Draft_Application_Stochastic_Volatility

1 Overview

In the following sections, we will try to replicate some of the analyses which were previously conducted on synthetic data, generated by a basic stochastic volatility model; this time, the same model specification and filters will be used on real data. However, as we will see, real data require different strategies to evaluate the effectiveness of several Sequential Monte Carlo techniques. Intuitively, such differences are due to the fact that now the latent process driving the Hidden Markov Model is not observed, nor do we know exactly its law of motion.

This part of the analysis is structured as follows: section 2 provides a review of recent research in the stochastic volatility literature, including extensions of the basic specification and the use of particle filters; section 3 describes the data which we employ and the transformations that were applied to them. In section 4 we briefly compare the series of filtered states that are obtained from different particle filters; moreover, several PFs are compared in terms of computational cost in approximating the exact filtering distribution implied by our SV model. Section **??** features a forecast comparison between a SV specification and a model that assumes constant mean and variance for the financial returns, namely a forecasting exercise that would not be possible were it not for particle filtering techniques. In section 5 we conclude by employing a non-parametric measure of realised volatility as a proxy for the latent volatility process, allowing a RMSE and MAE comparison between filters which is similar to that of section **??** for synthetic data.

2 A review of Stochastic Volatility models

In section **??** a simple specification of a Stochastic Volatility model was presented and used as the data generating process of a simulated dataset. Denoting by y_t the log returns at time t and by x_t the latent volatility stochastic process, such state-space model was presented as follows,

$$y_t|x_t \sim N(0, e^{x_t})$$
$$x_t|x_{t-1}, \alpha, \beta, \tau^2 \sim N(\alpha + \beta x_{t-1}, \tau^2)$$

This specification - which we will employ in this section on a real dataset - can be deemed as the standard version of the SV model. Throughout the recent decades, several extensions have been proposed, mostly (but not exclusively) in the field of financial econometrics. On the one hand, for simplicity we will stick to the standard version, one that will suffice for the scope of this work on Sequential Monte Carlo methods. On the other hand, it is worth mentioning some of these extensions: indeed, one can rather safely assume that, through their higher degree of sophistication, such extensions might well improve on the performance of our standard SV model in the analyses that will follow.

Stochastic Volatility Models, whose early formulation is commonly attributed to Taylor (1982, 1986), allow to account for time-varying and autoregressive volatility in financial returns, posing themselves as a valid alternative to ARCH (Engle 1982) or GARCH (Bollerslev 1986) models in dealing with non-constant volatility. Kim, Shephard and Chib (1998) define the canonical model for regularly spaced data as

$$y_t = \psi e^{\frac{x_t}{2}} \epsilon_t$$
$$x_{t+1} = \mu + \beta(x_t - \mu) + \tau\eta_t$$
$$x_1|\sigma, \beta \sim N\left(\mu, \frac{\sigma^2}{1-\beta^2}\right)$$
$$\epsilon_t \sim N(0,1)$$
$$\eta_t \sim N(0,1)$$

where the log volatility x_t is assumed to follow a stationary process ($|\beta| < 1$), ψ is a constant scaling factor, β is the persistence in the volatility and τ is the volatility of the log-volatility.[1] Most importantly, ϵ_t and η_t, the Gaussian white noise processes that drive the canonical model, are assumed to be uncorrelated. This latter assumption justifies another definition of this specification, namely "discrete SV model without leverage".[2]

In order to accommodate for leverage effects, namely capture the increase in volatility that follows a drop in the returns, the model can extended as in Omori

---

[1]For identifiability reasons, either $\psi = 1$ or $\mu = 0$. Kim, Shephard and Chib (1998) prefer $\psi = 1$.

[2]Note that in this review we do not focus on continuous time SV models. In fact, such models have attracted a considerable amount of research in financial econometrics and mathematical finance, especially after Hull and White (1987) considered stochastic volatility for option pricing. Arguably, the most influential model was then proposed for option pricing by Heston (1993), a SV model with leverage effects and square root diffusion driving volatility. Diffusion-based SV models enjoyed increasing popularity, see for example Barndorff-Nielsen and Shephard (2001), or Christoffersen et al. (2010), who investigate alternatives to the entrenched affine square root SV model. Eraker (2004) proposed a SV model with correlated jumps in prices and volatility, extending Heston's model, while Comte and Renault (1998) extended Hull and White's model as to feature long memory properties. More recently, Gatheral et al. (2018) further built on Comte and Renault's fractional SV model to propose the popular "rough volatility models" (see also Friz et al., 2021).

et al. (2007),

$$y_t = e^{\frac{x_t}{2}} \epsilon_t$$
$$x_{t+1} = \mu + \beta(x_t - \mu) + \eta_t$$
$$\begin{pmatrix} \epsilon_t \\ \eta_t \end{pmatrix} \Big| \rho, \tau \overset{i.i.d.}{\sim} N_2(\mathbf{0}, \Sigma), \ \Sigma = \begin{pmatrix} 1 & \rho\tau \\ \rho\tau & \tau^2 \end{pmatrix}$$

where $\rho < 0$ captures the negative correlation. Such specification, which can be referred to as "discrete SV model with leverage", captures the asymmetric response of volatility to returns of different signs, so that similar specifications are sometimes also deemed as "asymmetric SV models" (e.g. Harvey and Shephard 1996, Mao et al. 2020).

Note that so far the assumptions on ϵ_t and η_t implied that the returns are conditionally normally distributed. The discrete time SV models can also be extended to allow for heavy-tailed or asymmetric conditional returns distributions: symmetric or skewed Student-t, Generalised Hyperbolic (GH) distribution, Generalised Error Distribution (GED) and scale mixtures of normals feature as popular choices (Kim and Stoffer 2008, Nakajima and Omori 2012, Mao et al. 2020). In fact, SV with heavy tailed return distributions were shown to better meet empirical regularities like the leptokurtic distribution of the returns and slowly decaying autocorrelation functions of the squared returns (Liesenfeld and Jung 2000).

Assuming that ϵ_t follows a Student-t distribution, and exploiting the fact that ϵ_t can then be written as $\lambda_t^{-1/2}\zeta_t$, where $\zeta_t \sim N(0,1)$ and $v\lambda_t \sim \chi_v^2$ (Harvey et al. 1994, Chib et al. 2002), we have the following SV model with both fat tails and leverage effect (Jacquier et al., 2004),

$$y_t = e^{\frac{x_t}{2}} \lambda_t^{-1/2} \zeta_t$$
$$x_{t+1} = \alpha + \beta x_t + \eta_t$$
$$\begin{pmatrix} \zeta_t \\ \eta_t \end{pmatrix} \Big| \rho, \tau \overset{i.i.d.}{\sim} N_2(\mathbf{0}, \Sigma), \ \Sigma = \begin{pmatrix} 1 & \rho\tau \\ \rho\tau & \tau^2 \end{pmatrix}$$
$$v\lambda_t \sim \chi_v^2$$

To capture other elements of the behaviour of financial data, several other extensions have been proposed. For example, SV models have been extended to include conditional heteroskedasticity in the mean returns (Koopman and Hol Uspensky 2002) to capture potential volatility feedback effects, or to feature autoregressive moving average innovations (Chan 2013, Zhang et al. 2020), allowing better goodness of fit and out-of-sample forecasts.[3] [4] The model

---

[3]Dimitrakopoulos and Kolossiatis (2020) note that "the moving average component, the leverage effect and the conditional heteroscedasticity in mean have been considered separately in the stochastic volatility literature" and provide two specifications, one featuring an MA component and leverage effects, the other an MA component and conditional heteroskedasticity in mean.

