{pmatrix} \rho_1 & \ldots & \rho_m \end{pmatrix} \begin{pmatrix} 1 & \ldots & 0 \\ \ldots & 1 & \ldots \\ 0 & \ldots & 1 \end{pmatrix}^{-1} \begin{pmatrix} \rho_1 \\ \ldots \\ \rho_m \end{pmatrix} = 1 - \sum_{j=1}^{m} \rho_j^2. \end{aligned} \tag{A.3}$$

We can write $\eta_t | \varepsilon_{t-1}, \ldots, \varepsilon_{t-m}$ as $\eta_t = \sum_{j=1}^{m} \rho_j \varepsilon_{t-j} + \sqrt{1 - \sum_{j=1}^{m} \rho_j^2}\, w_t$, where $w_t \sim \mathcal{N}(0,1)$. Now, from the state equation $y_t = \exp\left\{\frac{h_t}{2}\right\}\varepsilon_t$, we can write $\varepsilon_t = \exp\left\{\frac{-h_t}{2}\right\} y_t$. Replacing ε_{t-j} with this expression leads to:

$$\eta_t = \sum_{j=1}^{m} \rho_j \exp\left\{\frac{-h_{t-j}}{2}\right\} y_{t-j} + \sqrt{1 - \sum_{j=1}^{m} \rho_j^2}\, w_t. \tag{A.4}$$

Thus, the state equation from model (A.1) can be written as:

$$h_t = c + \phi h_{t-1} + \sigma_n \sum_{j=1}^{m} \rho_j y_{t-j} \exp\left\{-\frac{h_{t-j}}{2}\right\} + \sigma_n \sqrt{\sum_{j=1}^{m} \rho_j^2}\, w_t, \tag{A.5}$$

For the observation equation, we compute the conditional distribution of ε_t given η_t and the previous observation shocks $\varepsilon_{t-1}, \ldots, \varepsilon_{t-m}$, which is normally distributed, $\eta_t | \varepsilon_{t-1}, \ldots, \varepsilon_{t-m} \sim \mathcal{N}(\mu_\varepsilon, \Sigma_\varepsilon)$, with mean and variance given by:

$$\begin{aligned}
\mu_\varepsilon &= 0 + \begin{pmatrix} \rho_0 & 0 & \ldots & 0 \end{pmatrix} \begin{pmatrix} 1 & \rho_1 & \ldots & \rho_m \\ \rho_1 & 1 & \ldots & 0 \\ \ldots & \ldots & 1 & \ldots \\ \rho_m & 0 & \ldots & 1 \end{pmatrix}^{-1} \begin{pmatrix} \eta_t \\ \varepsilon_{t-1} \\ \ldots \\ \varepsilon_{t-m} \end{pmatrix} \\
&= \frac{\rho_0}{(1 - \sum_{j=1}^{m} \rho_j^2)} \left(\eta_t - \sum_{j=1}^{m} \rho_j \varepsilon_{t-j} \right), \\
\Sigma_\varepsilon &= 1 - \begin{pmatrix} \rho_0 & 0 & \ldots & 0 \end{pmatrix} \begin{pmatrix} 1 & \rho_1 & \ldots & \rho_m \\ \rho_1 & 1 & \ldots & 0 \\ \ldots & \ldots & 1 & \ldots \\ \rho_m & 0 & \ldots & 1 \end{pmatrix}^{-1} \begin{pmatrix} \rho_0 \\ 0 \\ \ldots \\ 0 \end{pmatrix} = 1 - \frac{\rho_0^2}{1 - \sum_{j=1}^{m} \rho_j^2}.
\end{aligned} \tag{A.6}$$

