Stochastic Volatility with Stable Errors: Estimation, Filtering and Forecasting

Francisco Blasques, Siem Jan Koopman* and Karim Moussa

Vrije Universiteit Amsterdam and Tinbergen Institute, the Netherlands

This version: January 4, 2023

Preliminary version: please do not quote without permission.

Abstract

In this paper we consider the stochastic volatility model with a general stable error distribution. We discuss the challenges of parameter estimation, the filtering of the time-varying latent volatility and its forecasting. The indirect inference method is adopted for parameter estimation, while the recently developed extremum Monte Carlo method is used for filtering and forecasting. We argue that our approach is capable of handling the econometric challenges. We further provide the details of the methods used, and we present illustrations based on simulated data and a time series of Bitcoin returns.

Keywords: Filtering, Forecasting, Smoothing, Indirect Inference, Bitcoin.

1 Introduction

In the analysis of financial time series such as daily fluctuations in stock prices and exchange rates, return series will usually be approximately serially uncorrelated. However, they may not be serially independent because of dependence in the variance. In particular, it is often found that the observational error variance is subject to substantial variability over

*Corresponding author. E-mail: s.j.koopman@vu.nl.

time. This phenomenon is referred to as "volatility clustering" and was first recognised by Mandelbrot (1963b), who wrote that "large changes tend to be followed by large changes, of either sign, and small changes tend to be followed by small changes". This notion and its empirical relevance have led to the generalised autoregressive conditional heteroskedasticity (GARCH) model of Engle (1982) and Bollerslev (1986), which has been widely adopted in the econometrics and finance literature.

Another class of models that allow for the temporal variability in the variance of a financial return series is known as the stochastic volatility (SV) model. The SV model has a strong theoretical foundation in the finance theory on option pricing based on the work of the economists Black and Scholes. The development of SV models has been initiated by the work of Stephen Taylor, in particular in Taylor (1982) and Taylor (1986). It has also been recognised early on that the SV model has a strong connection with the class of nonlinear and non-Gaussian state space models; see Harvey, Ruiz, and Shephard (1994).

The estimation strategy for SV models has been widely regarded as an interesting and challenging problem. Many different econometric methods have been explored. The estimation approaches for SV models have been, for example, based on simple moment matching (Taylor, 1986), generalized method of moments (GMM; Melino & Turnbull, 1990), quasi-maximum likelihood (QML; Harvey et al., 1994), Bayesian methods such as Markov chain Monte Carlo (MCMC; Jacquier, Polson, & Rossi, 1994; Kim, Shephard, & Chib, 1998), the indirect inference method (Gourieroux, Monfort, & Renault, 1993; Lombardi & Calzolari, 2009), efficient method of moments (Gallant, Hsieh, & Tauchen, 1997), simulation-based maximum likelihood (Danielsson, 1994; Sandmann & Koopman, 1998), and direct maximum likelihood via numerical integration (Fridman & Harris, 1998). A more complete review on SV models, including their estimation methods, are collected in the book of Shephard (2005) and in the review chapter of Shephard and Andersen (2009).

Many financial return series are typically noisy and characterized by excess kurtosis and skewness. For example, crypto-currency return series, such as those of Bitcoin, are often contaminated by relatively many extreme values. The initial attempt to account for excess kurtosis in financial returns was by Mandelbrot (1963b). Although he assumed the return series to be independent and identically distributed (IID), he replaced the Gaussian assumption by the stable distribution for return series; see also Fama (1963); Mandelbrot

(1963a, 1967). The stable distribution has recently also been considered in the context of SV models by Vankov, Guindani, and Ensor (2019), and Lombardi and Calzolari (2009), who argue that the stable distribution for the SV model can be more appropriate for financial return series.

Although the flexible specification of the stable distribution is a convenient feature, its econometric treatment is challenging due to the fact that its probability density function (PDF) does not have a closed-form expression, and its second and higher moments do not exist. In Lombardi and Calzolari (2008), Garcia, Renault, and Veredas (2011) and Lombardi and Calzolari (2009) the method of indirect inference (II) of Gourieroux et al. (1993) is explored and is shown to be effective in parameter estimation for the stable distribution. In particular, the last of these contributions employs the symmetric stable distribution in the SV model. In this paper we further explore the II approach by considering the asymmetric stable distribution for the SV model. We show that the II method remains effective in this more general setting.