[4]Other extensions, though ones for which we do not report the specifications, feature

proposed by Koopman and Hol Uspensky is specified as

$$y_t = \nu_t + \psi e^{\frac{x_t}{2}} \epsilon_t$$
$$\nu_t = a + b y_{t-1} + d\psi^2 e^{x_t}$$
$$x_t = \beta x_{t-1} + \tau \eta_t$$
$$\begin{pmatrix} \epsilon_t \\ \eta_t \end{pmatrix} \overset{i.i.d.}{\sim} N_2\left(\mathbf{0}, \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}\right)$$

while the state space representation of the ARMA(p,q)-SV framework, as in Zhang et al. (2020), reads

$$y_t = \nu_t + \gamma_t$$
$$\gamma_t = \phi_1 \gamma_{t-1} + ... + \phi_p \gamma_{t-p} + u_t + \varphi_1 u_{t-1} + ... + \varphi_q u_{t-q}$$
$$u_t | x_t \sim N(0, e^{x_t})$$
$$x_t = x_{t-1} + \eta_t$$
$$\eta_t | \tau \sim N(0, \tau^2)$$

where the error terms u_t and η_t are independent across all leads and lags, while ν_t follows an unspecified time-varying process.

Interestingly enough, the path along which the SV models evolved coincides with that suggested by optimal portfolio findings. Johannes, Korteweg and Polson (2014) found that, in order to generate statistically significant portfolio improvements in a Bayesian learning problem, the model employed by the investor should incorporate both time-varying expected returns and stochastic volatility: indeed, either of these features alone did not lead to statistically significant gains with respect to employing models with time-constant expected returns and volatility.[5]

Finally, before moving to the application of Sequential Monte Carlo techniques to a SV model, we conclude by going through some references for the Bayesian analysis proposed for such models. Starting from the seminal work of Jacquier et al. (1994), the use of MCMC methods has become increasingly popular for parameter estimation and smoothing exercises in SV models (e.g. Kastner 2019, presenting the R package *stochvol* for Bayesian parameter estimation,

---

modelling the latent volatility process x_t as an ARFIMA process (Long Memory Stochastic Volatility model, Breidt et al. 1998) or as governed by a first-order Markov process (Markov Switching Stochastic Volatility model, So et al. 1998). Recently, Luo et al. (2018) incorporated neural networks in the stochastic volatility model (Neural Stochastic Volatility Model), while Xu and Chen (2021) employ deep learning models (Deep Stochastic Volatility Model).

[5]One could then argue that caution is needed when employing basic specifications of stochastic volatility models. For instance, Poon and Granger (2005) found that historical volatility and ARCH models both achieved better volatility forecasting performance than SV models. Similarly, Allen and McAleer (2020, see also Allen, 2020) found that, using realised volatility as benchmark, neither the canonical SV model or a GARCH(1,1) specification could forecast better than a simple form of historical volatility model.

and Chopin and Papaspiliopoulos 2020, who use MCMC to sample from the smoothing distribution of a SV model). As regards filtering exercises, the adoption of particle filters was rather rapid:[6] indeed, latent volatilities in Kim et al. (1998) were already filtered by employing the particle filter suggested in Pitt and Shephard (1999), paving the way for subsequent applications (*inter alia*) in Chib et al. (2006), Omori et al. (2007), Kim and Stoffer (2008) and Nakajima and Omori (2012).

3 Data Description

As previously mentioned, we will analyze the behaviour of the described filtering tools associated to a simple stochastic volatility model. In particular, we are evaluating the performance of such model on real data. For the observable process, namely financial returns in the SV model, we consider the continuously compounded daily returns (also called logarithmic returns) of three indices, S&P500, DOW JONES and STOXX50, in a time interval from June 1st, 2017 to May 30th, 2021. From these, we estimate the daily volatility, as computed by the model.

The indices have been selected as representatives of the global economic trends in the US and EU markets. Specifically, the S&P500 is a market-capitalization-weighted stock-price index tracing the performance of the 500 largest companies listed on US stock exchanges (NYSE and Nasdaq Exchange). The DOW JONES, instead, is a price-weighted stock-market index and accounts for the 30 major companies listed on US stock exchanges, characterized for being "blue-chip". Also the EURO STOXX 50 follows bluce-chip stocks representing leading firms in regions of the Eurozone.

The three plots below represent the time series of the log returns calculated for the three indices during our period of interest. In particular, each series is a sequence of daily observations representing the logarithm of the ratio between the closing price of the index for a given day and the closing price of the day before.

---

[6]Actually, SV models are now often used as straightforward applications of particle filters on non-linear state space models, see for example Andrieu et al. (2010), Douc et al. (2014) or Chopin and Papaspiliopoulos (2020).
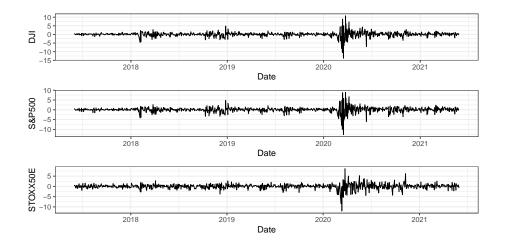
Figure 1: Close to close percentage log returns, 3 indices

At a first visual inspection, we can see a common behaviour in the volatility of all three indices. Especially, it is worth pointing out the similarity between the pattern of DJI and that of S&P500 (especially in terms of peaks), which both differ from the STOXX50E, presumably because the first two describe the US stock market, while the last index describes the European one. For example, a difference that can be seen by having a glance at the plots is that the US indices log returns displays differences in fluctuation magnitude across different periods that are more marked than for the EU index, for which log returns display, in general, wider fluctuations. Anyways, what stands out the most in all three plots is the very wide fluctuations present from February - March 2020 to around September 2020, which, very intuitively, are connected with the Covid-19 Pandemic Crisis.

Table 1: Summary statistics, close to close percentage log returns

	S&P500	Dow Jones	STOXX50E
Mean	0.05581	0.05002	0.01229
Standard Deviation	1.32564	1.38510	1.22951

As regards the series that we will employ in section 5, namely the series of realised volatilities which we will employ as a proxy for the true latent volatilities, the data were retrieved from the Oxford-Man Institute's Realized Library, which provides several different daily non-parametric measures of past volatility.[7] We consider a proper rescaling of the *rv5* series, *i.e.* model-free

---

[7]Such class of measures will be better presented in section 5, when reviewing the use of realised volatility estimators as proxies for latent volatilities.

daily volatility estimates based on 5 min intraday return intervals. Also in this case, the chosen interval spanned from June 1st, 2017 to May 30th, 2021. The plots below represent the realized volatility in the period of interest for all the three indices.[8] As usual, they show a rather similar pattern across all three indices, identifying peaks corresponding to periods of higher uncertainty (in particular referring to the 2020 Covid-19 Crisis outbreak and its impact over time). Once more, there is great similarity between the series for the DJI and the one for the S&P500, while the series of the STOXX50E appears slightly different, in terms of peaks.
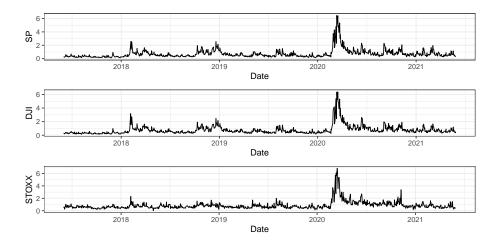
Figure 2: Rescaled realised daily volatility, 3 indices

3.1 Rescaling the data

When specifying the parameters of a stochastic volatility model, one should clarify whether the series of observed financial returns is one of log returns or of percentage log returns. Indeed, let y_t denote the returns at time t, so that a basic SV specification would prescribe $y_t|x_t \sim N(0, e^{x_t})$: clearly, for each period t, e^{x_t} needs to be higher to replicate the observed *percentage* log return rather than the observed log return. This implies that the latent volatility x should fluctuate around higher values when using percentage log returns as y_t. This feature, when adopting the aforementioned canonical SV model as defined in Kim et al. (1998), would be represented by a higher μ. For example, as noted in Kastner (2019), daily log returns often have a variance of 0.0001 or less, which implies that μ should lie somewhere around $\ln(0.0001) \approx -9$; instead, when considering daily percentage log returns, one then has a variance of 1, so that $\mu = \ln(1) = 0$. In other terms, one might end up with a basic specification of

---

[8]In the graph, the series $rv5$ is rescaled as specified in the upcoming section: we took the square root of each value, which is a daily variance measure, and multiplied it by 100.

the SV model when (roughly) calibrating the parameters of the canonical model on a wider set of financial returns.