Therefore, we can write

$$\begin{aligned}
\varepsilon_t &= \frac{\rho_0}{(1 - \sum_{j=1}^{m} \rho_j^2)} \left(\eta_t - \sum_{j=1}^{m} \rho_j \varepsilon_{t-j} \right) + \sqrt{1 - \frac{\rho_0^2}{1 - \sum_{j=1}^{m} \rho_j^2}}\, u_t \\
&= \frac{\rho_0}{(1 - \sum_{j=1}^{m} \rho_j^2)} \left(\eta_t - \sum_{j=1}^{m} \rho_j \exp\left\{\frac{-h_{t-j}}{2}\right\} y_{t-j} \right) + \sqrt{1 - \frac{\rho_0^2}{1 - \sum_{j=1}^{m} \rho_j^2}}\, u_t,
\end{aligned} \tag{A.7}$$

where $u_t \sim \mathcal{N}(0,1)$. Finally, rewriting from the second equation in model (A.1), we get $\eta_t = \frac{1}{\sigma_n}(h_t - c - \phi h_{t-1})$. Pluggin this into equation (A.7) gives the observation equation and plugging this equation for ε_t in the observation equation in (A.1) gives:

$$\begin{aligned}
y_t = \exp\left\{\frac{h_t}{2}\right\} &\left[\frac{\rho_0}{\sigma_n (1 - \sum_{j=1}^{m} \rho_j^2)} \left(h_t - c - \phi h_{t-1} - \sigma_n \sum_{j=1}^{m} \rho_j y_{t-j} \exp\left\{-\frac{h_{t-j}}{2}\right\} \right) \right] \\
&+ \exp\left\{\frac{h_t}{2}\right\} \sqrt{1 - \frac{\rho_0^2}{1 - \sum_{j=1}^{m} \rho_j^2}}\, u_t,
\end{aligned} \tag{A.8}$$

which is the observation equation of the state-space representation of the Catania model.

Appendix B Score and information quantities of the SV model

We have the following stochastic volatility model:

$$\begin{aligned} y_t &= \exp\left\{\frac{h_t}{2}\right\}\left(\sum_{j=0}^{m}\rho_j\eta_{t+j} + \sqrt{1-\sum_{j=0}^{m}\rho_j^2}\xi_t\right), \qquad \xi_t \sim \mathcal{N}(0,1) \\ \boldsymbol{\alpha}_t &= \boldsymbol{c} + \boldsymbol{\Phi}\boldsymbol{\alpha}_{t-1} + \boldsymbol{\xi}_t, \qquad \boldsymbol{\Xi}_t \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{Q}), \end{aligned} \tag{B.1}$$

such that conditional distribution of $y_t|\boldsymbol{\alpha}_t, \mathcal{I}_{t-1}$ is nonlinear and Gaussian, given by:

$$\begin{aligned} y_t|\boldsymbol{\alpha}_t, \mathcal{I}_{t-1} &\sim \mathcal{N}(\mu_{yt}, \sigma_{yt}^2), \\ \mu_{yt} &= \exp\left\{\frac{h_t}{2}\right\}\left(\sum_{j=0}^{m}\rho_j\eta_{t+j}\right) \\ \sigma_{yt}^2 &= \exp\{h_t\}\left(1-\sum_{j=0}^{m}\rho_j^2\right). \end{aligned} \tag{B.2}$$

For simplicity, we leave $\mathcal{I}_{t-1}$ out from here on. The probability density function of $y_t|\boldsymbol{\alpha}_t$ is given by the usual Gaussian pdf:

$$p(y_t|\boldsymbol{\alpha}_t) = \frac{1}{\sigma_{yt}\sqrt{2\pi}}\exp\left\{-\frac{1}{2}\frac{(y_t-\mu_{yt})^2}{\sigma_{yt}^2}\right\}, \tag{B.3}$$

with corresponding logarithm:

$$\ell(y_t|\boldsymbol{\alpha}_t) = -\log(\sigma_{yt}) - \frac{1}{2}\frac{(y_t-\mu_{yt})^2}{\sigma_{yt}^2} = -\frac{h_t}{2} - \frac{1}{2}\frac{(y_t-\mu_{yt})^2}{\sigma_{yt}^2}, \tag{B.4}$$

where we leave out the constants $-\frac{1}{2}\log(2\pi)$ and $-\frac{1}{2}\log\left(1-\sum_{j=0}^{m}\rho_j^2\right)$. To compute score and information quantities, we first compute the derivatives of the mean μ_{yt} and the variance σ_{yt}^2 with respect to the state variables:

$$\begin{aligned} \frac{\mathrm{d}\mu_{yt}}{\mathrm{d}h_t} &= \frac{1}{2}\exp\left\{\frac{h_t}{2}\right\}\left(\sum_{i=0}^{m}\rho_i\eta_{t+j}\right) = \frac{1}{2}\mu_{yt}, \\ \frac{\mathrm{d}\mu_{yt}}{\mathrm{d}\eta_{t+j}} &= \rho_j\exp\left\{\frac{h_t}{2}\right\}, && \text{for } j=0,\ldots,m, \\ \frac{\mathrm{d}\sigma_{yt}^2}{\mathrm{d}h_t} &= \exp\{h_t\}\left(1-\sum_{i=0}^{m}\rho_i^2\right) = \sigma_{yt}^2, \\ \frac{\mathrm{d}\sigma_{yt}^2}{\mathrm{d}\eta_{t+j}} &= 0 && \text{for } j=0,\ldots,m. \end{aligned} \tag{B.5}$$

Furthermore, we note that

$$\frac{\mathrm{d}\sigma_{yt}^{-2}}{\mathrm{d}h_t} = -\exp\{-h_t\}\left(1-\sum_{i=0}^{m}\rho_i^2\right)^{-1} = \sigma_{yt}^{-2} = -\frac{1}{\sigma_{yt}^2} \tag{B.6}$$

The score is given by:

$$\frac{\mathrm{d}\ell(y_t|\boldsymbol{\alpha}_t)}{\mathrm{d}\boldsymbol{\alpha}_t} = \left[\frac{\mathrm{d}\ell(y_t|\boldsymbol{\alpha}_t)}{\mathrm{d}h_t}, \frac{\mathrm{d}\ell(y_t|\boldsymbol{\alpha}_t)}{\mathrm{d}\eta_{t+m}}, \ldots, \frac{\mathrm{d}\ell(y_t|\boldsymbol{\alpha}_t)}{\mathrm{d}\eta_t}\right], \tag{B.7}$$

which can now easily be evaluated:

$$\begin{aligned}
\frac{\mathrm{d}\ell(y_t|\boldsymbol{\alpha}_t)}{\mathrm{d}h_t} &= -\frac{1}{2} - \frac{1}{2}(y_t-\mu_{yt})^2\frac{\mathrm{d}\sigma_{yt}^{-2}}{\mathrm{d}h_t} + \frac{(y_t-\mu_{yt})}{\sigma_{yt}^2}\frac{\mathrm{d}\mu_{yt}}{\mathrm{d}h_t} \\
&= -\frac{1}{2} + \frac{1}{2}\frac{(y_t-\mu_{yt})^2}{\sigma_{yt}^2} - \frac{1}{2}\frac{(y_t-\mu_{yt})}{\sigma_{yt}^2}\mu_{yt} \\
&= -\frac{1}{2} + \frac{1}{2}\frac{(y_t-\mu_{yt})}{\sigma_{yt}^2}y_t, \\
\frac{\mathrm{d}\ell(y_t|\boldsymbol{\alpha}_t)}{\mathrm{d}\eta_{t+j}} &= \frac{(y_t-\mu_{yt})}{\sigma_{yt}^2}\frac{\mathrm{d}\mu_{yt}}{\mathrm{d}\eta_{t+j}} = \frac{(y_t-\mu_{yt})}{\sigma_{yt}^2}\rho_j\exp\left\{\frac{h_t}{2}\right\} \\
&= \frac{(y_t-\mu_{yt})}{\sigma_{yt}}\frac{\rho_j}{\sqrt{1-\sum_{i=0}^{m}\rho_i^2}}, \qquad \text{for } j=0,\ldots,m.
\end{aligned} \tag{B.8}$$

