Volatility Takes the Lead

A New Approach to Stochastic Volatility Models

Jippe van Dunné - 545849

August 10, 2021

Abstract

We introduce a new stochastic volatility model with leading, contemporaneous and lagging correlation between volatility and return shocks as well as a dedicated parameter for the median of the returns. We show the latter to impact the estimated contemporaneous correlation through disentangling the mean and skewness of the returns. We use the recently proposed Bellman filter of Lange (2021) and the particle filter of Malik and Pitt (2011), the latter of which we find to perform less efficiently. The model is fitted on historical data from 1990 through 2020 of daily returns of seven stock indices. Results reveal that our model captures stylistic effects such as leverage, volatility feedback and risk premia through the lagging, contemporaneous and leading volatility correlation parameters. Furthermore, correlation between filtered volatility and VIX shocks doubles compared to models without the added parameters. Lastly, we analyse predictive power and show that, in some cases, our model outperforms more conventional volatility models.

Supervised by:
dr. Rutger-Jan Lange
Second assessor:
prof. dr. Dick van Dijk

Contents

1 Introduction

Volatility modeling plays a central role in financial econometrics and option pricing, since changes in volatility have a great impact on risk assessment and derivative prices. Therefore, accurate modeling is important for policymakers and financial market participants alike. Models for varying volatility fall in roughly two categories: (i) observation-driven models such as autoregressive conditional heteroskedasticity (ARCH), proposed by Engle (1982), and its generalized counterpart (GARCH), proposed by Bollerslev (1986), and (ii) stochastic volatility (SV) models as pioneered by Taylor (1986) as an alternative to the former category. Whereas (G)ARCH-type models make conditional volatility a deterministic function of its lag(s) and possibly lagged squared return(s), SV models allow this volatility to follow a latent, stochastic process. The latter has the advantage that is fits more naturally in modern economic theory and is analogous to continuous-time models. Furthermore, research has shown that SV models can price European call options on currencies more accurately than GARCH-type model (Melino & Turnbull, 1990), provide a better in-sample fit with fewer parameters (Danielsson, 1994; Kim et al., 1998) and provide more flexibility in capturing leptokurtosis and autocorrelation (Fridman & Harris, 1998; Geweke, 1989; Ruiz et al., 2001).

In its most basic form, the SV model generates a series of independent and identically distributed (i.i.d.) return shocks multiplied by the latent volatility state, which evolves according to an AR(1) process and its own i.i.d. shocks. However, this basic formulation is often seen as an oversimplification as it cannot capture market effects such as 'leverage', which is the phenomenon that negative returns seem to cause high volatility on the next day[1]. The reverse effect is called 'volatility feedback', where volatility increases and therefore volatility risk premia should increase, requiring an immediate price decline. To capture such effects, multiple model extensions have been suggested, which allow for dependence between return and state (i.e. volatility) innovations. To incorporate leverage, Harvey and Shephard (1996) propose an

[1]Classically, this effect is coined the leverage effect, because of a hypothesis developed in Black (1976) and Christie (1982). This hypothesis stipulates that negative returns drive down equity, which increases the leverage (debt/equity) of companies, which in turn increases the size of its returns. However, this hypothesis has been mostly discarded, see Bekaert and Wu (2000) for a comprehensive overview, but the name stuck.

intertemporal correlation parameter ρ_1 to create a one-day (as indicated by the subscript) delayed volatility response, where state innovations are correlated with the return innovations of the previous time step. Jacquier et al. (2004) suggest contemporaneous dependence through correlation parameter ρ_0, which implies that volatility shocks are contemporaneously related to return shocks. Furthermore this relation introduces skewness and therefore a non-zero mean in the returns, as the size and sign of returns impact each other. These two models are compared by Yu (2005), who finds that the intertemporal specification should be preferred from a likelihood perspective, dismissing contemporaneous correlation. He also suggests an encompassing model containing both ρ_1 and ρ_0, which he discards as it performs only marginally better than the strictly intertemporal model. Catania (2020) proposes a more flexible model, in which the number of correlation parameters and thus state-lags (i.e. $\rho_1, \rho_2, ..., \rho_n$) is determined by information criteria. Fitting this to daily index returns, he finds that more lags often lead to a better-quality model. He also finds that ρ_0 is significant, in contrast with the findings of Yu (2005). However, since contemporaneous correlation introduces skewness in the returns, it implicitly defines the mean of the returns. The models of Catania (2020), Jacquier et al. (2004), and Yu (2005) all neglect the mean of the returns as it is much smaller than its variance, but failure to include a dedicated mean parameter effectively constrains ρ_0.

For the above reason, we propose a new model that incorporates a model parameter for the median of the returns, which allows us to set the mean separately from the contemporaneous correlation. Furthermore, we allow for a freer intertemporal correlation structure by admitting leading volatility correlation ($\rho_{-1}, \rho_{-2}, ..., \rho_{-m}$) as opposed to strictly contemporaneous (ρ_0) and lagging ($\rho_1, \rho_2, ..., \rho_n$) correlation parameters. This added opposite intertemporal correlation has not yet been assessed to the best of our knowledge and implies that return innovations can be affected by past volatility shocks, but themselves can also influence future volatility shocks. We are of the belief that (strong) changes in volatility can affect future returns. Economically, this implies that market volatility can drive fear among investors, leading to a self-fulfilling prophecy of negative returns, in line with the volatility feedback hypothesis. This fits in with the classical finance hypothesis that an increase in volatility should lead to higher returns, as increased risk should lead to a higher premium. This risk premium is allowed for by an initial negative return,

following the rise in volatility, to 'pave the way' for future positive returns. As far as we are aware, this model is new, as these properties have not yet been assessed before in SV models.

I fit this model by maximum-likelihood optimisation as should be preferred according to Broto and Ruiz (2004) with an approximate filter and a simulated maximum-likelihood method. The approximate filter is the recently proposed Bellman filter by Lange (2021), which is built around a recursively computed (ergo efficient) approximate mode estimator. We choose this online filter for its quick calculation and estimation time since computational complexity is $O(s^3t)$ with s the dimension of the state and t the length of the time sample. Furthermore, the results of Lange (2021) show remarkable accuracy when compared with a state-of-the-art importance sampler and particle filter in the univariate case. According to Lange, this efficiency should extend to multidimensional cases and therefore allow us to evade 'the curse of dimensionality', which is the phenomenon that computational complexity scales exponentially with the state dimension (Bellman, 1957). The simulated maximum-likelihood method is the Sequential Monte Carlo method or particle filter of Malik and Pitt (2011), of which the likelihood estimate is continuous in the parameter space, allowing for gradient-based optimisation. In contrast with the Bellman filter, this filter is asymptotically exact as the number of particles goes to infinity. It estimates the state online by simulating multiple instances (particles) from the Markovian state transition equation and assessing their posterior likelihood with incoming data using the measurement equation. However, this method does suffer to a degree from the curse of dimensionality, as the number of required particles tends to increase quickly with the state dimension, though less than exponentially as would typically be the case for importance samplers[2].

In a simulation study, we find that parameters of five different data generating processes are estimated accurately, although the particle filter suffers from estimation problems, most likely due to the required number of particles not being reached within reasonable computation times and possibly the degeneracy of the state equation (Künsch, 2013; J. Liu & West, 2001). Further-

[2]The exact computational scaling of particle filters remains a disputed subject, as no theoretical bounds have been found with respect to the dimension. Research has shown that with a good proposal density and for certain filtering problems, particle filters can beat the curse of dimensionality, but not generally so. For a comprehensive overview, see Crisan and Doucet (2002) and Daum (2005).

more, this method requires high-quality starting values, which we find using the Bellman filter. The fact that the particle filter cannot be used as a standalone method is a serious drawback.

Fitting our model to daily returns of seven stock indices, we find that our model is able to jointly capture the stylistic leverage, volatility feedback and volatility premia with lagging, contemporaneous and leading correlation parameters. We find that the contemporaneous correlation is dominant in most indices, contradicting findings of Yu (2005). We believe this is due to the introduction of a median parameter, which disentangles the mean of the returns from the effect of the contemporaneous correlation parameter. Furthermore, we observe 66.3% correlation between filtered volatility shocks of the S&P 500 and changes in the VIX using our new model, compared to only 31.5% when leaving out the median parameter.

Finally, we compare predictive power using the loss functions of Patton (2011) and the test of Diebold and Mariano (1995) and West (1996) of our model to an SV model without leading volatility shocks, an SV model without correlation parameters and the asymmetric GARCH model of Glosten et al. (1993) and find that on the S&P 500 our model outperforms all the other models at the 5% significance level. On the FTSE 100, our model still comes out on top, but the performance differences narrow and the model cannot significantly beat all others. On two of the other five indices, our model performs the best of the SV models, but no significant results are found. These results suggest that, on certain indices, our model can contribute a valuable improvement in terms of forecasting. Furthermore, we conclude that our model provides economic insight into the relation between volatility and returns and the associated effects, which we believe to be unprecedented in the domain of stochastic volatility models.

2 The Model

Let y_t be the financial log return at time t, where t is in days. We then assume y_t to be generated according to the following discrete-time stochastic volatility process:

$$\begin{aligned} y_t &= \mu + \exp\left\{\frac{\lambda_t}{2}\right\} \cdot \widetilde{\varepsilon}_t, & \widetilde{\varepsilon}_t &\sim \mathrm{N}(0,1), \\ \lambda_t &= c + \phi\lambda_{t-1} + \sigma_\eta \cdot \eta_t, & \eta_t &\overset{iid}{\sim} \mathrm{N}(0,1), \\ \widetilde{\varepsilon}_t &= \sum_{i=-m}^{n} \rho_i \eta_{t+i} + \sqrt{1 - \sum_{i=-m}^{n} \rho_i^2} \cdot \varepsilon_t, & \varepsilon_t &\overset{iid}{\sim} \mathrm{N}(0,1), \end{aligned} \tag{1}$$

where μ accounts for the median of the returns, c an additive constant, $|\phi| < 1$ the persistence of the log-variance λ_t, $\sigma_\eta > 0$ the volatility of λ_t and $|\rho_i| < 1$ the correlations between return shock $\widetilde{\varepsilon}_t$ and state innovation η_{t+i} for $i = -m, m+1, ..., n$ with $m, n \geq 0$. Furthermore, we require $\sum_{i=-m}^{n} \rho_i^2 < 1$ to ensure that $\widetilde{\varepsilon}_t$ is real-valued. The correlations between the two standard normal random variables allow for shocks to affect each other at different points in time. Table 1 poses a small overview of the three central correlation parameters, their typical sign and economical interpretation. This includes the archetypal 'leverage effect', where a negative return shock at time t ($\widetilde{\varepsilon}_t < 0$) leads to a positive shock in volatility at $t+1$ ($\eta_{t+1} > 0$) through ρ_1. Contemporaneous correlation through ρ_0 can capture volatility feedback, as an increase in volatility can lead to an immediate price decline. Leading volatility correlation via ρ_{-1} is also possible, i.e. today's return is correlated with yesterday's volatility shock. Positive

Table 1: Correlation parameters from the model of Equation 1 for $m = 1, n = 1$ with their associated leading shock, possible economical interpretation and expected sign.

	Leading shock	Interpretation
ρ_1	return	leverage, typically < 0
ρ_0	simultaneous	volatility feedback, typically < 0
ρ_{-1}	volatility	delayed volatility feedback if < 0 risk premium if > 0

ρ_{-1} can show a risk premium for higher volatility or a delayed volatility feedback effect if this is negative. The specification of m and n determines by how many days a shock in the state λ_t is allowed to lead respectively lag a shock in the return y_t, with larger intertemporal differences representing more delayed effects.

This general stochastic volatility model encompasses some stochastic volatility models published in recent years. Constraining $m = 0$, $n = 0$, $\rho_0 = 0$ and $\mu = 0$ returns the classic stochastic volatility model with completely independent innovations and a zero median. The case $n = 1, m = 0$, $\rho_0 = 0$ and $\mu = 0$ gives the asymmetric stochastic volatility model of Harvey and Shephard (1996), which can account for intertemporal leverage. However, when we restrict $m = 0$, $n = 0$ and $\mu = 0$ but unfix ρ_0, we find the volatility model of Jacquier et al. (2004) with contemporaneous correlation. Yu (2005) unites the latter two in his encompassing model, which is a special case of the model presented here with constraints $m = 0, n = 1$ and $\mu = 0$. Recently, Catania (2020) studied stochastic volatility models with possibly longer lags (i.e. $n \geq 1$, $m = 0$, $\mu = 0$) and found that a better model fit can be acquired by this.

However, the existing models all neglect the non-zero mean introduced by the contemporaneous correlation parameter ρ_0, which it triggers through skewness in returns. Typically, we expect this parameter to be negative, as this implies that an increase in volatility tends to occur simultaneously with a negative return. However, since this negative skewness shifts the mean of the returns, we introduce the parameter μ to be able to set the mean and skewness independently.