A remaining issue of importance for the stable SV model is the estimation (or filtering) of the unobserved time-varying volatility, as well as the related task of forecasting volatility. The II method falls short of providing a method for extracting latent variables in the model. Gallant and Tauchen (1998) propose a somewhat involved procedure to estimate latent variables in the context of the II method. More general methods such as particle filtering, including the bootstrap filter (Gordon, Salmond, & Smith, 1993) and the auxiliary particle filter (Pitt & Shephard, 1999) cannot be applied since they require that the conditional observation density (or error density) exists in closed form; this is not the case for the stable distribution. To enable volatility extraction and forecasting for the stable SV model, we offer the solution of applying the Extremum Monte Carlo (XMC) filter, recently introduced by Blasques, Koopman, and Moussa (2022). The XMC filter applies generally to any dynamic model and does not require densities to be available in closed form.

The remainder of this paper has Section 2 to introduce and discuss the stable SV model, Section 3 to provide the details of the II estimation method for the asymmetric stable SV model, Section 4 for introducing the XMC filter and its application to the SV model, and Section 5 to illustrate our econometric approach to a Bitcoin time series.

2 Stochastic volatility model with stable distribution

For a time series of financial returns $y_{1:T} = (y_1, \dots, y_T)$, the SV model is given by

$$\begin{aligned} y_t &= \exp(\alpha_t/2)\varepsilon_t, & \varepsilon_t &\sim p(\varepsilon_t), \\ \alpha_{t+1} &= \mu + \phi(\alpha_t - \mu) + \sigma_\eta \eta_t, & \eta_t &\sim N(0,1), \end{aligned} \tag{1}$$

for $t = 1, \dots, T$, and where α_t represents the unobserved log variance, $p(\varepsilon_t)$ is the observation error density, and with coefficients $\mu \in \mathbb{R}$, $|\phi| < 1$ and $\sigma_\eta > 0$. Since the model is stationary, the initial distribution may be chosen as $\alpha_1 \sim N(\mu, \sigma_\eta^2/(1-\phi^2))$. In the initial development of the SV model, the observation error ε_t was assumed to be normally distributed. However, in later work it has been found that the Gaussian assumption is not sufficient in accounting for several aspects of the conditional distribution, such as excess kurtosis and skewness; see, for example, Gallant et al. (1997) and Durham (2006).

An alternative is to replace the normal distribution for the observation errors by the more general stable distribution (e.g. Casarin, 2004; Lombardi & Calzolari, 2009; Vankov et al., 2019). In this case, we set $p(\varepsilon_t) = S(a, b; 1)$, which denotes the first parametrization of the standard univariate stable distribution as in Nolan (2009), with tail index parameter $a \in (0, 2]$ and asymmetry parameter $b \in [-1, 1]$. Except for a few specific choices of the parameters, the density is not available in closed form, hence the characteristic function is used to describe the distribution:

$$\mathbb{E}[\exp(iu\varepsilon_t)] = \begin{cases} \exp\left(-|u|\left[1 + ib\frac{2}{\pi}(\operatorname{sgn} u)\log|u|\right]\right) & \text{if} \quad a = 1, \\ \exp\left(-|u|^a\left[1 - ib\tan(\frac{\pi a}{2})(\operatorname{sgn} u)\right]\right) & \text{else.} \end{cases}$$

For $a = 2$, the distribution reduces to the normal one, but for other values the distribution has finite moments only of orders less than a. An important property of the 1-parametrization is $\mathbb{E}[\varepsilon_t] = 0$ if $a > 1$, such that the mean is not impacted by the tail and asymmetry parameters. Since the variance does not exist for $a < 2$, the volatility of y_t thus refers to the scale $v_t = \exp(\alpha_t/2)$. To illustrate the impact of a on the time series y_t, we present two simulated time series from the stable SV model in Figure 1. It is clear that the coefficient a has a large impact on the tail of the observation density, especially given that the two values