Note that rescaling the log returns has analogous implications for the section in which we employ measures of realised volatility as a proxy for the latent volatility process. Let us denote by $rv5_t$ the daily realised variance at time t, which we obtain from the *rv5* series provided in the Oxford-Man database. Now, note that in a SV framework one should not directly compare $rv5_t$ to x_t, as $rv5_t$ can be rather seen as an estimate of the variance of the log returns at time t, *i.e.* of e^{x_t} in the SV model.[9] Moreover, by construction of $rv5_t$, the comparison should feature a SV model where the log returns did *not* undergo the percentage transformation.[10] Thus, in order to achieve a common scale with the percentage log returns, we need to multiply $rv5$ by 100^2. To ease the interpretation and representation of the comparison between differently approximated filtering distributions, we finally take the square root of the rescaled $rv5$, so that we compare it to $e^{x_t/2}$.

4 Comparing the filters

In this section the different particle filter techniques we discussed in earlier sections will be employed to approximate the filtering distribution that springs from a simple SV model applied to a real dataset. First, we provide a qualitative comparison between the filters, basing on visual analysis of the (approximated) filtering distributions. Subsequently, we move to a more quantitative approach, in order to assess the quality of the approximations as the number of particles increases. In a way, while in the first part most filters can be rather confidently assumed to be close to convergence, the second subsection will hint at possibly different computational costs across the competing filters.

Before moving to such analyses, it is worth mentioning that not all the filters we employ approximate the same filtering distribution $p(x_t|y_{1:t}, \psi)$, where ψ is a vector containing the parameters of the model. Indeed, the Liu and West filter actually approximates the distribution of a slightly different SV model, one that

---

[9]See for example Guidolin and Pedio (2021), who employ $rv10$, another measure of realised volatility provided in the Oxford-Man Library, as a proxy for the latent conditional variance of the daily log returns on the FTSE100 index.

[10]Indeed, this estimator is based on the sum of 5-minute intra-day squared returns: such returns are computed as the difference between the opening and closure price in an interval.

assumes an additional hierarchical level,

$$y_t|x_t \sim N(0, e^{x_t})$$
$$x_t|x_{t-1}, \alpha, \beta, \tau^2 \sim N(\alpha + \beta x_{t-1}, \tau^2)$$
$$x_0 \sim N(0, 100)$$
$$\alpha \sim N(\gamma, \zeta)$$
$$\beta \sim N(\pi, \phi)$$
$$\tau^2 \sim IG(\nu/2, \lambda\nu/2)$$

where $\gamma, \zeta, \pi, \phi, \nu$ and λ are assumed to be known.[11]}. The initial state is drawn from a normal distribution with commonly used parameters for percentage log returns, as reported by Kastner (2019).

For simplicity, in

$$y_t|x_t \sim N(0, e^{x_t})$$

$$x_t|x_{t-1}, \alpha, \beta, \tau^2 \sim N(\alpha + \beta x_{t-1}, \tau^2)$$

we set $\alpha = 0$, a value that in general, as we argued earlier, also has an empirical justification for percentage log returns. We choose $\beta = 0.99$, as one of the basic underlying ideas of SV models is considerable persistence in the volatility. Finally, we set $\tau^2 = 0.05$.

In the specification used by the LW filter, we have $\gamma = 0$ and $\pi = 0.99$, as to center the priors on the point guesses of the simpler model; ζ and ϕ are both set to 0.01, then allowing a 0.1 standard deviation on both α and γ. For τ^2, we choose an Inverse Gamma distribution whose expected value coincides with the point guess 0.05, by setting $\nu = 5$ and $\lambda = 0.05 * 3/5$.

4.1   Visual Analysis after Filtering

We now offer a preliminary comparison of the various filtering methods, based on a visual analysis of graphs representing, for each index, the time-varying volatility of the observations, obtained through the application of the different filtering techniques explored in the theoretical analysis. Plots that are compared represent the sequence of percentage log returns for a given index (the observations, in gray) and the sequence of filtered volatilities calculated using filtered states (in red) obtained through a specific filtering technique indicated in the caption of the plot.[12]

---

[11]In these empirical applications we keep the values of the parameter we employed in the synthetic data part.

[12]Such filtered volatilities are obtained as follows. For each t, let us denote by μ_t the mean of the generated particles; the filtered volatility is then defined as $e^{\mu_t/2}$. Note that the number of particles was set at 10000 for each filter: the conclusion of the next section will come back to such a choice.
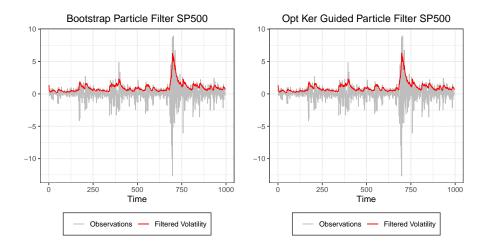
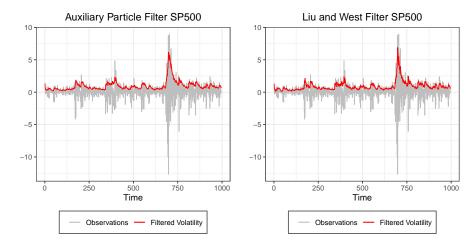Figure 3: Observations and filtered volatility, 3 indices

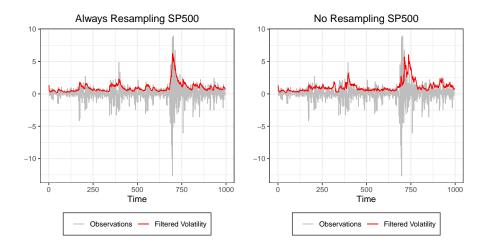Figure 4: Observations and filtered volatility, 3 indices

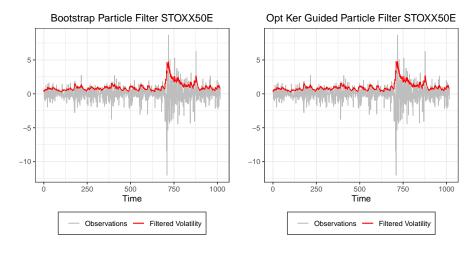Figure 5: Observations and filtered volatility, 3 indices

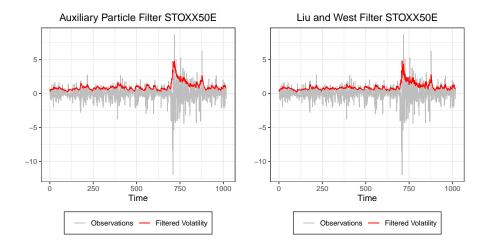Figure 6: Observations and filtered volatility, 3 indices