We are of the belief that volatility can lead returns and thus present a generalised model which cannot only account for leverage effects, but also volatility feedback and the associated risk premia. Furthermore, we suspect that the literature underestimates ρ_0 by constraining $\mu = 0$, effectively setting the mean and skewness of the returns by the same parameter. Another advantage of this new model is its linear and Gaussian state transition equation, which makes for a straightforward implementation. This is in contrast with the model of Catania (2020), in which return shocks are i.i.d. but state shocks are not, resulting in the log-variance following an ARMA(1,m) process. Furthermore, this model is able to capture the serial correlation in the return shocks $\widetilde{\varepsilon}_t$, which is a well-established phenomenon even on short horizons as shown

by e.g. Lo and MacKinlay (1988) and Campbell et al. (1993) and which the model of Catania (2020) cannot capture.

2.1 Statistical Properties of the Model

Unconditionally, the log-variance λ_t is normally distributed as

$$\lambda_t \sim \mathrm{N}\left(\frac{c}{1-\phi}, \sigma_\lambda^2\right), \qquad \sigma_\lambda^2 = \frac{\sigma_\eta^2}{1-\phi^2}. \tag{2}$$

Since λ_t follows an AR(1) process, its autocovariance function is $\gamma^\lambda(j) = \mathrm{cov}(\lambda_t, \lambda_{t-j})$ equals

$$\gamma^\lambda(j) = \phi^j \sigma_\lambda^2. \tag{3}$$

The autocovariance function of $\widetilde{\varepsilon}$, $\gamma^{\widetilde{\varepsilon}}(j) = \mathrm{cov}(\widetilde{\varepsilon}_t, \widetilde{\varepsilon}_{t-j})$ for $j \geq 1$, is given by

$$\gamma^{\widetilde{\varepsilon}}(j) = \begin{cases} \sum_{l=j-m}^{n} \rho_l \rho_{l-j}, & \text{if } j \leq m+n \\ 0, & \text{otherwise,} \end{cases} \tag{4}$$

which, conditional on λ_t and λ_{t-j}, is equal to the autocorrelation of the returns. This means that only under certain restrictions on the correlation parameters, this serial correlation is equal to zero. Generally, the autocovariance in the returns $\gamma^{y_t}(j) = \mathrm{cov}(y_t, y_{t-j})$ for $j \geq 1$ is expressed as

$$\gamma^{y_t}(j) = \begin{cases} \exp\left(\frac{\lambda_t + \lambda_{t-j}}{2}\right) \cdot \gamma^{\widetilde{\varepsilon}}(j), & \text{if } j \leq m+n \\ 0, & \text{otherwise.} \end{cases} \tag{5}$$

The expected value of y_t can be unequal to zero, due to the skewness in the returns introduced by ρ_0, but is hard to compute. We rewrite the first line of Equation 1 by substituting the AR-variable λ_t with the infinite sum of its historical innovations and $\widetilde{\varepsilon}$ for the last line of Equation 1 and rewrite its first sum, giving

$$\mathbb{E}[y_t] = \mu + \mathbb{E}\left[\exp\left\{\frac{1}{2}(c + \sigma_\eta \sum_{i=0}^{\infty} \phi^i \eta_{t-i})\right\} \cdot \left\{\sum_{i=-m}^{n} \rho_i \eta_{t+i} + \sqrt{1 - \sum_{i=-m}^{n} \rho_i^2} \varepsilon_t\right\}\right]. \tag{6}$$

From this equation, it can be seen that the relation between the state innovations η_{t-i} for $i = 0, ..., \infty$ in λ_t and η_{t+i} for $i = -m, ..., n$ in $\widetilde{\varepsilon}$ lead to a non-zero expectation on the right-hand side, depending on the values of ρ_i for $i = -m, ..., 0$. Parameter μ serves to let the optimiser freely estimate correlation between innovations, without implicitly fixing the mean of the returns.

2.2 State-space Formulation

For the estimation approaches we use, it is convenient to write the state transition equation of the model of Equation 1 in the state space form

$$\boldsymbol{\alpha}_t = \boldsymbol{c} + \boldsymbol{T}\boldsymbol{\alpha}_{t-1} + \boldsymbol{\eta}_t, \qquad \boldsymbol{\eta}_t \overset{iid}{\sim} \mathrm{N}(\mathbf{0}, \boldsymbol{Q}), \tag{7}$$

where $\boldsymbol{\alpha}_t$ represents the state vector, $\boldsymbol{c}$ the intercept vector, $\boldsymbol{T}$ the transition matrix and $\boldsymbol{\eta}_t$ the multivariate shock vector with covariance matrix $\boldsymbol{Q}$. The dimension of $\boldsymbol{\alpha}$ we call s, being equal to $n+m+2$, as at least λ_t and η_{t+1} are required in $\boldsymbol{\alpha}_t$. In the simple case of $n=1$ and $m=1$, this becomes

$$\begin{pmatrix} \lambda_t \\ \eta_{t+1} \\ \eta_t \\ \eta_{t-1} \end{pmatrix} = \begin{pmatrix} c \\ 0 \\ 0 \\ 0 \end{pmatrix} + \begin{pmatrix} \phi & \sigma_\eta & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} \cdot \begin{pmatrix} \lambda_{t-1} \\ \eta_t \\ \eta_{t-1} \\ \eta_{t-2} \end{pmatrix} + \begin{pmatrix} 0 \\ \eta_{t+1} \\ 0 \\ 0 \end{pmatrix}, \boldsymbol{Q} = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}. \tag{8}$$

It can be seen that part of the transition equation is deterministic, introducing degeneracy. The complete general state space formulation is rather hairy and can therefore be found in Appendix A.

3 Bellman Filter

The Bellman filter is an approximate filter based on the dynamic programming principle by Bellman (1957) applied to a mode estimator. It is a generalisation of the Kalman filter and applicable to a wider range of (non-linear non-Gaussian) models, such as our leveraged stochastic volatility model, while retaining the Kalman filter's computational simplicity. The latter is achieved by substituting high-dimensional numerical integrals associated with mean-based estimators for a dynamically programmed optimisation problem associated with mode-based estimators. This efficiency gain is traded for the fact that mode estimators are suboptimal for traditional loss functions such as the absolute or squared loss.

3.1 Concept

This filter estimates the state online in a recursive fashion. To illustrate this, let us follow Lange (2021) by defining the estimate of the time t hidden state $\boldsymbol{\alpha}_t$ (in Greek writing) as $\boldsymbol{a}_t$ (in Latin writing) and the value function as

$$V_t\left(\boldsymbol{a}_t\right) \equiv \max_{\boldsymbol{a}_{1:t-1} \in \mathbb{R}^{s \times (t-1)}} \ell(\boldsymbol{a}_{1:t}, y_{1:t}), \qquad \boldsymbol{a}_t \in \mathbb{R}^s, \tag{9a}$$

$$= \ell\left(y_t|\boldsymbol{a}_t\right) + \max_{\boldsymbol{a}_{t-1} \in \mathbb{R}^s} \left\{\ell(\boldsymbol{a}_t|\boldsymbol{a}_{t-1}) + V_{t-1}(\boldsymbol{a}_{t-1})\right\}, \tag{9b}$$

where the value V_t depends on the log-likelihood $\ell(\cdot)$ of the data $y_{1:t} \equiv (y_1, ..., y_t)$, the s-dimensional state estimate $\boldsymbol{a}_t$ and the arg max of $\boldsymbol{a}_{1:t-1} \equiv (\boldsymbol{a}_1, ..., \boldsymbol{a}_{t-1})$. Equation 9b is the Bellman equation applied to this function, from which it can be seen that this value can be calculated recursively through time, which makes this filtering method efficient.

Having assumed this value function, the Bellman-filtered state is defined as

$$\boldsymbol{a}_{t|t} \equiv \arg\max_{\boldsymbol{a}_t \in \mathbb{R}^s} \; V_t\left(\boldsymbol{a}_t\right), \tag{10}$$

where the subscript of the time t state estimate $\boldsymbol{a}_{t|t}$ indicates conditionality on y_t. This is the mode or 'maximum a posteriori' (MAP) estimate of the state $\boldsymbol{\alpha}_t$.

3.2 Filtering Algorithm

The filtering algorithm is presented in Algorithm 1. The first step is the initialisation of the variables, which we set to their unconditional mean (where $\mathbf{1}$ indicates the identity matrix of the appropriate size). After this preliminary step, we start filtering the state through time by alternately predicting, optimising and updating our state estimate.

For the prediction step, we compute the first-order condition of Equation 9b with respect to $\boldsymbol{a}_{t-1}$. For this, we need closed form expressions of $\ell(\boldsymbol{a}_t|\boldsymbol{a}_{t-1})$ and $V_{t-1}(\boldsymbol{a}_{t-1})$. Since the state transition equation of our model is linear and Gaussian, the former is given by

$$\ell\left(\boldsymbol{a}_t|\boldsymbol{a}_{t-1}\right) = -\frac{1}{2}\left(\boldsymbol{a}_t - \boldsymbol{c} - \boldsymbol{T}\boldsymbol{a}_{t-1}\right)' \boldsymbol{Q}^{-1}\left(\boldsymbol{a}_t - \boldsymbol{c} - \boldsymbol{T}\boldsymbol{a}_{t-1}\right) + \text{ constants}, \quad \boldsymbol{a}_{t-1}, \boldsymbol{a}_t \in \mathbb{R}^s. \tag{11}$$

However, the value function cannot be found exactly in general, an exception being the case when the model is linear and Gaussian. Then value function becomes multivariate quadratic and

we obtain Kalman's (1960) filter. Nonetheless, when this is not the case, the Bellman equation still holds, but an approximation is needed to solve it. Lange (2021) suggests a second-order polynomial approximation, which we follow:

$$V_{t-1}\left(\boldsymbol{a}_{t-1}\right) \approx -\frac{1}{2}\left(\boldsymbol{a}_{t-1}-\boldsymbol{a}_{t-1|t-1}\right)' \boldsymbol{I}_{t-1|t-1}\left(\boldsymbol{a}_{t-1}-\boldsymbol{a}_{t-1|t-1}\right)+ \text{ constants}, \quad \boldsymbol{a}_{t-1} \in \mathbb{R}^s, \tag{12}$$

where $\boldsymbol{a}_{t-1|t-1}$ and $\boldsymbol{I}_{t-1|t-1}$ are the posterior state estimate and its precision matrix of time $t-1$. This is not exact, but, according to Lange (2021), this approximation is generally sufficient around its peak, which is our region of interest. Substituting Equation 11 and 12 into Equation 9b, computing the first order condition and solving for $\boldsymbol{a}_{t-1}$, we get

$$\boldsymbol{a}_{t-1}^{*}=\left(\boldsymbol{I}_{t-1|t-1}+\boldsymbol{T}'\boldsymbol{Q}^{-1}\boldsymbol{T}\right)^{-1}\left\{\boldsymbol{I}_{t-1|t-1}\boldsymbol{a}_{t-1|t-1}+\boldsymbol{T}'\boldsymbol{Q}^{-1}\left(\boldsymbol{a}_{t}-\boldsymbol{c}\right)\right\}. \tag{13}$$

Substituting this back into Equation 9b, we get the value function that depends on the measurement likelihood $\ell(y_t|\boldsymbol{a}_t)$ and a regularising penalty term around the predicted state $\boldsymbol{a}_{t|t-1}$:

$$V_t\left(\boldsymbol{a}_t\right)=\ell\left(y_t|\boldsymbol{a}_t\right)-\frac{1}{2}\left(\boldsymbol{a}_t-\boldsymbol{a}_{t|t-1}\right)' \boldsymbol{I}_{t|t-1}\left(\boldsymbol{a}_t-\boldsymbol{a}_{t|t-1}\right)+ \text{ constants}, \quad \boldsymbol{a}_t \in \mathbb{R}^s. \tag{14}$$

Algorithm 1: Bellman filter for model with linear Gaussian state transition equation