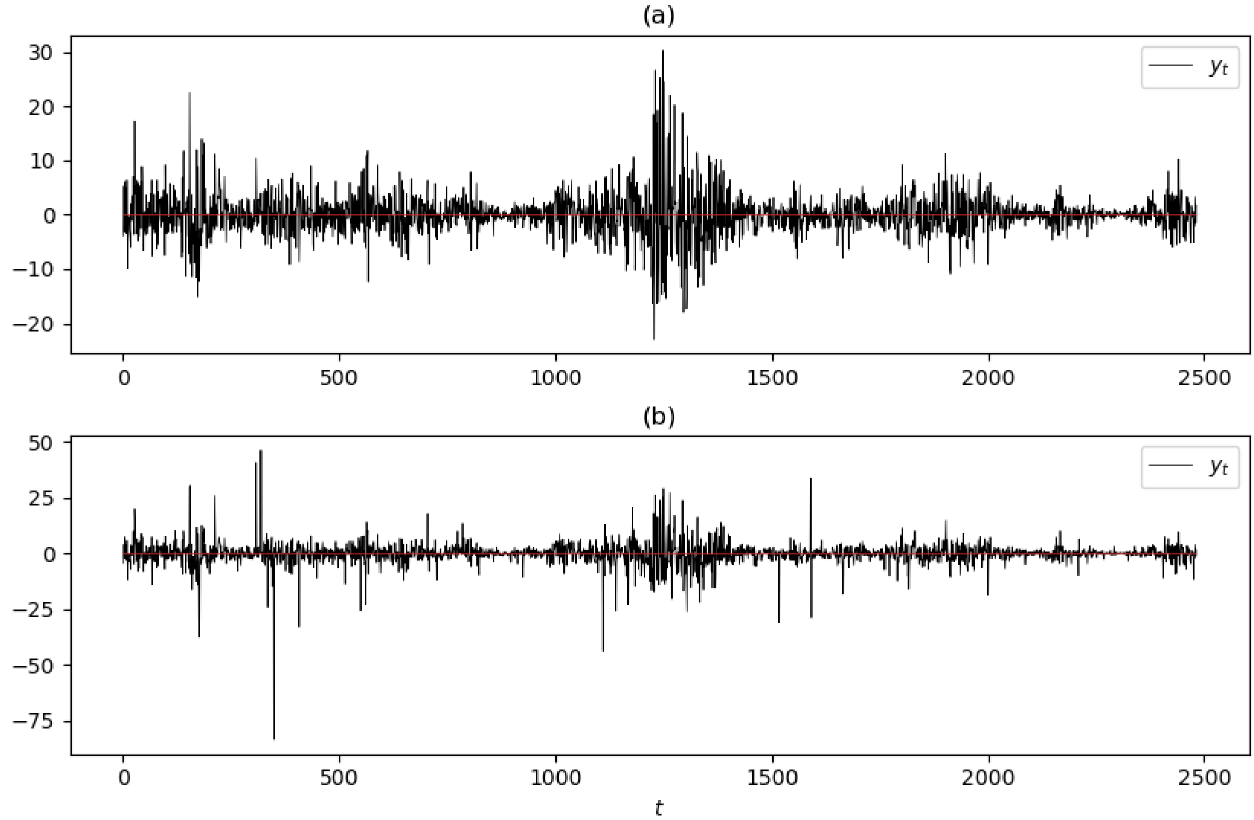

Figure 1: A simulated path of the observations from SV model (1) based on the following parameter values: (a) Gaussian errors ($a := 2$); (b) stable errors ($a := 1.677$).

are not far apart: $a = 2$ in panel (a) and $a = 1.677$ in panel (b).

Apart from its heavy tails for $a < 2$, the stable distribution has several important and intuitive properties. For instance, it is closed under summation, in the sense that any sum of IID stable variates is itself stable. The distribution also arises in the generalized central limit theorem (Gnedenko & Kolmogorov, 1954), which states that any limiting sum of IID variables must be stable. However, despite its intuitiveness, the practical use of the stable distribution has been somewhat limited, which may be due to its lack of analytic tractability. For example, the absence of an analytical expression for the probability density function $p(\varepsilon_t) = p(y_t|\alpha_t)$ implies that default methods of choice are often not directly available, such as maximum likelihood estimation for the parameters, and particle filters (Gordon et al., 1993; Kitagawa, 1996; Pitt & Shephard, 1999) or MCMC methods (e.g. Kim et al., 1998) for the extraction of time-varying volatility in the SV model.

The II method of Gourieroux et al. (1993) is explored for the estimation of the static parameters in the stable SV model. For volatility filtering and forecasting, the XMC filter, as recently introduced by Blasques et al. (2022), is considered because it is able to circumvent the issue of limited model tractability by using basic simulation and regression methods. The II estimation method is discussed in the next section while the XMC filter is explored in Section 4.