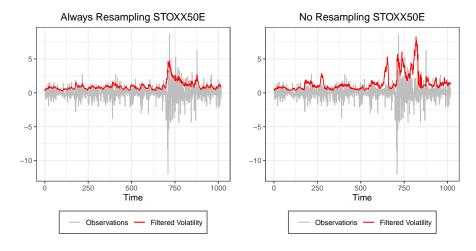Figure 7: Observations and filtered volatility, 3 indices

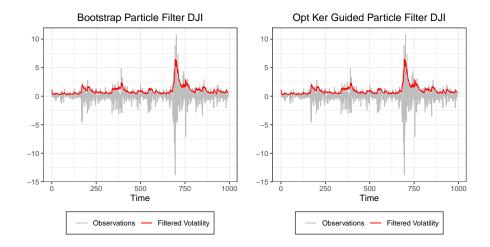Figure 8: Observations and filtered volatility, 3 indices

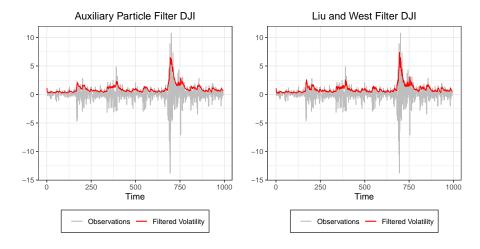Figure 9: Observations and filtered volatility, 3 indices

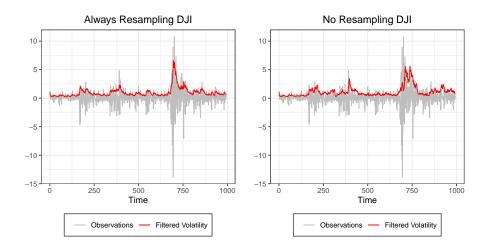Figure 10: Observations and filtered volatility, 3 indices

13

Figure 11: Observations and filtered volatility, 3 indices

First, a general point is that the sequence of percentage log returns are different across the three indices. The reasons why this holds can be associated with the index-construction method and the characteristics of the economy and industrial setting it represents. In particular, as previously noticed, there are extensive similarities between the DOW JONES Index and the S&P500, which can be visualized when looking at the percentage of log returns in the corresponding plots, since both are describing the US Economy. However, the STOXX50E differs quite extensively, especially in the different sensitivity to the Covid-19 Crisis and its consequences, representing the European Economy. Aside from this note, the considerations expressed below on the sequences "filtered volatilities" for the various techniques hold almost uniformly for all the three indices. Therefore, in the next paragraph we will be referring to the set of plots pertaining to one specific index (which could be any of the three).

Generally speaking, note that, the sequence of "filtered volatilities" obtained from any of the filtering techniques follows roughly similar paths across all the plots. Unsurprisingly, the filtered volatility appears wider in periods corresponding to wider fluctuations of the percentage log returns and lower in periods corresponding to smaller such fluctuations. For example, all sequences of filtered volatilities show peaks corresponding to the higher fluctuations due to the 2020 Covid-19 Crisis outbreak and the consequent increases in uncertainty for the later periods.

When focusing the attention on specific filters, first we can discuss the effect of resampling. Indeed, in the plot referring to the Sequential Importance Sampling procedure, the sequence of filtered volatilities is not able to reproduce the peaks detected by all the other filters, often pointing at completely different timing

14

and extents of such peaks. Then, if we instead look at the filtered volatilities obtained with resampling, we can appreciate how the filtered sequences for the Basic,[13] Bootstrap, Auxiliary and Optimal Kernel Guided particle filters give rather analogous results, with no differences appearing at a first visual inspection.

However, the sequence of filtered volatilities obtained by applying the Liu and West filter is somewhat different from the others. Indeed, it shows a "more detailed" sequence of states, meaning that the sequence of filtered volatilities is less smooth and features wider oscillations, with seemingly higher responsiveness to sudden changes in the observations. Therefore, we may expect that, compared to other filtering techniques, it can provide a better representation of the data.

4.2 Quality of approximation

Since we do not have access to the exact filtering distribution of the SV model, when assessing how rapidly different particle filters converge to it, we actually need to approximate even such target distribution as to have a benchmark.

Fortunately, convergence results about particle filters ensure that when the number of particles diverges, the approximation converges to the target distribution.[14] Thus, we run one of the particle filters with a very large number of particles, and assume that such approximation is close enough to the target to be used itself as the distribution the filters should tend to.

With such a benchmark, we can then compare how different filters approach it as the number of particles increases. We employ the Root Mean Squared Error and the Mean Absolute Error to measure discrepancies from the benchmark. Note that the Liu and West filter can not be included in such an analysis: trivially, the comparison assumes that the competing algorithms are approximating the same target distribution.[15]

We start by running the Bootstrap particle filter with N=50000, setting it as the benchmark. In the table below we report the RMSE and MAE measures one obtains by comparing the mean of the particles generated at each step by the Bootstrap PF with N=50000 and the mean[16] of the particles generated by other particles filters with N equal to 10, 100, 1000, 5000 or 10000.[17]

---

[13]By this name, we refer to a particle filter that features no adaptive resampling; rather, resampling is conducted at each step of the algorithm.

[14]Riferimento alla sezione di convergenza?.

[15]For an analogous reason, we do not need to conduct this analysis on all the indices, but we can rather focus on one, the SP500. Indeed, what matters in this section is not how well the model fits the data, but rather how different algorithms approach the same distribution yielded by the same model on a given set of data.

[16]As a possible limitation of our exercise, here we do not consider in detail the variance of such generated particles. Nevertheless, one can get at least an idea of the differences in the generated variances by comparing the confidence intervals that will be shown in the graphs below. Unsurprisingly, it will be apparent that including a resampling step is essential to avoid a collapsing effective sample size.

[17]We consider 5 algorithms: a Bootstrap PF, a Guided PF (with the optimal proposal

Table 2: RMSE and MAE, Bootstrap PF 50000 particles

	N	BPF	GPFOPT	APF	BAPF	SIS
RMSE	10	0.30180	0.33672	0.48223	0.45614	1.56629
RMSE	100	0.11646	0.09173	0.16480	0.17403	1.77055
RMSE	1000	0.02907	0.03357	0.05645	0.04786	1.17887
RMSE	5000	0.01512	0.01461	0.02657	0.02742	0.96810
RMSE	10000	0.01046	0.00957	0.02006	0.02085	0.93016
MAE	10	0.23428	0.25395	0.36488	0.36437	1.25687
MAE	100	0.08340	0.06764	0.11595	0.13520	1.39681
MAE	1000	0.02145	0.02448	0.03723	0.03834	0.86836
MAE	5000	0.01110	0.01027	0.01880	0.01940	0.72121
MAE	10000	0.00801	0.00720	0.01421	0.01482	0.71829

As we can see, almost every filter shows dramatic improvement already when passing from 10 particles to 100 particles, while the performance of the SIS algorithm actually gets worse. With N=10 or N=100, the Bootstrap PF and the Guided PF (with optimal proposal kernel) emerge quite clearly as the best performing filters; however, no conclusion should be drawn at this stage, as the large errors seem to suggest quite a wide margin from improvement even for the best performing filters.