Initialise: $\boldsymbol{a}_{0|0}=(\boldsymbol{1}-\boldsymbol{T})^{-1}\boldsymbol{c}$ and $\text{vec}(\boldsymbol{I}_{0|0}^{-1})=(\boldsymbol{1}-\boldsymbol{T}\otimes\boldsymbol{T})^{-1}\text{vec}(\boldsymbol{Q}^{-1})$

for t = *1:T* **do**

- *Predict:* $\boldsymbol{a}_{t|t-1}=\boldsymbol{c}+\boldsymbol{T}\boldsymbol{a}_{t-1|t-1}$ and $\boldsymbol{I}_{t|t-1}=\left(\boldsymbol{T}\boldsymbol{I}_{t-1|t-1}^{-1}\boldsymbol{T}'+\boldsymbol{Q}\right)^{-1}$
- *Optimise:* Set $\boldsymbol{a}_{t|t}^{(0)}=\boldsymbol{a}_{t|t-1}$ and $i=0$
- **for** $i<i_{\max}\vee\Delta^{(i-1)}<\epsilon$ **do**
  - $\boldsymbol{a}_{t|t}^{(i+1)}=\boldsymbol{a}_{t|t}^{(i)}+\left\{\boldsymbol{I}_{t|t-1}-\frac{\mathrm{d}^2\ell(y_t|\boldsymbol{a})}{\mathrm{d}\boldsymbol{a}\mathrm{d}\boldsymbol{a}'}\right\}^{-1}\left\{\frac{\mathrm{d}\ell(y_t|\boldsymbol{a})}{\mathrm{d}\boldsymbol{a}}-\boldsymbol{I}_{t|t-1}\left(\boldsymbol{a}-\boldsymbol{a}_{t|t-1}\right)\right\}\Big|_{a=a_{t|t}^{(i)}}$
  - $\Delta^{(i)}=\max\left|\left|\boldsymbol{a}_{t|t}^{(i+1)}-\boldsymbol{a}_{t|t}^{(i)}\right|\right|$
- **end**
- *Update:* $\boldsymbol{a}_{t|t}=\boldsymbol{a}_{t|t}^{(i)}$ and $\boldsymbol{I}_{t|t}=\boldsymbol{I}_{t|t-1}-\frac{\mathrm{d}^2\ell(y_t|\boldsymbol{a})}{\mathrm{d}\boldsymbol{a}\mathrm{d}\boldsymbol{a}'}\Big|_{\boldsymbol{a}=\boldsymbol{a}_{t|t}}$

end

Here, the state prediction and associated precision matrix $\boldsymbol{I}_{t|t-1}$ are defined as

$$\begin{aligned}
&\boldsymbol{a}_{t|t-1} = \boldsymbol{c} + \boldsymbol{T}\boldsymbol{a}_{t-1|t-1}, \\
&\boldsymbol{I}_{t|t-1} = \boldsymbol{Q}^{-1} - \boldsymbol{Q}^{-1}\boldsymbol{T}\left(\boldsymbol{I}_{t-1|t-1} + \boldsymbol{T}'\boldsymbol{Q}^{-1}\boldsymbol{T}\right)^{-1}\boldsymbol{T}'\boldsymbol{Q}^{-1} = \left(\boldsymbol{T}\boldsymbol{I}_{t-1|t-1}^{-1}\boldsymbol{T}' + \boldsymbol{Q}\right)^{-1},
\end{aligned} \tag{15}$$

where, in the last equality, we use the Woodbury matrix identity to obtain an expression that remains valid for singular $\boldsymbol{Q}$. It should be noted that these predicted quantities are the same as the Kalman filter's, but written in the information form. The regularisation term of Equation 14 introduces a bias towards the prediction, but reduces the variance of the maximisation step of Equation 9a, which James and Stein (1992) show may improve the efficiency of the estimator.

The optimisation step is executed by maximising the value function of Equation 14 with respect to $\boldsymbol{a}_t$. Whereas the Bellman filter's prediction was similar to the Kalman filter's, the optimisation step generally is not. Since our measurement equation is nonlinear, this maximisation is done numerically. Lange (2021) suggests using a simple iterative method such as Newton-Raphson, which we choose to employ here and reads

$$\boldsymbol{a}_{t|t}^{(i+1)} = \boldsymbol{a}_{t|t}^{(i)} + \left[-\frac{\mathrm{d}^2 V_t(\boldsymbol{a})}{\mathrm{d}\boldsymbol{a}\ \mathrm{d}\boldsymbol{a}'}\right]^{-1} \frac{\mathrm{d}V_t(\boldsymbol{a})}{\mathrm{d}\boldsymbol{a}}\bigg|_{\boldsymbol{a}=\boldsymbol{a}_{t|t}^{(i)}}, \tag{16}$$

where the first and second derivative of $V_t(\boldsymbol{a}_t)$ are respectively

$$\begin{aligned}
\frac{\mathrm{d}V_t(\boldsymbol{a})}{\mathrm{d}\boldsymbol{a}} &= \frac{\mathrm{d}\ell\left(y_t|\boldsymbol{a}\right)}{\mathrm{d}\boldsymbol{a}} - \boldsymbol{I}_{t|t-1}\left(\boldsymbol{a} - \boldsymbol{a}_{t|t-1}\right), \\
\frac{\mathrm{d}^2 V_t(\boldsymbol{a})}{\mathrm{d}\boldsymbol{a}\mathrm{d}\boldsymbol{a}'} &= \frac{\mathrm{d}^2\ell\left(y_t|\boldsymbol{a}\right)}{\mathrm{d}\boldsymbol{a}\mathrm{d}\boldsymbol{a}'} - \boldsymbol{I}_{t|t-1}.
\end{aligned} \tag{17}$$

The first- and second-order derivatives of the measurement log-likelihood with respect to the state estimate, as stated respectively in Equation 17, are referred to as the 'score' and the 'information'. For our model, these quantities are given in Appendix B. This optimisation scheme is executed until a maximum number of iteration $i_{\max}$, set to twenty, is reached or changes in the state become smaller than the precision variable ϵ, set to 10^{-5}. Generally, this precision is reached in fewer than five iterations.

The updating step consists of setting the 'optimal' state estimate $\boldsymbol{a}_{t|t}^{(i)}$ to be the filtered state estimate $\boldsymbol{a}_{t|t}$ and calculating the corresponding precision matrix $\boldsymbol{I}_{t|t}$ as the sum of the predicted information $\boldsymbol{I}_{t|t-1}$ and the realised information evaluated at filtered state $\boldsymbol{a}_{t|t}$.

3.3 Parameter Estimation

The estimation of the model parameters is done by maximising the decomposed log-likelihood, output from the Bellman filter. Lange (2021) introduces this decomposition for a single measurement as

$$\ell\left(y_t|\mathcal{F}_{t-1}\right) = \ell\left(y_t, \boldsymbol{\alpha}_t|\mathcal{F}_{t-1}\right) - \ell\left(\boldsymbol{\alpha}_t|y_t, \mathcal{F}_{t-1}\right) = \ell\left(y_t|\boldsymbol{\alpha}_t\right) + \ell\left(\boldsymbol{\alpha}_t|\mathcal{F}_{t-1}\right) - \ell\left(\boldsymbol{\alpha}_t|\mathcal{F}_t\right). \tag{18}$$

Since $\boldsymbol{\alpha}_t$ is latent, we evaluate the log-likelihood at the filtered state estimate $\boldsymbol{a}_{t|t}$ which, with some rearranging, yields

$$\ell\left(y_t|\mathcal{F}_{t-1}\right) = \ell\left(y_t|\boldsymbol{\alpha}_t\right)\big|_{\boldsymbol{\alpha}_t=\boldsymbol{a}_{t|t}} - \left\{\ell\left(\boldsymbol{\alpha}_t|\mathcal{F}_t\right) - \ell\left(\boldsymbol{\alpha}_t|\mathcal{F}_{t-1}\right)\right\}\big|_{\boldsymbol{\alpha}_t=\boldsymbol{a}_{t|t}}, \tag{19}$$

where $\ell\left(y_t|\boldsymbol{\alpha}_t\right)\big|_{\boldsymbol{\alpha}_t=\boldsymbol{a}_{t|t}}$ is the log-likelihood of a measurement given the filtered state estimate and therefore gives a measure of fit of the filtered state. The second term, which is between curly brackets, is the realised Kullback-Leibler (1951) divergence: a penalty term which minimises the distance between predicted and updates state estimates. This regularises the parameters and prevents over-fitting. However, unless the model is linear and Gaussian, this term generally cannot be evaluated in closed form. We follow Lange (2021) again in using a second-order polynomial approximation:

$$\begin{aligned}
\ell\left(\boldsymbol{\alpha}_t|\mathcal{F}_t\right) &\approx \frac{1}{2}\log\det\left\{\boldsymbol{I}_{t|t}/(2\pi)\right\} - \frac{1}{2}\left(\boldsymbol{\alpha}_t - \boldsymbol{a}_{t|t}\right)'\boldsymbol{I}_{t|t}\left(\boldsymbol{\alpha}_t - \boldsymbol{a}_{t|t}\right), \\
\ell\left(\boldsymbol{\alpha}_t|\mathcal{F}_{t-1}\right) &\approx \frac{1}{2}\log\det\left\{\boldsymbol{I}_{t|t-1}/(2\pi)\right\} - \frac{1}{2}\left(\boldsymbol{\alpha}_t - \boldsymbol{a}_{t|t-1}\right)'\boldsymbol{I}_{t|t-1}\left(\boldsymbol{\alpha}_t - \boldsymbol{a}_{t|t-1}\right).
\end{aligned} \tag{20}$$

Substituting these expressions back into Equation 19, summing over all measurements and taking the argmax with respect to the model parameters collected in the vector $\boldsymbol{\theta}$, we get

$$\widehat{\boldsymbol{\theta}} = \arg\max_{\boldsymbol{\theta}} \sum_{t=1}^{T} \left\{\ell\left(y_t|\boldsymbol{a}_{t|t}\right) + \frac{1}{2}\log\det\left(\boldsymbol{I}_{t|t}^{-1}\boldsymbol{I}_{t|t-1}\right) - \frac{1}{2}\left(\boldsymbol{a}_{t|t} - \boldsymbol{a}_{t|t-1}\right)'\boldsymbol{I}_{t|t-1}\left(\boldsymbol{a}_{t|t} - \boldsymbol{a}_{t|t-1}\right)\right\}, \tag{21}$$

which we define as our proposed maximum likelihood estimator and requires the measurement likelihood as in Equation 1 and various outputs from the Bellman filter. Lange (2021) argues that the estimator of Equation 21 is only slightly more computationally demanding than maximum-likelihood estimation of the Kalman filter, the single difference being the number of

required optimisation steps per time step. Whereas the Kalman filter requires only one step, the Bellman filter can require more, though typically not more than five, meaning estimation of the Bellman filter is almost as computationally inexpensive as that of the Kalman filter.

4 Particle Filter

The particle filter we apply is the continuous sampling importance resampling (CSIR) algorithm from Malik and Pitt (2011). Particle filtering is similar to sequential importance sampling as introduced by Hammersley and Morton (1954) and first applied in econometrics by Hendry and Richard (1992), where a latent state is 'guessed' numerous times with 'particles' and estimated a posteriori conditionally on the observations. The difference is a resampling step, which was added by Gordon et al. (1993). This decreases the number of particles required as they are used more efficiently. This is an important advantage, as for importance sampling the number of required particles increases exponentially with the dimension of the state. Under certain conditions, it has been shown that particle filters can beat this curse of dimensionality. The particle filter from Malik and Pitt (2011) is essentially the traditional particle filter from Gordon et al. (1993) where the resampling step is changed to ensure that the approximated likelihood is continuous in the parameters, which allows for gradient-based maximisation of the estimated likelihood.

4.1 Concept

A particle filter estimates the state by the online simulation of numerous possible states called 'particles' and assessing their likelihood based on measurements to approximate the integral

$$p\left(\boldsymbol{\alpha}_{t+1}|\mathcal{F}_{t+1}\right) \propto p\left(y_{t+1}|\boldsymbol{\alpha}_{t+1}\right) \int_D p\left(\boldsymbol{\alpha}_{t+1}|\boldsymbol{\alpha}_t\right) p\left(\boldsymbol{\alpha}_t|\mathcal{F}_t\right) \mathrm{d}\boldsymbol{\alpha}_t \tag{22}$$

where y_t is the one-dimensional measurement, $\mathcal{F}_t$ the filtration of the process up to and including time t and D is the s-dimensional domain of the time t state $\boldsymbol{\alpha}_t = (\lambda_t, \eta_{t+n}, ..., \eta_{t-m})'$ with a straightforward sampling-resampling method. The approximated posterior distribution is the

empirical filtering density, given by

$$\widehat{p}\left(\boldsymbol{\alpha}_{t+1}|\mathcal{F}_{t+1}\right) \propto p\left(y_{t+1}|\boldsymbol{\alpha}_{t+1}\right) \sum_{k=1}^{N} \pi_t^k p\left(\boldsymbol{\alpha}_{t+1}|\boldsymbol{x}_t^k\right), \tag{23}$$

where $\boldsymbol{x}_t^k$ is the kth s-dimensional particle, representing a possible time t state, π_t^k its associated relative probability or 'weight' and N the total number of particles. Smith and Gelfand (1992) prove this estimator to converge to the required density of Equation 22 as $N \to \infty$. This method therefore requires only two assumptions: that the measurement density is computable and that samples can be drawn from the Markovian transition density, which are the particles $\boldsymbol{x}^k$ for $k = 1, ..., N$. Since some particles may be a better estimate of the actual state than others, many particles are required to ensure a reasonable estimate. Particles are assigned weights equal to the relative likelihood of that state estimate being correct given the measurement. The filtered state is estimated a posteriori with a weighted mean of the particles. Since the weights are multiplied with their likelihood and normalised at each time step, certain particles attain a very high weight (i.e. likely state estimates), whereas the weights of most particles become negligible (i.e. unlikely state estimates). To avoid this degeneracy, particles are resampled according to the distribution of their weights, after which their weights are reset to $1/N$.