Next, we calculate the realized information matrix, which is given by:

$$-\frac{\mathrm{d}^2\ell(y_t|\boldsymbol{\alpha}_t)}{\mathrm{d}\boldsymbol{\alpha}_t\mathrm{d}\boldsymbol{\alpha}_t'} = \begin{bmatrix}
-\frac{\mathrm{d}^2\ell(y_t|\boldsymbol{\alpha}_t)}{\mathrm{d}h_t^2} & -\frac{\mathrm{d}^2\ell(y_t|\boldsymbol{\alpha}_t)}{\mathrm{d}h_t\mathrm{d}\eta_{t+m}} & \cdots & -\frac{\mathrm{d}^2\ell(y_t|\boldsymbol{\alpha}_t)}{\mathrm{d}h_t\mathrm{d}\eta_t} \\
-\frac{\mathrm{d}^2\ell(y_t|\boldsymbol{\alpha}_t)}{\mathrm{d}h_t\mathrm{d}\eta_{t+m}} & -\frac{\mathrm{d}^2\ell(y_t|\boldsymbol{\alpha}_t)}{\mathrm{d}\eta_{t+m}^2} & \cdots & -\frac{\mathrm{d}^2\ell(y_t|\boldsymbol{\alpha}_t)}{\mathrm{d}\eta_{t+m}\mathrm{d}\eta_t} \\
\cdots & \cdots & \cdots & \cdots \\
-\frac{\mathrm{d}^2\ell(y_t|\boldsymbol{\alpha}_t)}{\mathrm{d}h_t\mathrm{d}\eta_t} & -\frac{\mathrm{d}^2\ell(y_t|\boldsymbol{\alpha}_t)}{\mathrm{d}\eta_{t+m}\mathrm{d}\eta_t} & \cdots & -\frac{\mathrm{d}^2\ell(y_t|\boldsymbol{\alpha}_t)}{\mathrm{d}\eta_t^2}
\end{bmatrix}. \tag{B.9}$$

For which we compute the elements by:

$$
\begin{aligned}
-\frac{\mathrm{d}^2\ell(y_t|\boldsymbol{\alpha}_t)}{\mathrm{d}h_t^2} &= -\left(\frac{1}{2}(y_t-\mu_{yt})y_t\frac{\mathrm{d}\sigma_{yt}^{-2}}{\mathrm{d}h_t} - \frac{1}{2}\frac{y_t}{\sigma_{yt}^2}\frac{\mathrm{d}\mu_{yt}}{\mathrm{d}h_t}\right) = \frac{1}{2}\frac{(y_t-\mu_{yt})y_t}{\sigma_{yt}^2} + \frac{1}{4}\frac{y_t\mu_{yt}}{\sigma_{yt}^2} \\
&= \frac{1}{2}\frac{y_t^2}{\sigma_{yt}^2} - \frac{1}{4}\frac{y_t\mu_{yt}}{\sigma_{yt}^2}, \\
-\frac{\mathrm{d}^2\ell(y_t|\boldsymbol{\alpha}_t)}{\mathrm{d}h_t\mathrm{d}\eta_{t+j}} &= -\left(-\frac{1}{2}\frac{y_t}{\sigma_{yt}^2}\frac{\mathrm{d}\mu_{yt}}{\mathrm{d}\eta_{t+j}}\right) = \frac{1}{2}\frac{y_t}{\sigma_{yt}^2}\rho_j\exp\left\{\frac{h_t}{2}\right\} \\
&= \frac{1}{2}\frac{y_t}{\sigma_{yt}}\frac{\rho_j}{\sqrt{1-\sum_{i=0}^m\rho_i^2}}, \qquad \text{for } j=0,\ldots,m, \\
-\frac{\mathrm{d}^2\ell(y_t|\boldsymbol{\alpha}_t)}{\mathrm{d}\eta_{t+j}^2} &= -\left(-\frac{\rho_j}{\sqrt{1-\sum_{i=0}^m\rho_i^2}}\frac{1}{\sigma_{yt}}\frac{\mathrm{d}\mu_{yt}}{\mathrm{d}\eta_{t+j}}\right) = \frac{\rho_j}{\sqrt{1-\sum_{i=0}^m\rho_i^2}}\frac{1}{\sigma_{yt}}\rho_j\exp\left\{\frac{h_t}{2}\right\} \\
&= \frac{\rho_j^2}{1-\sum_{i=0}^m\rho_i^2}, \qquad \text{for } j=0,\ldots,m, \\
-\frac{\mathrm{d}^2\ell(y_t|\boldsymbol{\alpha}_t)}{\mathrm{d}\eta_{t+j}\mathrm{d}\eta_{t+k}} &= -\left(-\frac{\rho_j}{\sqrt{1-\sum_{i=0}^m\rho_i^2}}\frac{1}{\sigma_{yt}}\frac{\mathrm{d}\mu_{yt}}{\mathrm{d}\eta_{t+k}}\right) = \frac{\rho_j}{\sqrt{1-\sum_{i=0}^m\rho_i^2}}\frac{1}{\sigma_{yt}}\rho_k\exp\left\{\frac{h_t}{2}\right\} \\
&= \frac{\rho_j\rho_k}{1-\sum_{i=0}^m\rho_i^2}, \qquad \text{for } j=0,\ldots,m, k=0,\ldots,m, j\neq k.
\end{aligned}
\tag{B.10}
$$