As it turns out, such suggestion is confirmed by running the algorithms with N=1000. Passing from 100 to 1000 particles more than halves the RMSEs across the 5 competing filters, with similarly pronounced gains in the MAEs; at N=1000, the Bootstrap and the Guided particle filters still hold their lead over the rest of the field, keeping a shrinking hedge over the Auxiliary PF, which shows most impressive gains when passing from 100 to 1000 particles. Further sizable gains for every filter are observed when increasing the number of particles to 5000, a level at which both the RMSE and the MAE agree on a slight hedge of the Guided and the Bootstrap filters over the Auxiliary and the Basic PFs. Nevertheless, we would argue that already for N=5000, but even more clearly for N=10000, the difference between the Bootstrap, the Guided and the Auxiliary particle filters is so small that none of these filters can be straightforwardly chosen over the others, a figure which visual inspection of the graphs below can confirm. Interestingly enough, no gains seem to arise from resampling at each step with respect to adaptive resampling (note that the Basic PF is preferred to the APF only twice, by the RMSE), whereas the need of some

---

kernel), an Auxiliary PF, a "Basic PF" and a Sequential Importance Sampling algorithm. By "Basic PF" we refer to a Bootstrap PF that features resampling at each step, whereas resampling never happens in the SIS algorithm: one can then consider these 2 algorithms as extreme cases of an adaptive resampling algorithm, namely one that features resampling when ESS is non negative (*i.e.* always) or when $ESS < 0$ (*i.e.* never). In this sense, we use the shorthand "particle filters" when referring to all the 5 algorithms. Finally, note that the Auxiliary Particle Filter employed here does not coincide with that of the section with synthetic data, which employed a different kernel implied by optimality results.

degree of resampling is apparent, as shown by the awful performance of the SIS algorithm with any number of particles. Moreover, the RMSE and the MAE show infrequent disagreement, delivering a rather clear picture of the comparison.
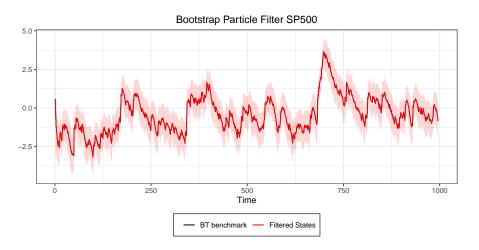

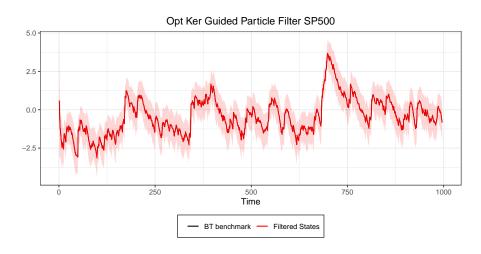Figure 12: Approximations at N=10000, Bootstrap benchmark

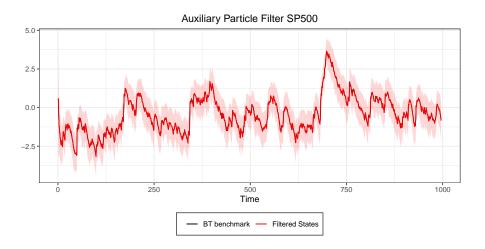Figure 13: Approximations at N=10000, Bootstrap benchmark

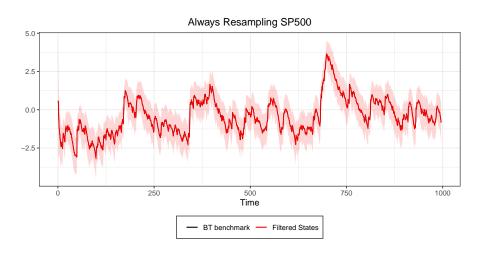Figure 14: Approximations at N=10000, Bootstrap benchmark

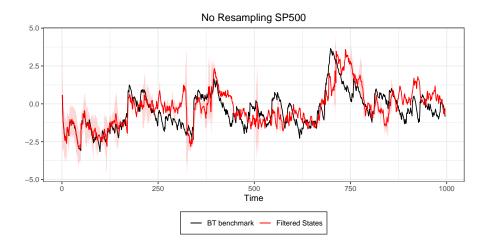Figure 15: Approximations at N=10000, Bootstrap benchmark

18

Figure 16: Approximations at N=10000, Bootstrap benchmark

Now, one might argue that the relative performance of the filters is influenced by the choice of the Bootstrap filter as the benchmark algorithm when run with N=50000: intuitively, if the convergence phase were not reached, the Bootstrap filter (which did in fact come out as one of the best performing algorithms) would clearly find it easier to replicate the benchmark we set.[18]
As a robustness check, we thus conduct the same analysis employing the Auxiliary Particle Filter as the benchmark, keeping the number of particles set to 50000 as for the benchmark BT filter.

Table 3: RMSE and MAE, Auxiliary PF 50000 particles

	N	BPF	GPFOPT	APF	BAPF	SIS
RMSE	10	0.29984	0.33650	0.48285	0.45556	1.56518
RMSE	100	0.11720	0.09527	0.16655	0.17153	1.77045
RMSE	1000	0.03340	0.03979	0.05696	0.04847	1.17822
RMSE	5000	0.02214	0.02074	0.03056	0.02916	0.96819
RMSE	10000	0.01880	0.01745	0.01929	0.02212	0.92985
MAE	10	0.23342	0.25410	0.36516	0.36344	1.25483
MAE	100	0.08307	0.06927	0.11743	0.13432	1.39720
MAE	1000	0.02415	0.02793	0.03692	0.03878	0.86665
MAE	5000	0.01444	0.01360	0.02031	0.02076	0.72037
MAE	10000	0.01185	0.01150	0.01438	0.01495	0.71820

Using the APF as the benchmark does not deliver a different picture than the one we previously had. Leaving the SIS algorithm aside (whose analogous

---

[18]Note that all the filters, including the BT filter run with 50000 particles, shared the same seed in the R software, ruling randomness out of these considerations.

19

performance comes as no surprise), it is interesting to note how once again the Auxiliary particle filter actually fares worse than the Bootstrap and the Guided PFs at each threshold (though very slightly so for the higher ones). Its performance ameliorates dramatically at N=1000, while it still shows some sizable improvements even when passing from N=5000 to N=10000, almost halving its errors. Moreover, note that the mere fact that at N=10000 the Guided and the Bootstrap particle filters approximate better the APF benchmark suggests that the APF still has some margins for relative improvement even at high levels of N, pointing at higher computational costs than the competing filters: trivially, since at N=50000 its RMS and MA difference with respect to this benchmark must be zero, the APF overtakes the Guided and the Bootstrap PFs at N >> 10000.

Table 4: Discrepancy between Bootstrap and Auxiliary PFs

	RMSE	MAE
N=50000	0.01518	0.00868
N=100000	0.01374	0.00647

By comparing the two benchmark approximations with N=50000, we can see that some discrepancy between competing approximations still persists at high numbers of particles. Conveniently - as a different outcome would point at a theoretically-unsound difference between the limiting distributions of the filters - running the Auxiliary and the Bootstrap filters with 100000 particles testifies a shrinking difference between the approximations - at increasingly large computational costs, though.

Overall, we should not overplay tiny differences.[19] Rather, we prefer to stress the points that emerged from both benchmark analyses and we can therefore support quite confidently.
First, if on the one hand it is clear that some degree of resampling is essential, on the other we found no gains from resampling at each period.
Second, when the number of particles is set at 5000 or more, neither benchmark could single out an algorithm that clearly outperformed the others, when restricting the analysis to procedures that feature adaptive resampling rules.