4.2 Filtering Algorithm

The filtering procedure is depicted in Algorithm 2. At time $t = 0$, we initialise the N particles $\boldsymbol{x}_0^1, ..., \boldsymbol{x}_0^N$, from a diffuse Gaussian distribution $p(\boldsymbol{x}_0) \sim N(\boldsymbol{\mu}_\alpha, 10 \cdot \mathbf{1})$, where $\boldsymbol{\mu}_\alpha$ is the unconditional mean $(\mathbf{1} - \boldsymbol{T})^{-1}\boldsymbol{c}$ and $\mathbf{1}$ the $s \times s$ identity matrix. Subsequently, we assign the particles weights, gathered in the vector $\boldsymbol{\pi}_0$ with length N, with each element equal to $1/N$. After initialisation, we predict the particles for the next time step, $\widetilde{\boldsymbol{x}}_1^1, ..., \widetilde{\boldsymbol{x}}_1^N$ by sampling from their corresponding distributions $p(\boldsymbol{x}_1|\boldsymbol{x}_0^1), ..., p(\boldsymbol{x}_1|\boldsymbol{x}_0^N)$. Following the prediction step comes the correction step, in which we scale the weights with the likelihood of the measurement given the particles, such that $\boldsymbol{\pi}_1 \propto [p(y_t|\widetilde{\boldsymbol{x}}_1^1), ..., p(y_t|\widetilde{\boldsymbol{x}}_1^N)]' \odot \boldsymbol{\pi}_0$. Subsequently, we estimate the hidden state $\boldsymbol{\alpha}_t$ with $\boldsymbol{a}_t$ being the weighted mean of the particles. Chopin (2004) shows it is important to perform estimation before resampling, as the former is more efficient.

Lastly, we consider resampling the particles from their weights distribution, to remove unlikely

state estimates. We do not resample at every time step, which is a departure from Malik and Pitt (2011), because the sorting required for the resampling step is of operation order $O(N \log N)$ and thus computationally intensive and introduces unnecessary extra Monte Carlo variation (Durbin & Koopman, 2012). Therefore, resampling is triggered when the estimated degeneracy of the particles exceeds a certain threshold, meaning that most of the weight has accumulated to a small selection of particles, effectively rendering the other particles irrelevant. This degeneracy leads to waste of memory and computing power and less accurate results. We measure degeneracy with an estimate of the 'effective sample size' as described by J. S. Liu and Chen (1995):

$$\widehat{N_{eff,t}} = \frac{1}{\boldsymbol{w}_t' \cdot \boldsymbol{w}_t}, \tag{24}$$

where $\boldsymbol{w}_t$ is the time t unnormalised weights vector from the correction step of Algorithm 2. Following the 'rule of thumb', we resample when $\widehat{N_{eff}} < 0.5$. In our application, this occurs every 10 - 50 time steps depending on the parameters, sharply decreasing the computation time.

Algorithm 2: CSIR

for $k = 1, ..., N$ **do**
  Initialise $\boldsymbol{x}_0^k$ from $p(\boldsymbol{x}_0)$ and set $\pi_0^k = \frac{1}{N}$
end
for $t = 1, ..., T$ **do**
  for $k = 1, ..., N$ **do**
    Predict: $\widetilde{\boldsymbol{x}}_t^k \sim p(\boldsymbol{x}_t | \boldsymbol{x}_{t-1}^k)$
  end
  Correct: $\boldsymbol{\pi}_t = (\sum_{i=1}^N w_t^i)^{-1} \cdot \boldsymbol{w}_t$, where $\boldsymbol{w}_t = [p(y_t|\widetilde{\boldsymbol{x}}_t^1), ..., p(y_t|\widetilde{\boldsymbol{x}}_t^N)]' \odot \boldsymbol{\pi}_{t-1}$
  Estimate: $\boldsymbol{a}_t = [\widetilde{\boldsymbol{x}}_t^1, ..., \widetilde{\boldsymbol{x}}_t^N] \cdot \boldsymbol{\pi}_t$
  Resample: $\boldsymbol{x}_t^k$ from approximate empirical cdf $\widetilde{F}(\boldsymbol{x})$
end

4.3 Continuous Resampling

Instead of resampling the particles from their weighted empirical cumulative distribution function (cdf) $\widehat{F}_N(x) = \sum_{k=1}^{N} \pi^k H\left(x - \widetilde{x}^k\right)$ with $H(\cdot)$ being the unit or Heaviside step function, we sample from its continuous approximation. In a univariate state setting, Malik and Pitt (2011) propose this continuous approximate empirical cdf to be

$$\widetilde{F}_N(x) = \sum_{k=0}^{N} \lambda^k G_k \left(\frac{x - x^{(k)}}{x^{(k+1)} - x^{(k)}} \right), \tag{25}$$

where $G_k(z)$ is a non-decreasing distribution function on the interval $[0, 1]$ for $k = 1, ..., N-1$ and equal to 0 or 1 for $k = 0$ and $k = N$ respectively. We follow Malik and Pitt (2011) in their choice of the uniform distribution function such that $G(z) = z$ for straightforward calculations. λ^k is the interpolated particle weight, with $\lambda^k = (\pi^{k+1} + \pi^k)/2$, $\lambda^0 = \pi^1/2$ and $\lambda^N = \pi^N/2$. Effectively, this function interpolates the empirical cumulative distribution function, passing through all the midpoints between steps. Malik and Pitt (2011) extend the asymptotic convergence result of Del Moral (2004) by proving that as $N \to \infty$, $\widetilde{F}_N(x) \to \widehat{F}_N(x) \to F_N(x)$.

To make this originally univariate resampling procedure applicable to our model, we apply the multivariate extension using conditionally independent dimensions suggested by Malik and Pitt (2011), whom we follow in this section. First, we construct the weighted empirical cdf of the first state dimension, which we call x_1, and split into P quantiles, which we call partitions R_p for $p = 1, ..., P$. The number of particles in every partition is then equal to $N_p = N/P$, such that $R_p = \{\widetilde{x}_{1,p}^1, ..., \widetilde{x}_{1,p}^{N_p}\}$. We renormalise the weights of these particles by $\widetilde{\pi}_p^j = P \cdot \pi_p^j$ with $j = 2, .., N_p - 1$. To ensure that $\sum_{j=1}^{N_p} \pi_p^j = 1$, the weights of the endpoints $\widetilde{\pi}_p^1$ and $\widetilde{\pi}_p^{N_p}$ are adjusted, such that

$$\widetilde{\pi}_p^1 = P \cdot \left(\widehat{F}_N\left(x_{1,p}^1\right) - (p-1)/P \right), \qquad \widetilde{\pi}_p^{N_p} = P \cdot \left(p/P - \widehat{F}_N\left(x_{1,p}^{N_p - 1}\right) \right). \tag{26}$$

Second, we sample N_p times from every partition p for every dimension $1, ..., s$ using the continuous approximate cdf of Equation 25, which we apply individually per dimension and partition. This yields independent samples $\{x_{1,p}^j, ..., x_{s,p}^j\}$ with $j = 1, ..., N_p$ and $p = 1, ..., P$. Third, we

randomly link samples, one from every dimension, within one partition to form multivariate samples $\boldsymbol{x}_p^j$. Here, we make the assumption that within a partition, the state dimensions are independent. According to Malik and Pitt (2011), this assumption becomes less troublesome as the number of partitions increases, but the continuous approximation will be less accurate, as this is only based on the particles in the partition. Conversely, having a small number of partitions will ensure continuity will be accurate, but the independence assumption can become straining. However, since in our case the state dimensions are almost independent in the state transition equation, we choose P to be 10, such that every partition contains 500 particles. We believe this to be a good compromise between a reasonably accurate continuous approximation, while keeping the independence assumption acceptable, but this has to be investigated further.

4.4 Parameter Estimation

The log-likelihood given the model parameters $\boldsymbol{\theta}$ is estimated by

$$\log \widehat{L}_N(\boldsymbol{\theta}) = \sum_{t=1}^{T} \log \widehat{p}\left(y_t|\boldsymbol{\theta}; \mathcal{F}_{t-1}\right) = \sum_{t=1}^{T} \log\left(\frac{1}{N}\sum_{k=1}^{N} w_t^k\right), \tag{27}$$

where w_t^k is the unnormalised weight of the particle $\widetilde{\boldsymbol{x}}_t^k$, computed in the correction step of Algorithm 2. The maximum likelihood parameters are then defined as:

$$\widehat{\boldsymbol{\theta}} = \arg\max_{\boldsymbol{\theta}} \sum_{t=1}^{T} \log\left(\frac{1}{N}\sum_{k=1}^{N} w_t^k\right). \tag{28}$$

This optimisation is executed well given good starting values. However, without this initial point, the optimiser can get stranded in local maxima, which deteriorates the practicality of this method.

5 Simulation Studies

We investigate simulation results of the two estimators in finite samples by comparing different settings for the parameter vector $\boldsymbol{\theta} = (\mu, c, \phi, \sigma_\eta, \boldsymbol{\rho}')'$ where $\boldsymbol{\rho} = (\rho_n, ..., \rho_{-m})'$ with their estimated counterparts. Table 2 shows the different scenarios we consider as data generating processes for the simulation studies.

The values of μ, c, ϕ and σ_η are from Harvey and Shephard (1996) and the specifications of scenarios 2 and 3 are taken from Catania (2020), who follows Harvey and Shephard (1996) in the same fashion. Note that scenario 1 considers the 'classic' intertemporal specification for the stochastic volatility model with leverage of Harvey and Shephard (1996), whereas scenario 2 considers the contemporaneous specification of Jacquier et al. (2004). Scenarios 4 and 5 are newly considered, as they involve volatility leads as well as lags.

5.1 Parameter Estimates

For every scenario, we estimate our model one hundred times using both estimation methods on generated samples of length $T = 5000$, which Catania (2020) finds to generally produce good results for estimates from his implementations of the particle filter of Malik and Pitt (2011) and the quasi-maximum likelihood estimator of Harvey and Shephard (1996). The number of particles is set to $N = 5000$, being a practical constraint, as more particles produced infeasible computation times. However, simulations show this number to provide a filtering accuracy comparable to that of the Bellman filter.

Table 2: Scenarios with their associated coefficient values.

Scenario	1	2	3	4	5
μ	0	0	0	0	0
c	0	0	0	0	0
ϕ	0.975	0.975	0.975	0.975	0.975
σ_η	0.1	0.1	0.1	0.1	0.1
ρ_2	0	0	−0.3	−0.3	−0.3
ρ_1	−0.5	0	−0.5	−0.5	−0.5
ρ_0	0	−0.8	−0.8	−0.7	−0.7
ρ_{-1}	0	0	0	−0.2	−0.2
ρ_{-2}	0	0	0	0	−0.1

Starting values for the maximum likelihood optimisation for the Bellman filter are found by first optimising $(\widehat{\mu}, \widehat{c}, \widehat{\phi}, \widehat{\sigma_\eta})$ keeping $\widehat{\boldsymbol{\rho}} = \mathbf{0}$, thereby not allowing any correlation between shocks. In practice, this produces a good initial point and these values are changed only slightly in the final optimisation. After this, we perform a grid search over $\boldsymbol{\rho}$ to form a complete set of starting values for the final unconstrained optimisation. Parameter estimation for the particle filter is more difficult as the optimiser tends to get stuck in local maxima of $\boldsymbol{\theta}$ despite the added continuous resampling step. This may be caused by an insufficient number of particles due to the state dimension being too large or the degeneracy in the state transition equation, which can cause particle filters to struggle (Künsch, 2013; J. Liu & West, 2001). Therefore, we resort to using the 'optimal' parameters from the Bellman filter as starting values, which in practice we find to work well but is a drawback of the particle filter. Furthermore, this vastly decreases the required estimation time and makes this method feasible, as general starting values are not guaranteed to lead to good estimates and require a lot of computing capacity.