Finally, the expected information matrix is given by:

$$
E\left[-\frac{\mathrm{d}^2\ell(y_t|\boldsymbol{\alpha}_t)}{\mathrm{d}\boldsymbol{\alpha}_t\mathrm{d}\boldsymbol{\alpha}_t'}\middle|\boldsymbol{\alpha}_t\right] = E\left[\left[\begin{array}{cccc}
-\frac{\mathrm{d}^2\ell(y_t|\boldsymbol{\alpha}_t)}{\mathrm{d}h_t^2} & -\frac{\mathrm{d}^2\ell(y_t|\boldsymbol{\alpha}_t)}{\mathrm{d}h_t\mathrm{d}\eta_{t+m}} & \cdots & -\frac{\mathrm{d}^2\ell(y_t|\boldsymbol{\alpha}_t)}{\mathrm{d}h_t\mathrm{d}\eta_t} \\
-\frac{\mathrm{d}^2\ell(y_t|\boldsymbol{\alpha}_t)}{\mathrm{d}h_t\mathrm{d}\eta_{t+m}} & -\frac{\mathrm{d}^2\ell(y_t|\boldsymbol{\alpha}_t)}{\mathrm{d}\eta_{t+m}^2} & \cdots & -\frac{\mathrm{d}^2\ell(y_t|\boldsymbol{\alpha}_t)}{\mathrm{d}\eta_{t+m}\mathrm{d}\eta_t} \\
\cdots & \cdots & \cdots & \cdots \\
-\frac{\mathrm{d}^2\ell(y_t|\boldsymbol{\alpha}_t)}{\mathrm{d}h_t\mathrm{d}\eta_t} & -\frac{\mathrm{d}^2\ell(y_t|\boldsymbol{\alpha}_t)}{\mathrm{d}\eta_{t+m}\mathrm{d}\eta_t} & \cdots & -\frac{\mathrm{d}^2\ell(y_t|\boldsymbol{\alpha}_t)}{\mathrm{d}\eta_t^2}
\end{array}\right]\middle|\boldsymbol{\alpha}_t\right],
\tag{B.11}
$$

for which we only have to take the expectations of the elements of the realized information, given in B.12.

$$
\begin{aligned}
E\left[-\frac{\mathrm{d}^2\ell(y_t|\boldsymbol{\alpha}_t)}{\mathrm{d}h_t^2}\middle|\boldsymbol{\alpha}_t\right] &= \frac{1}{2}\frac{(\mu_{yt}^2+\sigma_{yt}^2)}{\sigma_{yt}^2} - \frac{1}{4}\frac{\mu_{yt}^2}{\sigma_{yt}^2} = \frac{1}{2} + \frac{1}{4}\frac{\mu_{yt}^2}{\sigma_{yt}^2}, \\
E\left[-\frac{\mathrm{d}^2\ell(y_t|\boldsymbol{\alpha}_t)}{\mathrm{d}h_t\mathrm{d}\eta_{t+j}}\middle|\boldsymbol{\alpha}_t\right] &= \frac{1}{2}\frac{\mu_{yt}}{\sigma_{yt}}\frac{\rho_j}{\sqrt{1-\sum_{i=0}^m\rho_i^2}}, \qquad \text{for } j=0,\ldots,m, \\
E\left[-\frac{\mathrm{d}^2\ell(y_t|\boldsymbol{\alpha}_t)}{\mathrm{d}\eta_{t+j}^2}\middle|\boldsymbol{\alpha}_t\right] &= \frac{\rho_j^2}{1-\sum_{i=0}^m\rho_i^2}, \qquad \text{for } j=0,\ldots,m, \\
E\left[-\frac{\mathrm{d}^2\ell(y_t|\boldsymbol{\alpha}_t)}{\mathrm{d}\eta_{t+j}\mathrm{d}\eta_{t+k}}\middle|\boldsymbol{\alpha}_t\right] &= \frac{\rho_j\rho_k}{1-\sum_{i=0}^m\rho_i^2}, \qquad \text{for } j=0,\ldots,m, k=0,\ldots,m, j\neq k.
\end{aligned}
\tag{B.12}
$$