---

[19]Indeed, the mere extent of such difference also depends on the real dataset which is used for the comparison, implying different target distributions. For example, running the same experiment using the STOXX50E dataset, we have similar differences between the filters at N=50000, with RMS difference between the Bootstrap and the Auxiliary equal to 0.016121989 and mean absolute difference equal to 0.009306925. The extent of the difference clearly depends on the chosen parameters too: for example, at N=50000, choosing $\alpha = -0.0031$, $\beta = 0.9951$ and $\tau^2 = 0.0074$, using SP500 data yields RMS difference equal to 0.04408 and MA difference equal to 0.01668, while using STOXX50E data we have 0.01236279 RMS difference and 0.008028818 mean absolute difference.

Third, when the number of particles was set to 10, 100 or 1000, the Bootstrap particle filter and the Guided particle filter with optimal proposal kernel significantly emerged as the best performing algorithms for our SV model, suggesting that their computational cost is lower than that of the Auxiliary particle filter, which needed more particles to "get rolling".

Finally, some discrepancy between different approximations persisted even when the number of particles was set to 50000 or 100000. However, the narrow extent of the RMSEs and MAEs with respect to both benchmarks supports the idea that at N=10000 the best-performing algorithms have already entered the phase of convergence. Even in light of these markedly more computationally expensive benchmarks, we then choose N=10000 as the number of particles which we routinely employ in this empirical application.

5 Realised volatility as benchmark

In section 4 we compared several filtered distributions, briefly pointing out three main behaviours, namely that of the SIS algorithm, that of the Liu and West particle filter and that of the remaining filters. However, we did not have metrics that could help us clearly point at one behaviour as the closest to reality. In this section, we will employ realised volatility as one such metric.

A realised volatility measure - the realised variance based on 5 min intraday return intervals ($rv5$ series in the Oxford-Man Realized Library) - will be used as a proxy for the latent volatility process driving the log returns on the 3 indices presented in the earlier sections. As we previously mentioned, a transformation is needed as to compare filtered states and realised volatility, namely one turning both into estimates of the standard deviation of the percentage log returns. For each t, $rv5_t$ is then turned into $rv5_t^{0.5} * 100$, while $e^{x_t/2}$ (where x_t is the filtered state) is computed by taking the mean of the generated particles and then applying the $exp(\cdot/2)$ transformation.

Throughout this section, we will deem as the best performing algorithm that coming closest to the proxy, either "literally" in a graphical framework or in a Root Mean Squared Error or Mean Absolute Error sense. Clearly enough, this approach builds on the assumption that realised variance is indeed a good proxy for latent volatility in real financial series.

A considerable amount of research in the 1990s and 2000s has focused on the properties of realised volatility measures, let them be realised variances or realised kernels. A thorough review of such literature can be found on the Oxford-Man Institute Realized volatility website, as their database specifically focuses these two measures.[20]

---

[20]Here we would mention the work by Barndorff-Nielsen and Shephard (2002), who focus on the asymptotic properties of the realised volatility error in stochastic volatility models, namely the difference between realized volatility and the "discretised integrated volatility", also called "actual volatility".

As Shephard and Sheppard (2010) themselves put it, in a paper that also presents the methodology that backs their Library, "such statistics are based on a variety of theoretically sound non-parametric estimators of the daily variation of prices". Thanks to their theoretical justification and their being model-free estimates, in recent years realised volatility measures have been frequently used in financial applications as proxies for different types of latent volatility, providing a valuable (and arguably) preferable alternative to using squared returns (Hansen and Lunde, 2006).[21]

Nevertheless, we shall not indulge into forgetting the inherent limitations of using a proxy for a latent process. In particular, realised volatility should not be seen as a perfect proxy of latent volatility (as, for example, realised measures ignore overnight price variations), but rather as a comparatively solid one. Therefore, in this section conclusions will be drawn and put forward only when supported by large differences between models and filters, while we will keep a neutral stance when such support turns out to be mild.

5.1 Comparing the filters

In order to evaluate the different filters according to the evaluation tool represented by the realized volatility, we start once more from the visual inspection of the plots. For any of the three indices, below graphs are shown including both the filtered volatility obtained from the filtered states (for any of the Sequential MC methods described in the theoretical section) and the corresponding rescaled realised volatility series.[22] As compared to the visual analysis we conducted in section 4, we can now use such a proxy to determine the relative performance of the competing filters.

Once more, most algorithms used (aside from SIS and Liu-West) seem to provide indistinguishable results. The SIS algorithm appears completely inadequate to describe the "true" volatility process: as we argued in section XX, the degeneracy of its weights on the particles prevents it from delivering a good enough approximation of the target distribution (*check whether this is the correct intuition*). On top of that, it does not seem the case that the SIS algorithm, by falling short of the SV target distribution, ends up fitting the proxy better.[23]

On the other hand, the Liu and West filter offers the best performance among all filters, at least visually. This suggests that, since the Liu and West particle

---

[21]See for example Kambouroudis et al. (2016), Buncic and Gisler (2016), Gatheral et al.(2018), Allen and McAleer (2020), Guidolin and Pedio (2021).

[22]For simplicity,troughout the rest of this section we will omit "rescaled" when refering to the rescaled realised volatility.

[23]As we will now see, this seems instead to be the case for our preferred algorithms, namely those adopting adaptive resampling rules: a better approximation of the SV target filtering distribution yields slightly higher RMSEs and MAEs than those resulting from less accurate MC approximations.

filter is built on a different underlying model than the other filters considered, this model may offer a better representation of the data analyzed. Once more, these considerations hold for all of the three indices.
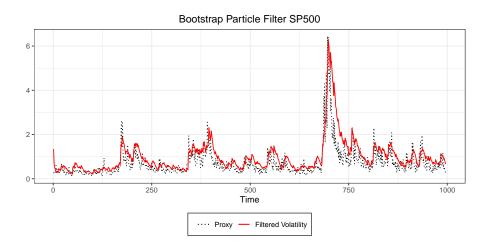
Bootstrap Particle Filter SP500

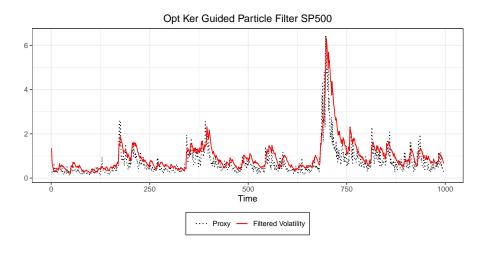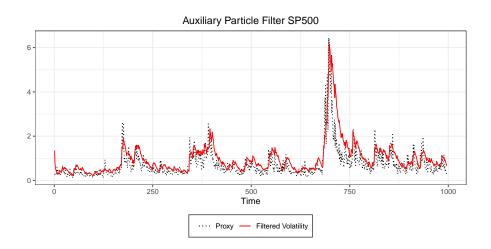Figure 17: Filtered volatility and realised volatility

Opt Ker Guided Particle Filter SP500

Figure 18: Filtered volatility and realised volatility

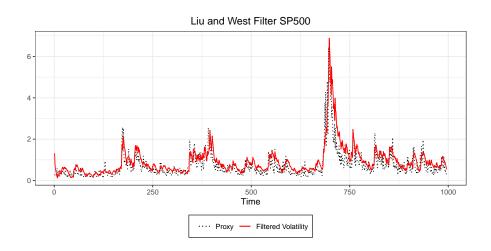Figure 19: Filtered volatility and realised volatility

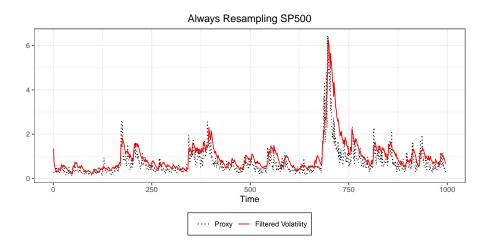Figure 20: Filtered volatility and realised volatility