Average estimate biases of the one hundred replicates with their bootstrapped standard errors are depicted in Table 3 and 4 for the Bellman and particle filter respectively. At the bottom, the mean absolute error (MAE) of the filtered state innovations and log-variances is reported, which is a measure of the quality of the filtration. MAEs are calculated as

$$\text{MAE } \eta_{t|t} = \frac{1}{R \cdot T} \sum_{i=1}^{R} \sum_{t=1}^{T} |\eta_{t|t}^{(i)} - \eta_t^{(i)}|, \qquad \text{MAE } \lambda_{t|t} = \frac{1}{R \cdot T} \sum_{i=1}^{R} \sum_{t=1}^{T} |\lambda_{t|t}^{(i)} - \lambda_t^{(i)}|, \tag{29}$$

where R is the number of replications, T is the length of each sample, $\eta_t^{(i)}$ denotes the time t log-variance shock of replication i, $\eta_{t|t}^{(i)}$ its associated filtered estimate and ditto for the log-variances $\lambda_t^{(i)}$ and $\lambda_{t|t}^{(i)}$. The 'naive' estimate being the unconditional mean can be used as a benchmark, which yields an MAE of $\sqrt{2/\pi} \cdot \sigma \approx 0.798 \cdot \sigma$ with σ being the standard deviation of the estimated variable.

From Table 3, it can be seen that the maximum-likelihood optimiser for the Bellman filter is able to estimate the coefficients remarkably well under correct model-specification, with the exception of $\widehat{\mu}$ which seems positively biased. This can be due to the interplay between the non-zero mean introduced by skewness in the returns through ρ_0 and μ, making these hard

to estimate. Standard errors are increasing with the number of parameters in the model, which is illustrated by the relatively high average standard errors of scenario 5 compared to the relatively low average standard errors of scenarios 1 and 2. Also, standard errors for the parameter estimates in $\widehat{\boldsymbol{\rho}}$ are generally larger than those in $(\widehat{\mu}, \widehat{c}, \widehat{\phi}, \widehat{\sigma_\eta})$, suggesting that the correlation parameters are more difficult to estimate. These effects together yield the highest standard error: $\widehat{\rho_{-1}}$ in scenario 5 is equal to -0.267 with a standard error of 0.150, which shows that estimates of this parameter are quite volatile. MAEs of the filtered log-variance shocks $\eta_{t|t}$ are on average found at around 0.569, which indicates that these shocks are hard to filter, seen as their standard deviation is set at 1. However, this is lower than 0.798, showing out-performance of the naive estimate. The average MAE of $\lambda_{t|t}$ is 0.137, which is smaller than $0.798 \cdot \sigma_\lambda = 0.359$ with σ_λ being the unconditional log-variance standard deviation of

Table 3: Average coefficient estimate biases with bootstrapped standard errors and MAEs of the filtered state shocks and states for the five scenarios using the Bellman filter.

	Bellman filter				
Scenario	1	2	3	4	5
$\widehat{\mu}$	0.001 (0.014)	0.041 (0.016)	0.035 (0.026)	0.057 (0.037)	0.080 (0.077)
$\widehat{c}$	0.001 (0.002)	−0.003 (0.002)	−0.002 (0.002)	−0.003 (0.002)	−0.005 (0.004)
$\widehat{\phi}$	−0.002 (0.005)	−0.001 (0.004)	0.001 (0.002)	0.002 (0.003)	−0.003 (0.015)
$\widehat{\sigma_\eta}$	0.003 (0.010)	0.008 (0.004)	0.001 (0.007)	−0.001 (0.008)	−0.004 (0.015)
$\widehat{\rho_2}$	x	x	0.000 (0.031)	0.036 (0.054)	0.073 (0.092)
$\widehat{\rho_1}$	0.045 (0.065)	x	0.001 (0.013)	0.046 (0.058)	0.072 (0.132)
$\widehat{\rho_0}$	x	0.006 (0.069)	0.000 (0.050)	0.048 (0.084)	−0.038 (0.099)
$\widehat{\rho_{-1}}$	x	x	x	−0.043 (0.090)	−0.067 (0.150)
$\widehat{\rho_{-2}}$	x	x	x	x	−0.052 (0.114)
MAE $\eta_{t\|t}$	0.695	0.489	0.470	0.583	0.606
MAE $\lambda_{t\|t}$	0.228	0.201	0.057	0.089	0.112

Equation 2. For both MAEs, it seems that the number of parameters increases the value of the loss function, with the exception of scenario 1, which yields the highest values. This could be caused by a relatively low amount of information in the returns on the state shocks in this data generating process. Finally, the mean correlation between true and filtered shocks is equal to 69%, suggesting that our filtering of the process is quite accurate.

Table 4 contains the estimates, standard errors and MAEs for the particle filter simulation results. It can be seen this method also finds coefficient estimates within a reasonable margin of error, although this declines as the number of parameters increases. (c,ϕ,σ_η) are almost always found within one standard error, but also for this filter it seems that $\widehat{\mu}$ is positively biased. Estimates for $\boldsymbol{\rho}$ are quite accurate for scenarios 1 through 3, but deteriorate for four and five, the latter being the most extreme. Here, standard errors indicate that estimates are

Table 4: Average coefficient estimate biases with bootstrapped standard errors and MAEs of the filtered state shocks and states for the five scenarios using the particle filter.

	Particle filter				
Scenario	1	2	3	4	5
$\widehat{\mu}$	0.001 (0.016)	0.041 (0.016)	0.037 (0.027)	0.050 (0.035)	0.073 (0.083)
$\widehat{c}$	0.001 (0.002)	−0.003 (0.002)	−0.002 (0.002)	−0.003 (0.002)	−0.004 (0.005)
$\widehat{\phi}$	−0.038 (0.011)	−0.032 (0.011)	0.003 (0.003)	0.002 (0.004)	0.002 (0.004)
$\widehat{\sigma_\eta}$	−0.012 (0.013)	−0.012 (0.012)	−0.003 (0.008)	−0.003 (0.008)	0.000 (0.012)
$\widehat{\rho_2}$	x	x	−0.03 (0.033)	0.031 (0.061)	0.062 (0.112)
$\widehat{\rho_1}$	0.023 (0.082)	x	0.002 (0.016)	0.031 (0.061)	0.081 (0.147)
$\widehat{\rho_0}$	x	−0.014 (0.075)	0.053 (0.033)	0.066 (0.068)	0.008 (0.085)
$\widehat{\rho_{-1}}$	x	x	x	−0.045 (0.094)	−0.100 (0.160)
$\widehat{\rho_{-2}}$	x	x	x	x	−0.063 (0.136)
MAE $\eta_{t\|t}$	0.770	0.494	0.483	0.589	0.609
MAE $\lambda_{t\|t}$	0.315	0.272	0.091	0.108	0.108

more volatile than for the Bellman filter, although their absolute biases seem to be slightly smaller on average. The same trends are visible in this table as in that of the Bellman filter, leading to the largest standard errors for $\boldsymbol{\rho}$ in the high-dimensional scenarios. MAEs are very comparable, albeit slightly higher on average, to those of the Bellman filter, which shows us that the accuracy of the filtered shocks of these two filters are nearly identical. Average correlation between true and filtered shocks is found to be 72%, a slight increase over the Bellman filter. Furthermore, correlation between the Bellman filtered shocks and the particle filtered shocks is on average equal to 91.7%, indicating that they offer a comparable filtration of the process.

5.2 Model Misspecification

We examine the performance of the model under misspecification for scenarios 1 and 2, the first stipulating $\rho_1 = -0.5$, the second $\rho_0 = -0.8$ and both all other correlation parameters to zero. Our research interest is how the model compensates for its misspecification through its correlation parameter estimates, which is relevant when selecting a model specification for the empirical study. Since the previous section showed the Bellman and particle filter to produce very similar estimates, with the latter not only requiring the Bellman's estimates as starting values, but also a higher computational load, we solely use the Bellman filter for estimation in this section. Apart from this, the procedure is exactly the same as in the previous section.

In Table 5, average coefficient estimates for $\boldsymbol{\rho}$ with MAE values for state shocks are given for

Table 5: Average correlation parameter estimates under model misspecification with associated bootstrapped standard errors and MAEs of the filtered state shocks.

	Scenario 1 ($\rho_1 = -0.5$)			Scenario 2 ($\rho_0 = -0.8$)		
Constraint	$\widehat{\rho_1}$	$\widehat{\rho_0}$	MAE $\eta_{t\|t}$	$\widehat{\rho_1}$	$\widehat{\rho_0}$	MAE $\eta_{t\|t}$
none	-0.472 (0.082)	-0.008 (0.051)	0.697	0.000 (0.017)	-0.818 (0.028)	0.486
$\widehat{\rho_1} = 0$	-	-0.443 (0.089)	0.873	-	-0.794 (0.069)	0.489
$\widehat{\rho_0} = 0$	-0.465 (0.065)	-	0.695	-0.713 (0.059)	-	0.977

scenario 1 and 2, given different constraints. First we observe scenario 1, where we see that the constraint $\rho_0 = 0$ corresponds to the correctly specified model with results from Table 3 and the lowest MAE of 0.695. When unconstraining ρ_0, we find estimates for both parameters with $\widehat{\rho_1}$ slightly closer to its true value of -0.5 and $\widehat{\rho_0}$ of a negligible value. The standard error of ρ_1 widens slightly, which might indicate that the optimiser uses the extra model flexibility to shift weight between the two parameters to best fit every specific replication. However, the letting go of the $\rho_0 = 0$ constraint does not seem to deliver a significantly worse performing model, as coefficient estimates are still accurate within one standard error and MAE has increased only slightly to 0.697. Constraining $\rho_1 = 0$, forces the optimiser to (incorrectly) only use ρ_0 to capture the dependence between state and return shocks. ρ_0 is estimated at -0.443 with a relatively large standard error. It seems to be 'pulled' towards the value of ρ_1, due to the only small intertemporal distance. The average MAE value shoots up to 0.873, indicating that this model not only performs worse than the unconstrained and correctly constrained models, but also the naive benchmark.

Moving on to scenario 2, we again observe the constraint $\rho_1 = 0$ to return the same results as in Table 3. Unconstraining ρ_1 yields an estimate of 0.000 with a small standard error, indicating that the optimiser accurately recognises that the associated correlation is not present in the data. The estimate for ρ_0 increases slightly from -0.794 to -0.818, being somewhat further from the true value of -0.8. Its standard error surprisingly becomes smaller, which might be caused by some variation in ρ_0 being absorbed by the (now free) ρ_1, but this remains unclear. The MAE for the unconstrained scenario is marginally smaller than that of the correct specification, but this seems negligible. Imposing the incorrect constraint of $\rho_0 = 0$ yields an estimate for ρ_1 of -0.713, showing again that the optimiser tries to siphon some of the contemporaneous correlation to the intertemporal parameter ρ_1, making this close to but smaller then the value of ρ_0. Under this constraint, the mean MAE sharply increases to 0.977, which reveals the average absolute error of the filtered state shocks to double under this specification with respect to the others and failing the benchmark.

We conclude from this that adding extra correlation parameters can decrease the in-sample

performance of the model slightly and make coefficient estimates less accurate. However, leaving out correlation parameters that are of importance does so much more extreme. Therefore, we would advise to be hesitant to preferring a smaller model over a larger from a semantic perspective.

6 Empirical Results

6.1 Data

We obtained daily data of adjusted close-to-close returns of seven stock indices: the AEX from 10 Dec 1992 through 30 Dec 2020 (7152 observations), DAX from 3 Jan 1990 through 30 Dec 2020 (7832 observations), FTSE 100 from 3 Jan 1990 through 30 Dec 2020 (7858 observations), Hang Seng index (HSI) from 3 Jan 1990 through 29 Dec 2020 (7648 observations), Nikkei 225 (N225) from 4 Jan 1990 through 30 Dec 2020 (7611 observations), NASDAQ from 3 Jan 1990 through 30 Dec 2020 (7810 observations) and the S&P 500 from 3 Jan 1990 through 30 Dec 2020 (7810 observations) from Yahoo Finance and transform them to daily percentage returns by taking the log, first differences and multiplying by 100%.

These indices are selected as they are from different regions and therefore can reflect different markets. Furthermore, the indices differ in their construction, as the S&P is a broad index of 500 large-cap American companies, whereas the AEX only consists of 25 companies. Lastly, indices differ in the sectors they cover, e.g. the DAX is mostly an industry-driven index, but the NASDAQ, with over 3000 listed securities, almost exclusively consists of tech companies.

Furthermore, we acquired data on realised variance from the Oxford-Man Institute of Quantitative Finance of the same seven stock indices from 3 Jan 2000 up until the end of each respective index' dataset of returns. From this variance dataset, we use the daily realised variance based on 5 minute interval return data as a proxy for variance in the stock index returns. Lastly, we acquired historical data of the Cboe Volatility Index (VIX), which is a measure of expected volatility based on the S&P 500, from 3 Jan 1990 through 30 Dec 2020 (7810 observations) from Yahoo Finance.