Appendix C Estimation on foreign exchange returns

Table C.1: Estimated parameters from the Catania model and the new SV model on foreign exchange returns

New SV model						Catania model					
c	ϕ	σ_n	ρ_0	ρ_1	ρ_2	c	ϕ	σ_n	ρ_0	ρ_1	ρ_2
Panel A: United States dollar vs British pound sterling (USD/Pound)											
-0.192	0.981	0.135	-	-	-	-0.170	0.984	0.120	-	-	-
0.042	0.004	0.014	-	-	-	0.069	0.007	0.010	-	-	-
-0.204	0.980	0.137	0.155	-	-	-0.189	0.979	0.125	0.197	-	-
0.041	0.004	0.013	0.050	-	-	0.004	0.001	0.004	0.005	-	-
-0.204	0.980	0.137	-	0.087	-	-0.179	0.983	0.124	-	0.091	-
0.042	0.004	0.014	-	0.053	-	0.050	0.005	0.012	-	0.080	-
-0.206	0.980	0.137	0.145	0.026	-	-0.170	0.984	0.122	0.149	-0.055	-
0.043	0.004	0.014	0.054	0.050	-	0.002	0.001	0.004	0.015	0.010	-
-0.206	0.980	0.138	-	0.078	0.015	-0.173	0.984	0.125	-	0.183	-0.108
0.042	0.004	0.013	-	0.061	0.059	0.033	0.003	0.004	-	0.142	0.062
-0.205	0.980	0.063	0.556	0.204	-0.431	-0.199	0.973	0.219	0.128	0.183	-0.159
0.001	0.000	0.006	0.012	0.009	0.013	0.007	0.001	0.008	0.017	0.021	0.028
Panel B: Japanese yen vs United States dollar (YEN/USD)											
-0.427	0.928	0.287	-	-	-	-0.432	0.958	0.201	-	-	-
0.053	0.025	0.055	-	-	-	0.028	0.003	0.008	-	-	-
-0.429	0.958	0.261	-0.212	-	-	-0.462	0.956	0.208	-0.199	-	-
0.036	0.013	0.033	0.036	-	-	0.002	0.002	0.006	0.002	-	-
-0.506	0.950	0.301	-	-0.142	-	-0.458	0.955	0.206	-	-0.160	-
0.089	0.018	0.042	-	0.038	-	0.033	0.003	0.028	-	0.051	-
-0.456	0.955	0.267	-0.199	-0.021	-	-0.409	0.956	0.202	-0.260	-0.032	-
0.054	0.015	0.037	0.045	0.041	-	0.016	0.002	0.008	0.039	0.026	-
-0.499	0.951	0.300	-	-0.187	0.056	-0.436	0.958	0.213	-	-0.323	0.191
0.076	0.017	0.039	-	0.048	0.043	0.028	0.003	0.013	-	0.053	0.060
-0.448	0.955	0.245	-0.499	-0.211	0.446	-0.432	0.956	0.324	-0.279	-0.078	0.182
0.068	0.007	0.017	0.013	0.010	0.011	0.077	0.021	0.008	0.019	0.024	0.028

NOTE: This table shows parameter estimations of the stochastic volatility models from Section 4.1 and Section 4.2 on returns of two major foreign exchanges. We vary the leverage effects in the stochastic volatility models between zero and three. The estimated parameters are shown in this table, with below their estimated standard errors using MATLAB's fminunc function.