24

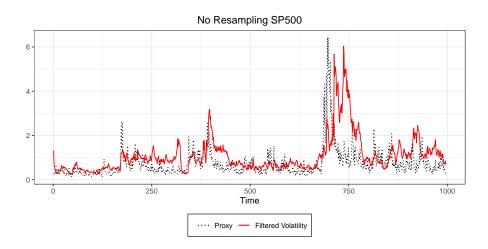Figure 21: Filtered volatility and realised volatility

Figure 22: Filtered volatility and realised volatility

Figure 23: Filtered volatility and realised volatility

Figure 24: Filtered volatility and realised volatility

Figure 25: Filtered volatility and realised volatility

Figure 26: Filtered volatility and realised volatility

Figure 27: Filtered volatility and realised volatility

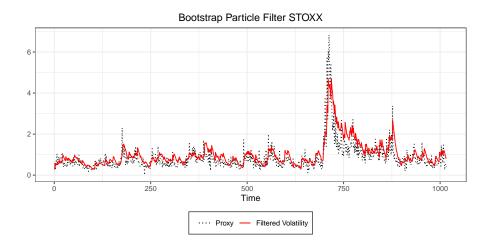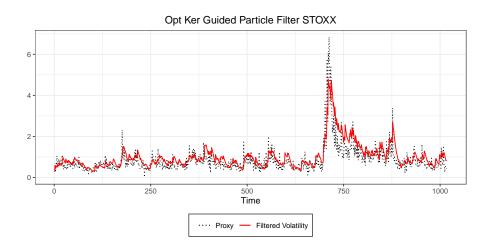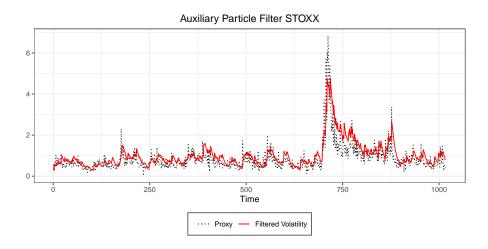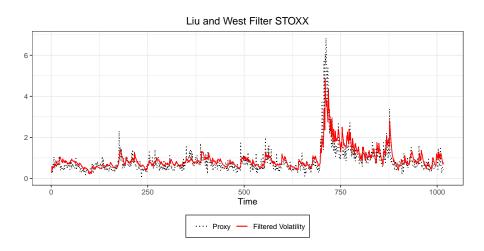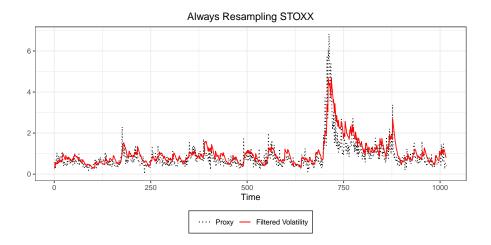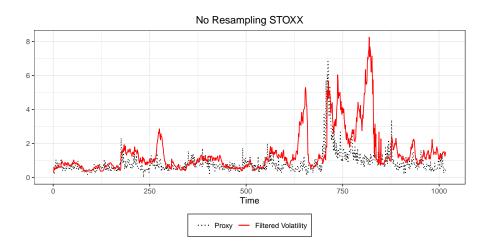
Figure 28: Filtered volatility and realised volatility

The above intuitions are confirmed by looking at the table below, where the RMSE and MAE are shown for each filter (run with a number of particles equal to 10000), considering the realised volatility as benchmark. Indeed, for any index, most of the filters give rather similar results in terms of RMSE and MAE. For any index and for any measure, the worst performing is the SIS algorithm, while the best performing is the Liu and West particle filter. Although the latter algorithm was outperformed in the analysis we conducted on synthetic data, in fact we expected such reversal of fortunes: intuitively, while the rest of the field is built on a model with fixed guesses for the unknown

parameters, the Liu and West filters enjoys more flexibility thanks to the higher variability implied by its additional hierarchical level and by its parameter learning, two features that should both pay off in real data analyses.

Table 5: RMSE and MAE, realised volatility as benchmark

	N	BPF	GPFOPT	APF	BAPF	LW	SIS
RMSE SP500	10000	0.4779	0.4777	0.4760	0.4730	0.3989	0.8987
MAE SP500	10000	0.3249	0.3247	0.3234	0.3235	0.2805	0.5509
RMSE STOXX	10000	0.3952	0.3961	0.3962	0.3969	0.3526	1.2853
MAE STOXX	10000	0.2776	0.2780	0.2790	0.2781	0.2374	0.7285
RMSE DJI	10000	0.4892	0.4877	0.4910	0.4896	0.4317	0.8973
MAE DJI	10000	0.3131	0.3129	0.3135	0.3138	0.2758	0.5558

Finally, we can see from the table below the RMSE and MAE obtained with a number of particles equal to 50000, for the SP500 and the EuroStoxx50 indices. We consider the Bootstrap particle filter and the Auxiliary particle filter, which we employed in section 4; in addition to these, we add the Liu and West filter, which did not feature in that section of this work.

In the case of the SP500 index, although the number of particles increases, both RMSE and MAE increase (although by very small amounts) for the Auxiliary Particle Filter, while they slightly decrease for the Bootstrap PF. Instead, for the Liu and West filter we see a sizable deterioration in the performance: this suggests that the parameter learning process embedded in the LWF is actually pointing in a different direction than that of the proxy, suggesting that either the proxy does not capture well the latent volatility, or a SV model may not be the best tool to disclose such latent process.[24]

In the case of the EuroStoxx50 index, while the Bootstrap filter slightly increases both RMSE and MAE when N is raised to 50000, the APF shows mild improvements in its MAE performance, while the RMSE criterion points in the opposite direction. The criteria disagree also in the case of the LW filter, as the MAE mildly prefers the performance of N=10000, while the RMSE detects minor improvements for N=50000.

Table 6: RMSE and MAE, realised volatility as benchmark (SP500 and STOXX50E)

	N	BPF	APF	LWF

---

[24]Note that at N=1000, the LWF yields a 0.3794359 RMSE and a 0.2687336 MAE. In words, worse approximations of the target distribution of the LWF are actually best at replicating the proxy for the SP500. A possible interpretation goes, once again, as that reported in the main text.

RMSE SP500	50000	0.47646	0.47628	0.42007
MAE SP500	50000	0.32425	0.32380	0.28735
RMSE STOXX	50000	0.39640	0.39787	0.34862
MAE STOXX	50000	0.27768	0.27658	0.23809

Nevertheless, as we previously argued, we should not let tiny differences drive our conclusions. The idea behind employing also such 50000 particles approximations (*i.e* more precise and computationally expensive ones) is to check whether the relative performance that we assessed with N=10000 might be significantly influenced by issues of convergence among the filters: possibly, by further approaching the exact filtering distribution implied by the basic SV model, the filters built on such specification would improve their relative performance. This does not seem to be the case, even though the use of a proxy of the latent volatility process leaves room for discussion, as such conclusions rely on the assumption that such proxy is indeed a good one.[25]

Rather, what this exercise highlights (under its assumptions) is a need for more flexibility and variability in the basic SV model: the underlying SV model of the LW filter meets this request, but clearly other solutions are possible, and have been argued for in the field of financial econometrics and mathematical finance, as we noted in the literary review section.

6 Bibliography

Allen, D. E., & McAleer, M. (2020). Do we need stochastic volatility and generalised autoregressive conditional heteroscedasticity? Comparing squared end-of-day returns on FTSE. Risks, 8(1), 12.