6.2 Full-scale Model Estimation

We estimate our model on logarithmic returns of historical closing prices of the stock indices AEX, DAX, FTSE 100, HSI, NASDAQ, N225 and S&P 500. To ward off inaccuracies in parameter estimates and filtered states, we start with estimating the full-scale model considered, which contains five correlation parameters ρ_2 through ρ_{-2} with both the Bellman filter and the particle filter. Since estimates from these two methods are very similar, the estimates from the Bellman filter are depicted in Table 6, whereas the particle filter estimates can be found in Appendix C together with numerically estimated standard errors for the Bellman coefficients. These standard errors could not be estimated for the particle filter, possibly due to insufficient smoothness of the likelihood in the parameters. Statistical insignificance at 5% confidence is calculated by assuming normality and indicated with a star ($\star$) in Table 6.

The results show conventional values for $\widehat{c}$, $\widehat{\phi}$ and $\widehat{\sigma_\eta}$, falling in line with previous research. The estimates for μ are all positive and around 0.1, indicating a positive median daily return of 0.1%. Values for $\widehat{\rho_2}$ and $\widehat{\rho_{-2}}$ show the estimated correlation between volatility shocks and

Table 6: Coefficient estimates from the Bellman filter for historical returns of seven stock indices. A star ($\star$) indicates insignificance at the 5% level, assuming normality.

	AEX	DAX	FTSE 100	HSI	NASDAQ	N225	S&P 500
$\widehat{\mu}$	0.125	0.093	0.051	0.099	0.179	0.064	0.109
$\widehat{c}$	$-0.001^\star$	0.006	$-0.002^\star$	0.011	$0.000^\star$	0.018	-0.008
$\widehat{\phi}$	0.988	0.981	0.986	0.982	0.987	0.967	0.984
$\widehat{\sigma_\eta}$	0.204	0.199	0.171	0.182	0.237	0.236	0.238
$\widehat{\rho_2}$	$0.056^\star$	0.328	0.351	$0.043^\star$	0.124	$-0.054^\star$	$0.008^\star$
$\widehat{\rho_1}$	$-0.085^\star$	-0.595	-0.622	-0.257	$-0.038^\star$	-0.281	-0.115
$\widehat{\rho_0}$	-0.758	-0.262	-0.288	-0.415	-0.759	-0.500	-0.812
$\widehat{\rho_{-1}}$	$0.054^\star$	-0.110	$-0.118^\star$	0.102	$-0.012^\star$	0.291	0.154
$\widehat{\rho_{-2}}$	$-0.053^\star$	$-0.043^\star$	$-0.036^\star$	-0.092	-0.103	-0.102	$-0.010^\star$

return shocks of two days earlier and later respectively. They are incidentally found at notable values, but do not seem to imply a strong relation. Exceptions are the DAX and FTSE 100, where the effect of return shocks on volatility seems to not be fully incorporated within one day. This could contain a volatility overreaction, where the leverage effect of ρ_1 overshoots on the first day, leading to a relation between a negative return of two days ago and a decrease in volatility today.

$\widehat{\rho_1}$, $\widehat{\rho_0}$ and $\widehat{\rho_{-1}}$ tell an interesting story as they could be interpreted as representing three well-known volatility phenomena: leverage, volatility feedback and the volatility risk premium respectively. The estimates for ρ_1 show that in most of the indices the leverage effect seems to be present, as there is a consistent negative relation between returns of yesterday and volatility shocks of today, implying an increase in market distress the day after a negative return. ρ_0 captures the contemporaneous relation between the shocks, also showing negative correlation estimates between the two. This falls in line with the volatility feedback hypothesis which implies that a rise in volatility requires an immediate decline in price to allow for volatility risk premia. Some markets seem to be subjected more to one of these effects than the other as the relative sizes of $\widehat{\rho_1}$ and $\widehat{\rho_0}$ interchange depending on the index. This could suggest that some stock indices incorporate volatility changes faster than others. ρ_{-1} shows the relation between the volatility shock of today and the return of tomorrow. This relation has not been considered before, but can capture self-fulfilling prophecies of nervous investors and volatility risk premia. The results differ per market, but seem economically significant. For some indices, this relation is found to be negative, whereas for others this is positive. A positive relation combined with volatility feedback implies a premium for investors after a dip as volatility rises: a negative return leads to a volatility increase on the same day, which leads to a 'bounce-back' return on the next day. However, a negative relation could mean that declining volatility leads to positive returns in the future. Furthermore, this could reflect delayed volatility feedback, where price decline stemming from a volatility surge is not fully incorporated immediately, but is also visible a day later.

We conclude that the correlation structures suggested by Table 6 are not unambiguous over

different indices. These indices can be treated as reflecting different markets, e.g. the DAX is an industry-heavy index, whereas the NASDAQ almost exclusively considers tech companies. It remains speculative to link economical interpretation to index differences and Table 6, but a possible interpretation is that indices with a strong contemporary correlation show a larger volatility risk premium. On the other hand, indices where 'slower' leverage dominates through ρ_1, also seem to experience a delayed volatility feedback effects through negative ρ_{-1} estimates. It is possible that indices where $\widehat{\rho_0}$ is found to be very large incorporate changes in volatility faster and therefore leave less of the stylistic volatility effects to be absorbed by ρ_1 and ρ_{-1}. Furthermore, the results for these indices are more likely to suggest the existence of a risk premium for volatility through positive $\widehat{\rho_{-1}}$. Table 6 suggests that the 'slower' indices are the DAX and FTSE 100 and 'faster' indices are the AEX, NASDAQ and S&P 500, as the latter show estimated $\widehat{\rho_0}$ of around 80%, whereas the former reach only 27% on average and almost vice versa for the intertemporal ρ_1. Moreover, the ρ_{-1} estimate for the AEX and S&P 500 show positive risk premia for volatility, for the NASDAQ only a negligible amount and for the DAX and FTSE it suggests a delayed volatility feedback effects.

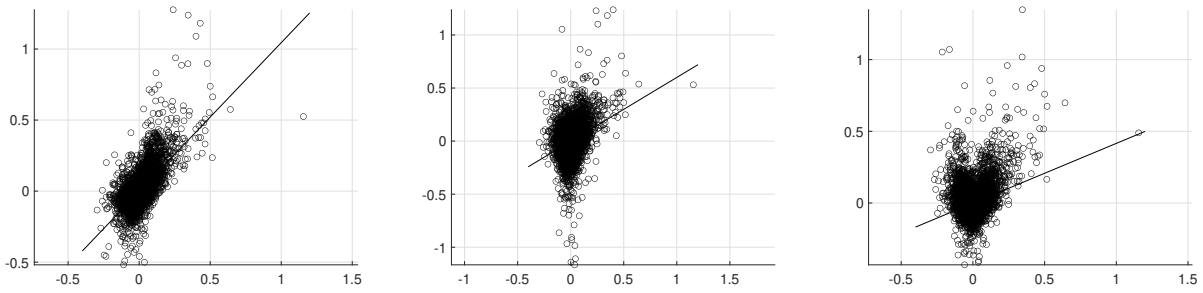
6.2.1 Implied mean, skew and autocorrelation of S&P 500

We estimate the mean and skew of the returns of the S&P 500 numerically through simulation with the estimated model, since this is difficult in closed form as shown in subsection 2.1. The estimate of the mean is found at 0.032%, lower than the median of 0.109%, exactly equal to $\widehat{\mu}$. The skewness of the simulated returns equals -1.186, pointing to a strong negative skew with negative returns being larger in size. This tells us that the non-zero mean, which stems from the skewness induced by ρ_0, is compensated for by $\widehat{\mu}$ yielding a slightly positive mean return. This corresponds to the empirical estimates of the mean, median and skew, which are 0.030, 0.058 and -0.410 respectively. We conclude that the skew is somewhat overestimated, but mean and median are reasonably correct. Furthermore, we calculate the autocorrelation of the return shocks implied by $\widehat{\boldsymbol{\rho}}$ with Equation 4, which yields $\widehat{\boldsymbol{\gamma}}^{\widetilde{\varepsilon}} = (-0.034, -0.016, 0.002, 0.000)'$, sorted from lag one to four. This approximately falls in line with the empirical autocorrelation $\boldsymbol{\gamma}^{\widetilde{\varepsilon}} = (-0.056, -0.058, 0.020, -0.011)'$ and the negative autocorrelation typical for index returns.

6.2.2 Impact of μ

We find that omitting μ, as is done in e.g. Catania (2020), can lead to strongly underestimating ρ_0, as the negative mean implied by the negative skewness cannot be offset by increasing μ. This possibly explains why previous research sometimes finds ρ_0 to be irrelevant. It is our belief that this is incorrect and that this parameter plays an important, in some markets central role in the relation between stock prices and market volatility.

To support this claim, we estimate our full-scale model, filter the volatility shocks with the Bellman filter and calculate the correlation with shocks of the VIX. We compare this result with a model that neglect this parameter by setting $\mu = 0$ and with a 'simple' unleveraged model by setting $\boldsymbol{\rho} = \mathbf{0}$ but μ unconstrained. The latter reaches a correlation of 28.2%; adding the correlation parameters but constraining $\mu = 0$ improves this to 31.5%. However, adding μ back to obtain our full-scale model yields 66.3%, a strong improvement. Furthermore, the scatter plots in Figure 1 illustrate this difference, as the points and ordinary least squares lines line up straighter in Figure 1a than in 1b and 1c. This leads us the believe that these parameters are of vital importance when incorporating correlation between return and volatility shocks.



(a) Full-scale model (b) $\mu = 0$ (c) $\boldsymbol{\rho} = \mathbf{0}$

Figure 1: Three scatterplots with ordinary least squares lines of the VIX shocks on the x-axis and the filtered volatility shocks on the y-axis under three different model specifications.

6.3 Model Selection

We find an indication of parameters' relevancy by calculating the Akaike information criteria for different model sizes and depict them in Table 7.

From Table 7, it can be seen that the information criteria mostly point to models with one volatility lag and either one or two leads. This roughly corresponds to the results of Table 6, as they show the three 'central' correlation parameters to be of the most importance. Moreover, no index yields the lowest information score with the model without leading volatility. Lastly, there does not seem to be a structural difference between Bellman and particle filter selection.

For the AEX, both methods point to a model with one lag and lead, which is striking since all correlation parameters except the contemporaneous were found to be insignificant. DAX and FTSE results are similar, both once indicating $n = 1, m = 1$ and once $n = 1, m = 2$. Surprising, as in Table 6 they show significant values for $\widehat{\rho_2}$, but not for $\widehat{\rho_{-2}}$. It could be that, although these leading parameters are not found at notable values, their likelihood gains offset their penalties in the AIC, whereas the lagging ones do not, but this is speculative. The criteria for both the HSI and the NASDAQ reflect their significant 'slow' correlation parameters $\widehat{\rho_2}$ and $\widehat{\rho_{-2}}$ as they point to large models. Lastly, the N225 and S&P 500 waddle between the simple $n = 1, m = 1$ model and adding one lagging or leading parameter respectively. This falls in line with their strong central correlation parameters of Table 6.

Table 7: Akaike information criterion values estimated using the Bellman and particle filter. A star ($\star$) indicates the lowest values for method-index pairs.

	AEX	DAX	FTSE 100	HSI	NASDAQ	N225	S&P 500
$n = 2, m = 2$				P	B		
$n = 2, m = 1$						B	
$n = 1, m = 2$		P	B	B	P		P
$n = 1, m = 1$	B&P	B	P			P	B
$n = 2, m = 0$							

We conclude that differences found between full-scale model parameter estimates found in Table 6 may translate into different model selections, again reflecting our suspicion that the relation between volatility and returns differs between indices.

6.4 Predictive Performance

6.4.1 Loss functions

In Table 8, we compare out-of-sample volatility forecasts from 4 Jan 2010 through 30 Dec 2020 of our new stochastic volatility model (SV1), which is specified corresponding to the results of Table 7, with a more traditional volatility model (SV2: $m = 0, n = 2$) in the style of Catania (2020) as it allows for contemporaneous and lagging, but not leading volatility correlation. Furthermore, we compare it with the performance of a completely 'unleveraged' stochastic volatility model (SV3: $m = 0, n = 0, \rho_0 = 0$), where return and state shocks are independent. To round out our comparison, we also analyse the performance of the asymmetric or 'leveraged' GARCH(1,1) model of Glosten et al. (1993), which is given by

$$\begin{aligned} y_t &= \mu + \sigma_t \cdot \varepsilon_t, \\ \sigma_{t+1}^2 &= \omega + (\alpha + \beta H(y_t - \mu))\,(y_t - \mu)^2 + \gamma \sigma_t^2. \end{aligned} \tag{30}$$

where $H(\cdot)$ is the Heaviside step function to differentiate between positive and negative returns. The future variance σ_{t+1}^2 is a deterministic function of the current log-variance λ_t and return y_t, setting this model apart from stochastic volatility models.

For a quantitative comparison, we apply the robust loss functions from Patton (2011), which he shows to work well for measuring predictive performance of volatility with imperfect proxies. These loss functions are the mean squared error (MSE) and quasi-likelihood (QLIKE) given respectively as

$$\begin{aligned} \text{MSE} &= \left(\widehat{\sigma}_t^2 - \exp \lambda_{t|t-1}\right)^2, \\ \text{QLIKE} &= \lambda_{t|t-1} + \frac{\widehat{\sigma}_t^2}{\exp \lambda_{t|t-1}}, \end{aligned} \tag{31}$$

where $\widehat{\sigma}_t$ is the proxy for the true variance at time t, in our case being the realised variance, and $\lambda_{t|t-1}$ the prediction of the log-variance. Since the MSE is sensitive to outliers, results can

become distorted. For this reason, we remove squared errors exceeding a value of one thousand, which typically happens thrice: once on 24 Aug 2015 during a flash crash, and twice in March 2020 during the COVID-19 outbreak. Effectively, we treat these events as being completely unpredictable.

From Table 8, it can be seen that the predictions of the particle filter are almost always out-

Table 8: Mean squared error (MSE) and quasi-likelihood (QLIKE) values for the Bellman and particle filtered log-variance λ_t and the realised variance. SV1 is our new model specified according to Table 7, SV2 the model without leading volatility correlation, SV3 the unleveraged model and asymmetric GARCH(1,1). A star ($\star$) indicates the lowest value for each method-index pair.

		MSE				QLIKE			
Index	Filter	SV1	SV2	SV3	GARCH	SV1	SV2	SV3	GARCH
AEX	Bellman	2.271*	2.283	2.585	2.436	0.865	0.863*	0.859	1.004
	Particle	2.535*	2.623	4.138		1.273*	1.338	4.668	
DAX	Bellman	3.141	3.122*	3.682	3.668	1.150	1.134*	1.146	1.256
	Particle	4.253	4.179*	6.002		1.637	1.503*	4.652	
FTSE 100	Bellman	1.840*	1.927	2.147	1.962	0.594*	0.602	0.613	0.756
	Particle	1.951	1.943*	2.467		0.620*	0.638	2.083	
HSI	Bellman	2.518*	2.520	2.393	2.184	1.291	1.288	1.225*	1.212
	Particle	3.235*	3.354	4.011		2.605*	2.644	6.290	
NASDAQ	Bellman	2.716	2.636*	3.230	2.973	0.917	0.912*	0.935	1.002
	Particle	3.917	3.457*	5.601		1.205	1.190*	3.117	
N225	Bellman	4.283*	4.284	4.699	4.736	1.177*	1.177	1.207	1.309
	Particle	5.539	5.191*	7.176		1.699	1.621*	3.586	
S&P 500	Bellman	2.230*	2.362	3.004	4.138	0.436*	0.439	0.535	0.779
	Particle	2.559*	2.742	3.586		0.462	0.443*	0.812	

performed by the Bellman filter. In terms of MSE, the best value using the full-scale model SV1 is attained eight times and using SV2 six times. In terms of QLIKE, this shift to six versus seven, with one point for SV3 on the HSI. These results indicate that superior predictive performance can be attained by adding volatility leads on certain indices. However, it should be noted that for the HSI, asymmetric GARCH seems unbeatable, yielding the lowest losses. For this instance, we suspect the SV models to lack robustness, possibly due to regime shifts, a strong point of (G)ARCH-type models. Nonetheless, it is interesting to note that asymmetric GARCH, being a model with fewer parameters than the stochastic volatility models, should be more robust out-of-sample, but typically does not outperform a leveraged SV model.

6.4.2 Diebold-Mariano-West test

To measure statistical significance of differences in predictive power between the models while accounting for serial correlation in the loss differential, we use the test for comparing predictive accuracy by Diebold and Mariano (1995) and West (1996). The test statistic is given by

$$\frac{\bar{d}}{\sqrt{\widehat{\sigma}_d^2/T}} \sim \mathrm{N}(0,1), \tag{32}$$

where $\bar{d}$ is the average difference between the squared loss of two predictions e_t^2 and r_t^2:

$$\bar{d} = \frac{1}{T}\sum_{t=1}^{T}\left(e_t^2 - r_t^2\right), \tag{33}$$

and $\widehat{\sigma}_d$ the autocorrelation corrected volatility of $\bar{d}$

$$\widehat{\sigma}_d^2 = \widehat{\gamma}^d(0) + 2\sum_{j=1}^{h}\widehat{\gamma}^d(j), \tag{34}$$

where $\widehat{\gamma}^d(j)$ is the empirical estimate of the jth autocovariance of $\bar{d}$. We found significant autocorrelation in $\bar{d}$ up to four days, which is why we choose $h = 4$.

Resulting p-values for the Diebold-Mariano-West tests are given in Table 9 for the Bellman and particle filter predictions. For the Bellman filter, we see that on the S&P 500, SV1 outperforms the other models at the 5% significance level. On the FTSE 100, we also see significant performance increases, but GARCH catches up. On the other indices, results differ. On the AEX

Table 9: p-values from the Diebold-Mariano-West tests for the Bellman and particle filter predictions for volatility with models SV1 being the AIC selected model of Table 7, SV2 the model without leading volatility correlation, SV3 the unleveraged model, and asymmetric GARCH(1,1). Values larger than 0.5 indicate that the model the to left outperforms the model above and lower vice versa.

		Bellman filter			Particle filter		
		SV2	SV3	GARCH	SV2	SV3	GARCH
AEX	SV1	0.847	0.954	0.358	0.582	1.000	0.127
	SV2		0.960	0.352		1.000	0.123
	SV3			0.254			0.017
DAX	SV1	0.177	0.985	0.969	0.093	0.996	0.084
	SV2		0.991	0.965		0.997	0.107
	SV3			0.497			0.000
FTSE 100	SV1	0.951	0.938	0.891	0.299	0.163	0.696
	SV2		0.864	0.861		0.167	0.726
	SV3			0.805			0.846
HSI	SV1	0.319	0.009	0.000	0.692	0.960	0.000
	SV2		0.007	0.000		0.894	0.000
	SV3			0.000			0.000
NASDAQ	SV1	0.020	0.980	0.230	0.826	0.668	0.139
	SV2		0.983	0.307		0.639	0.138
	SV3			0.099			0.137
N225	SV1	0.689	0.976	0.997	0.239	0.975	0.989
	SV2		0.976	0.997		0.998	0.994
	SV3			0.998			0.859
S&P 500	SV1	0.982	0.998	0.972	0.904	1.000	0.993
	SV2		0.999	0.971		0.999	0.993
	SV3			0.941			0.939

and N225, we still find SV1 to perform the best among the SV models, but not significantly so. On the DAX, HSI and NASDAQ, SV2 defeats SV1, on the latter at 5% significance. GARCH dominates on the HSI and to a lesser extent on the NASDAQ.

Moving to the particle filter results, we see that the gap between SV1 and SV2 prediction performances narrows, indicating that particle filter predictive performance benefits less from adding leading volatility correlation. However, for the S&P 500 we see that SV3 is clearly defeated by the other SV models, as opposed to for the FTSE 100. Lastly, we notice that here the GARCH filter is on average more likely to outperform SV models, as p-values fall with the exception of the S&P 500.

7 Conclusion

Our new stochastic volatility model extends classic volatility models by accounting for both leading and lagging correlation between return and volatility shocks. Furthermore, we add a median parameter to the measurement equation, to be able to independently set the skewness and the mean of the returns as well as effectively unconstraining the contemporaneous correlation parameter ρ_0. These changes allow the model to more fully capture the relation between price fluctuations and their volatility, as not only the classic 'leverage' effect can be captured, where volatility shocks follow return shocks, but also 'volatility feedback', where a sudden increase in volatility requires a decline in price to facilitate future volatility risk premia. Lastly, leading volatility correlation parameters can capture this premium as they show the generally positive relation between a rise in volatility yesterday and the return shocks of today.

We investigated the performance of this model using the recently proposed Bellman filter by Lange (2021) and the particle filter of Malik and Pitt (2011). Generally, we found these filters to give similar results, but estimating and filtering with the Bellman filter is notably quicker compared to the particle filter. Moreover, the particle filter requires good starting values, which we produced with the Bellman filter. Without these starting points, the optimiser got stuck in local maxima of the particle filter's estimated likelihood function.

In our empirical study of the returns of seven stock indices from 1990 through 2020, we found that the added parameters are a relevant addition to the stochastic volatility model. Generally, the median is found at a positive value, the distribution has a notable negative skew and the typical negative autocorrelation in returns is captured. Furthermore, we calculated correlation of our filtered volatility shocks of the S&P 500 with daily changes in the VIX and found that adding both parameters more than doubled correlation percentage with respect to leaving either one out.

Lastly, we compared predictive performance of our model through robust loss functions and Diebold-Mariano-West tests with a stochastic volatility model without leading volatility correlation in the style of Catania (2020), an 'unleveraged' model and the asymmetric GARCH model of Glosten et al. (1993) and conclude that out-of-sample predictions of our new model can significantly outperform other stochastic volatility models and asymmetric GARCH on the S&P 500 using the Bellman filter, and to a lesser extent on the FTSE 100. On the five other indices, our new model outperformed the model without leads twice, but not significantly so. We suspect these differences to be due to the different behaviour of volatility on these indices, as implied by the correlation estimates of Table 6. However, from a semantic perspective, our model gives, to the best of our knowledge, insight into the evolution and interplay of returns and volatility that is new to the field of stochastic volatility models.

From this point, there are multiple paths for future research. For example, it would be interesting to compare the performance of the model of Catania (2020) with that of our model. For this, the Bellman filter implementation needs to be changed slightly, as the state transition equation of this model is degenerate and non-linear. Furthermore, it is appealing to study the implementation of non-Gaussian distributions in our model, a good example being a Student's t-distribution for the returns, as financial returns are generally heavy-tailed. Lastly, an implementation on high-frequency data could yield interesting results, as this allows the researcher to 'zoom in' on the intertemporal relation between return and volatility.

References

Bekaert, G., & Wu, G. (2000). Asymmetric volatility and risk in equity markets. *The Review of Financial Studies*, *13*(1), 1–42.

Bellman, R. E. (1957). *Dynamic programming*. Princeton University Press.

Black, F. (1976). Studies of stock market volatility changes. *Proceedings of the 1976 Meetings of the American Statistical Association, Business and Economic Statistics Section*, 177–181.

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

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

Campbell, J. Y., Grossman, S. J., & Wang, J. (1993). Trading volume and serial correlation in stock returns. *The Quarterly Journal of Economics*, *108*(4), 905–939.

Catania, L. (2020). A stochastic volatility model with a general leverage specification. *Journal of Business & Economic Statistics*. Advance online publication.

Chopin, N. (2004). Central limit theorem for sequential Monte Carlo methods and its application to Bayesian inference. *Annals of Statistics*, *32*(6), 2385–2411.

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.

Crisan, D., & Doucet, A. (2002). A survey of convergence results on particle filtering methods for practitioners. *IEEE Transactions on Signal Processing*, *50*(3), 736–746.

Danielsson, J. (1994). Stochastic volatility in asset prices estimation with simulated maximum likelihood. *Journal of Econometrics*, *64*(1-2), 375–400.

Daum, F. (2005). Nonlinear filters: Beyond the Kalman filter. *IEEE Aerospace and Electronic Systems Magazine*, *20*(8), 57–69.

Del Moral, P. (2004). *Feynman-Kac formulae: Genealogical and interacting particle systems with applications*. Springer.

Diebold, F. X., & Mariano, R. S. (1995). Comparing predictive accuracy. *Journal of Business & Economic Statistics*, *13*(3), 253–263.

Durbin, J., & 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*, *50*(4), 987–1007.

Fridman, M., & Harris, L. (1998). A maximum likelihood approach for non-Gaussian stochastic volatility models. *Journal of Business & Economic Statistics*, *16*(3), 284–291.

Geweke, J. (1989). Bayesian inference in econometric models using Monte Carlo integration. *Econometrica*, *57*(6), 1317–1339.

Glosten, L. R., Jagannathan, R., & 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.

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

Hammersley, J. M., & Morton, K. W. (1954). Poor man's Monte Carlo. *Journal of the Royal Statistical Society: Series B (Methodological)*, *16*(1), 23–38.

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

Hendry, D. F., & Richard, J.-F. (1992). Likelihood evaluation for dynamic latent variables models. In H. M. Amman, D. A. Belsley, & L. F. Pau (Eds.), *Computational economics and econometrics* (pp. 3–17). Springer.