Allen, D. E. (2020). Stochastic Volatility and GARCH: Do Squared End-of-Day Returns Provide Similar Information?. Journal of Risk and Financial

---

[25]Indeed, it can be argued that the issue is almost the opposite one, as in some cases the filters replicate the proxy better when they are further from convergence. For example, note how comparing the following table for the SP500 dataset with that of the main text reveals that some filters perform better when N=1000 as compared to N=10000, even though we know that at N=10000 they replicate their target distribution better.

Table 7: RMSE and MAE, realised volatility as benchmark (SPX)

	N	BPF	GPFOPT	APF	BAPF	LWF	SIS
RMSE SP500	1000	0.47871	0.48019	0.46912	0.47682	0.37944	0.93192
MAE SP500	1000	0.32426	0.32565	0.31965	0.32437	0.26873	0.63744

Management, 13(9), 202.

Andrieu, C., Doucet, A., & Holenstein, R. (2010). Particle markov chain monte carlo methods. Journal of the Royal Statistical Society: Series B (Statistical Methodology), 72(3), 269-342.

Barndorff-Nielsen, O. E., & Shephard, N. (2001). Non-Gaussian Ornstein–Uhlenbeck-based models and some of their uses in financial economics. Journal of the Royal Statistical Society: Series B (Statistical Methodology), 63(2), 167-241.

Barndorff-Nielsen, O. E., & Shephard, N. (2002). Econometric analysis of realized volatility and its use in estimating stochastic volatility models. Journal of the Royal Statistical Society: Series B (Statistical Methodology), 64(2), 253-280.

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. Journal of econometrics, 31(3), 307-327.

Breidt, F. J., Crato, N., & De Lima, P. (1998). The detection and estimation of long memory in stochastic volatility. Journal of econometrics, 83(1-2), 325-348.

Buncic, D., & Gisler, K. I. (2016). Global equity market volatility spillovers: A broader role for the United States. International Journal of Forecasting, 32(4), 1317-1339.

Chan, J. C. (2013). Moving average stochastic volatility models with application to inflation forecast. Journal of Econometrics, 176(2), 162-172.

Chib, S., Nardari, F., & Shephard, N. (2002). Markov chain Monte Carlo methods for stochastic volatility models. Journal of Econometrics, 108(2), 281-316.

Chib, S., Nardari, F., & Shephard, N. (2006). Analysis of high dimensional multivariate stochastic volatility models. Journal of Econometrics, 134(2), 341-371.

Chopin, N., & Papaspiliopoulos, O. (2020). An introduction to sequential Monte Carlo (Vol. 4). Springer.

Comte, F., & Renault, E. (1998). Long memory in continuous-time stochastic volatility models. Mathematical finance, 8(4), 291-323.

Christoffersen, P., Jacobs, K., & Mimouni, K. (2010). Volatility dynamics for the S&P500: Evidence from realized volatility, daily returns, and option prices. The Review of Financial Studies, 23(8), 3141-3189.

Dimitrakopoulos, S., & Kolossiatis, M. (2020). Bayesian analysis of moving average stochastic volatility models: modeling in-mean effects and leverage for financial time series. Econometric Reviews, 39(4), 319-343.

Douc, R., Moulines, E., & Stoffer, D. (2014). Nonlinear time series: Theory, methods and applications with R examples. CRC press.

Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. Econometrica: Journal of the econometric society, 987-1007.

Eraker, B. (2004). Do stock prices and volatility jump? Reconciling evidence from spot and option prices. The Journal of Finance, 59(3), 1367-1403.

Friz, P. K., Gassiat, P., & Pigato, P. (2021). Precise asymptotics: robust stochastic volatility models. The Annals of Applied Probability, 31(2), 896-940.

Gatheral, J., Jaisson, T., & Rosenbaum, M. (2018). Volatility is rough. Quantitative finance, 18(6), 933-949.

Guidolin, M., & Pedio, M. (2021). Media Attention vs. Sentiment as Drivers of Conditional Volatility Predictions: An Application to Brexit. Finance Research Letters, 101943.

Hansen, P. R., & Lunde, A. (2006). Consistent ranking of volatility models. Journal of Econometrics, 131(1-2), 97-121.

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

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

Heber, Gerd, Asger Lunde, Neil Shephard and Kevin Sheppard (2009). Oxford-Man Institute's realized library. Oxford-Man Institute, University of Oxford, version 0.3. Available at: https://realized.oxford-man.ox.ac.uk/.

Heston, S. L. (1993). A closed-form solution for options with stochastic volatility with applications to bond and currency options. The review of financial studies, 6(2), 327-343.

Hull, J., & White, A. (1987). The pricing of options on assets with stochastic volatilities. The journal of finance, 42(2), 281-300.

Jacquier, E., Polson, N. G., & Rossi, P. (1994). Bayesian analysis of stochastic volatility models. Journal of Business and Economic Statistics, 12(4), 371-389.

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.

Johannes, M., Korteweg, A., & Polson, N. (2014). Sequential learning, predictability, and optimal portfolio returns. The Journal of Finance, 69(2), 611-644.

Kambouroudis, D. S., McMillan, D. G., & Tsakou, K. (2016). Forecasting stock return volatility: a comparison of GARCH, Implied volatility, and realized volatility models. Journal of Futures Markets, 36(12), 1127-1163.

Kastner, G. (2019). Dealing with stochastic volatility in time series using the R package stochvol. arXiv preprint arXiv:1906.12134.

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.

Kim, J., & Stoffer, D. S. (2008). Fitting stochastic volatility models in the presence of irregular sampling via particle methods and the EM algorithm. Journal of time series analysis, 29(5), 811-833.

Koopman, S. J., & Hol Uspensky, E. (2002). The stochastic volatility in mean model: empirical evidence from international stock markets. Journal of applied Econometrics, 17(6), 667-689.

Liesenfeld, R., & Jung, R. C. (2000). Stochastic volatility models: conditional normality versus heavy-tailed distributions. Journal of applied Econometrics, 15(2), 137-160.

Luo, R., Zhang, W., Xu, X., & Wang, J. (2018, April). A neural stochastic volatility model. In Thirty-second AAAI conference on artificial intelligence.

Mao, X., Czellar, V., Ruiz, E., & Veiga, H. (2020). Asymmetric stochastic volatility models: Properties and particle filter-based simulated maximum likelihood estimation. Econometrics and Statistics, 13, 84-105.

Nakajima, J., & Omori, Y. (2012). Stochastic volatility model with leverage and asymmetrically heavy-tailed error using GH skew Student's t-distribution. Computational Statistics & Data Analysis, 56(11), 3690-3704.

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

Pitt, M. K., & Shephard, N. (1999). Filtering via simulation: Auxiliary particle filters. Journal of the American statistical association, 94(446), 590-599.

Poon, S. H., & Granger, C. (2005). Practical issues in forecasting volatility. Financial analysts journal, 61(1), 45-56.

Shephard, N., & Sheppard, K. (2010). Realising the future: forecasting with high-frequency-based volatility (HEAVY) models. Journal of Applied Econometrics, 25(2), 197-231.

So, M. E. P., Lam, K., & Li, W. K. (1998). A stochastic volatility model with Markov switching. Journal of Business & Economic Statistics, 16(2), 244-253.

Xu, X., & Chen, Y. (2021). Deep Stochastic Volatility Model. arXiv preprint arXiv:2102.12658.

Zhang, B., Chan, J. C., & Cross, J. L. (2020). Stochastic volatility models with ARMA innovations: An application to G7 inflation forecasts. International Journal of Forecasting, 36(4), 1318-1328.