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

James, W., & Stein, C. (1992). Estimation with quadratic loss. In S. Kotz & N. L. Johnson (Eds.), *Breakthroughs in statistics* (pp. 443–460). Springer.

Kalman, R. E. (1960). A new approach to linear filtering and prediction problems. *Journal of Basic Engineering*, *82*(1), 35–45.

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

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

Künsch, H. R. (2013). Particle filters. *Bernoulli*, *19*(4), 1391–1403.

Lange, R.-J. (2021). *Bellman filtering for state-space models*, Tinbergen Institute Discussion Paper 2020-052/III.

Liu, J., & West, M. (2001). Combined parameter and state estimation in simulation-based filtering. In A. Doucet, N. De Freitas, & N. Gordon (Eds.), *Sequential Monte Carlo methods in practice* (pp. 197–223). Springer.

Liu, J. S., & Chen, R. (1995). Blind deconvolution via sequential imputations. *Journal of the American Statistical Association*, *90*(430), 567–576.

Lo, A. W., & MacKinlay, A. C. (1988). Stock market prices do not follow random walks: Evidence from a simple specification test. *The Review of Financial Studies*, *1*(1), 41–66.

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

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

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

Ruiz, E., Peña, D., & Carnero, M. A. (2001). *Is stochastic volatility more flexible than GARCH?* (Tech. rep.). Universidad Carlos III de Madrid. Departamento de Estadística.

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

Taylor, S. J. (1986). *Modelling financial time series*. World Scientific.

West, K. D. (1996). Asymptotic inference about predictive ability. *Econometrica*, *64*(5), 1067–1084.

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

Appendix A State space form general model

Below, the complete general state space formulation of the state transition equation of Equation 1 is given.

$$\begin{pmatrix} \lambda_t \\ \eta_{t+n} \\ \eta_{t+n-1} \\ \vdots \\ \eta_{t+1} \\ \eta_t \\ \eta_{t-1} \\ \vdots \\ \eta_{t-m+1} \\ \eta_{t-m} \end{pmatrix} = \begin{pmatrix} c \\ 0 \\ 0 \\ \vdots \\ 0 \\ 0 \\ 0 \\ \vdots \\ 0 \\ 0 \end{pmatrix} + \begin{pmatrix} \phi & 0 & 0 & & \sigma_\eta & 0 & 0 & & 0 & 0 \\ 0 & 0 & 0 & \dots & 0 & 0 & 0 & \dots & 0 & 0 \\ 0 & 1 & 0 & & 0 & 0 & 0 & & 0 & 0 \\ & \vdots & & \ddots & & \vdots & & \ddots & \vdots & \\ 0 & 0 & 0 & & 0 & 0 & 0 & & 0 & 0 \\ 0 & 0 & 0 & \dots & 1 & 0 & 0 & \dots & 0 & 0 \\ 0 & 0 & 0 & & 0 & 1 & 0 & & 0 & 0 \\ & \vdots & & \ddots & & \vdots & & \ddots & \vdots & \\ 0 & 0 & 0 & & 0 & 0 & 0 & & 0 & 0 \\ 0 & 0 & 0 & \dots & 0 & 0 & 0 & \dots & 1 & 0 \end{pmatrix} \cdot \begin{pmatrix} \lambda_{t-1} \\ \eta_{t+n-1} \\ \eta_{t+n-2} \\ \vdots \\ \eta_t \\ \eta_{t-1} \\ \eta_{t-2} \\ \vdots \\ \eta_{t-m} \\ \eta_{t-m-1} \end{pmatrix} + \begin{pmatrix} 0 \\ \eta_{t+n} \\ 0 \\ \vdots \\ 0 \\ 0 \\ 0 \\ \vdots \\ 0 \\ 0 \end{pmatrix} \tag{35}$$

Appendix B Score and information formulas

In this appendix, we give the formulas for calculating the realised score and information of a measurement y_t with respect to a state estimate $\boldsymbol{a}_t$, as well as their expected counterparts, which can be used for different numerical optimisation methods, such as Fisher's scoring method. Conditional on $\boldsymbol{a}_t = [\lambda_t \ \boldsymbol{\eta}_t]'$, y_t is normally distributed $y_t \sim \mathrm{N}(m_t, s_t^2)$ with

$$\begin{aligned} m_t &= \mathbb{E}\left[y_t|\lambda_t, \boldsymbol{\eta_t}\right] = \mu + \boldsymbol{\rho}'\boldsymbol{\eta_t} \cdot \exp(\lambda_t/2), \\ s_t^2 &= \mathbb{V}\left(y_t|\lambda_t, \boldsymbol{\eta_t}\right) = \left(1 - \boldsymbol{\rho}'\boldsymbol{\rho}\right) \cdot \exp(\lambda_t). \end{aligned} \tag{36}$$

Using this notation, the realised score, which is defined as the first derivative of the measurement log-likelihood conditional on the state $\ell(y_t|\boldsymbol{\alpha}_t)$ with respect to the state itself, is

$$\frac{\mathrm{d}\ell(y_t|\boldsymbol{\alpha}_t)}{\mathrm{d}\boldsymbol{\alpha}_t} = \begin{pmatrix} \frac{\mathrm{d}\ell(y_t|\lambda_t, \boldsymbol{\eta_t})}{\mathrm{d}\lambda_t} \\ \frac{\mathrm{d}\ell(y_t|\lambda_t, \boldsymbol{\eta_t})}{\mathrm{d}\boldsymbol{\eta_t}} \end{pmatrix} = \begin{pmatrix} \frac{y_t(y_t - m_t)}{2s_t^2} - \frac{1}{2} \\ \frac{\boldsymbol{\rho}}{\sqrt{1-\boldsymbol{\rho}'\boldsymbol{\rho}}} \cdot \frac{y_t - m_t}{s_t} \end{pmatrix}, \tag{37}$$

which in expectation is

$$\mathbb{E}\left[\frac{\mathrm{d}\ell(y_t|\boldsymbol{\alpha}_t)}{\mathrm{d}\boldsymbol{\alpha}_t}\right] = \mathbb{E}\left[\begin{pmatrix} \frac{\mathrm{d}\ell(y_t|\lambda_t,\boldsymbol{\eta_t})}{\mathrm{d}\lambda_t} \\ \frac{\mathrm{d}\ell(y_t|\lambda_t,\boldsymbol{\eta_t})}{\mathrm{d}\boldsymbol{\eta_t}} \end{pmatrix}\right] = \begin{pmatrix} 0 \\ 0 \end{pmatrix}. \tag{38}$$

The second derivative with respect to the state is the realised information and given by

$$-\frac{\mathrm{d}^2\ell(y_t|\boldsymbol{\alpha}_t)}{\mathrm{d}\boldsymbol{\alpha}_t\mathrm{d}\boldsymbol{\alpha}_t'} = -\begin{pmatrix} \frac{\mathrm{d}^2\ell(y_t|\lambda_t,\boldsymbol{\eta_t})}{\mathrm{d}\lambda_t^2} & \frac{\mathrm{d}^2\ell(y_t|\lambda_t,\boldsymbol{\eta_t})}{\mathrm{d}\lambda_t\mathrm{d}\boldsymbol{\eta}_t'} \\ \frac{\mathrm{d}^2\ell(y_t|\lambda_t,\boldsymbol{\eta_t})}{\mathrm{d}\lambda_t\mathrm{d}\boldsymbol{\eta}_t} & \frac{\mathrm{d}^2\ell(y_t|\lambda_t,\boldsymbol{\eta_t})}{\mathrm{d}\boldsymbol{\eta_t}\mathrm{d}\boldsymbol{\eta_t}'} \end{pmatrix} = \begin{pmatrix} \frac{y_t^2}{2s_t^2} - \frac{y_t m_t}{4s_t^2} & \frac{\boldsymbol{\rho}'}{\sqrt{1-\boldsymbol{\rho}'\boldsymbol{\rho}}} \cdot \frac{y_t}{2s_t} \\ \frac{\boldsymbol{\rho}}{\sqrt{1-\boldsymbol{\rho}'\boldsymbol{\rho}}} \cdot \frac{y_t}{2s_t} & \frac{\boldsymbol{\rho}\boldsymbol{\rho}'}{1-\boldsymbol{\rho}'\boldsymbol{\rho}} \end{pmatrix}, \tag{39}$$

which in expectation is equal to

$$\mathbb{E}\left[-\frac{\mathrm{d}^2\ell(y_t|\boldsymbol{\alpha}_t)}{\mathrm{d}\boldsymbol{\alpha}_t\mathrm{d}\boldsymbol{\alpha}_t'}\right] = \mathbb{E}\left[-\begin{pmatrix} \frac{\mathrm{d}^2\ell(y_t|\lambda_t,\boldsymbol{\eta_t})}{\mathrm{d}\lambda_t^2} & \frac{\mathrm{d}^2\ell(y_t|\lambda_t,\boldsymbol{\eta_t})}{\mathrm{d}\lambda_t\mathrm{d}\boldsymbol{\eta}_t'} \\ \frac{\mathrm{d}^2\ell(y_t|\lambda_t,\boldsymbol{\eta_t})}{\mathrm{d}\lambda_t\mathrm{d}\boldsymbol{\eta}_t} & \frac{\mathrm{d}^2\ell(y_t|\lambda_t,\boldsymbol{\eta_t})}{\mathrm{d}\boldsymbol{\eta_t}\mathrm{d}\boldsymbol{\eta_t}'} \end{pmatrix}\right] = \begin{pmatrix} \frac{1}{2} + \frac{m_t^2}{4s_t^2} & \frac{\boldsymbol{\rho}'}{\sqrt{1-\boldsymbol{\rho}'\boldsymbol{\rho}}} \cdot \frac{m_t}{2s_t} \\ \frac{\boldsymbol{\rho}}{\sqrt{1-\boldsymbol{\rho}'\boldsymbol{\rho}}} \cdot \frac{m_t}{2s_t} & \frac{\boldsymbol{\rho}\boldsymbol{\rho}'}{1-\boldsymbol{\rho}'\boldsymbol{\rho}} \end{pmatrix}. \tag{40}$$

Appendix C Empirical particle filter coefficient estimates and Bellman standard errors

Table 10 contains the parameter estimates for the full-scale model on daily index returns of the AEX, DAX, FTSE 100, HSI, NASDAQ, N225 and the S&P 500 using the particle filter. They are similar to the estimates produced by using the Bellman filter, as depicted in Table 6.

Table 11 contains the by finite differences numerically approximated standard errors for the Bellman filter parameter estimates of Table 6.

Table 10: Coefficient estimates from the particle filter for historical returns of seven stock indices.

	AEX	DAX	FTSE 100	HSI	NASDAQ	N225	S&P 500
$\widehat{\mu}$	0.133	0.094	0.050	0.100	0.188	0.067	0.100
$\widehat{c}$	−0.001	0.006	−0.003	0.011	0.000	0.018	−0.008
$\widehat{\phi}$	0.942	0.933	0.945	0.947	0.938	0.921	0.965
$\widehat{\sigma_\eta}$	0.195	0.176	0.142	0.159	0.215	0.212	0.256
$\widehat{\rho_2}$	0.058	0.322	0.370	0.044	0.127	−0.055	0.009
$\widehat{\rho_1}$	−0.084	−0.603	−0.623	−0.257	−0.039	−0.287	−0.126
$\widehat{\rho_0}$	−0.787	−0.274	−0.307	−0.413	−0.760	−0.507	−0.890
$\widehat{\rho_{-1}}$	0.055	−0.117	−0.124	0.107	−0.013	0.292	0.157
$\widehat{\rho_{-2}}$	−0.057	−0.046	−0.038	−0.094	−0.109	−0.108	−0.011

Table 11: Coefficients' standard errors from the Bellman filter for historical returns of seven stock indices.

	AEX	DAX	FTSE 100	HSI	NASDAQ	N225	S&P 500
$\widehat{\mu}$	0.016	0.016	0.011	0.020	0.015	0.019	0.011
$\widehat{c}$	0.003	0.003	0.002	0.004	0.003	0.004	0.002
$\widehat{\phi}$	0.128	0.107	0.106	0.101	0.139	0.084	0.115
$\widehat{\sigma_\eta}$	0.074	0.067	0.070	0.082	0.077	0.073	0.070
$\widehat{\rho_2}$	0.049	0.039	0.052	0.047	0.040	0.050	0.047
$\widehat{\rho_1}$	0.067	0.051	0.026	0.050	0.053	0.060	0.038
$\widehat{\rho_0}$	0.080	0.073	0.027	0.081	0.071	0.047	0.079
$\widehat{\rho_{-1}}$	0.068	0.051	0.069	0.041	0.037	0.026	0.039
$\widehat{\rho_{-2}}$	0.047	0.043	0.038	0.035	0.039	0.023	0.053