3 Indirect inference parameter estimation

3.1 Indirect inference for the asymmetric stable SV model

In this section we propose an indirect inference (II; Gourieroux et al., 1993) estimator for the parameters of the SV model (1) with asymmetric stable errors. The main idea behind the II method is to estimate the parameters of a model of interest, the structural model, by simulating data for various candidate values of its parameter vector, θ. The estimate is then chosen as the value for which the simulated data are most "similar" to the observed data. This similarity is expressed through auxiliary parameters, β, which are the arguments of a corresponding auxiliary objective function $Q(\beta; y_{1:T})$. The first step is to estimate β by optimizing over the auxiliary objective function based on the observed data:

$$\widehat{\beta} \in \arg\max Q(\beta; y_{1:T}).$$

Noting that the "score" $\partial Q/\partial\beta$ equals zero when evaluated at $\widehat{\beta}$ and $y_{1:T}$, *score-based* II uses this property to define an estimator for the structural parameters using $M \in \mathbb{N}$ simulated paths $y_{1:T}^{(j)}(\theta), j = 1, \ldots, M$:

$$\widehat{\theta} \in \arg\min s(\theta)'\Sigma s(\theta), \qquad s(\theta) = \frac{1}{M}\sum_{j=1}^{M} \frac{\partial Q\left(\beta; y_{1:T}^{(j)}(\theta)\right)}{\partial\beta}\bigg|_{\beta=\widehat{\beta}}, \tag{2}$$

with Σ a positive definite weighting matrix. The estimator of the structural parameters is thus defined as the minimizer of a quadratic form determined by the average score over the M simulated paths.

In our case, M was set to 10 simulated paths and the weighting matrix to the identity matrix I_5. Since the number of structural parameters and the number of auxiliary parameters are equal, the optimization problem (2) is just identified (and therefore equivalent to a multivariate root-finding problem), which makes the choice of weighting matrix irrelevant. Furthermore, we have $\theta = (\mu, \phi, \sigma_\eta, a, b)'$ as the vector of structural parameters and Q as the average log likelihood of the auxiliary model that we describe next. Combining the insights from Lombardi and Calzolari (2008), who use a Skew-t distribution as auxiliary model for the stable distribution, and Lombardi and Calzolari (2009), who use a GARCH(1,1) model with Student's t errors as auxiliary model for several symmetric SV models, we propose the following GARCH(1,1) model with Skew-t errors as the auxiliary model:

$$\begin{aligned} y_t &= \sigma_t \xi_t, & \xi_t &\sim CST(\nu, \rho), \\ \sigma_{t+1}^2 &= \omega + \gamma y_t^2 + \delta \sigma_t^2, \end{aligned} \tag{3}$$

where ω, γ, δ, $\nu > 0$, $\rho \in \mathbb{R}$ are static coefficients and $CST(\nu, \rho)$ denotes the centered version of the Skew-t distribution from Azzalini and Capitanio (2003). Following Lombardi and Calzolari (2008) and Lombardi and Calzolari (2009), we additionally impose $\nu < 20$, as well as $a > 1$ in the structural model.

3.2 Illustration for a Bitcoin log return series

We apply the proposed II estimator to the Bitcoin data from Figure 2 (a). The observations are the centered daily log returns of the BTCUSD exchange rate multiplied by 100, starting from September 9$^{\text{th}}$, 2014, to July 7$^{\text{th}}$, 2021 ($T = 2481$).

Table 1 presents the parameter estimates for both the structural and auxiliary model. The parameter constraints are incorporated by reparameterizing the optimization problem, so that optimization could take place over unrestricted transformations of the parameters (e.g. $\widetilde{\omega}$: $\omega = \exp(\widetilde{\omega})$). The tail index estimates indicate that frequent extreme values are to

Table 1: Parameter estimates for the structural model in (1) with stable errors and auxiliary model in (3) based on the daily log returns of Bitcoin.

Parameters	Tail	Asymmetry	Location (scale)	Autocorr. (scale)	Other
Structural	a	b	μ	ϕ	σ_η
	1.677	−0.061	1.910	0.994	0.165
Auxiliary	ν	ρ	ω	δ	γ
	2.508	−0.058	0.035	0.873	0.050

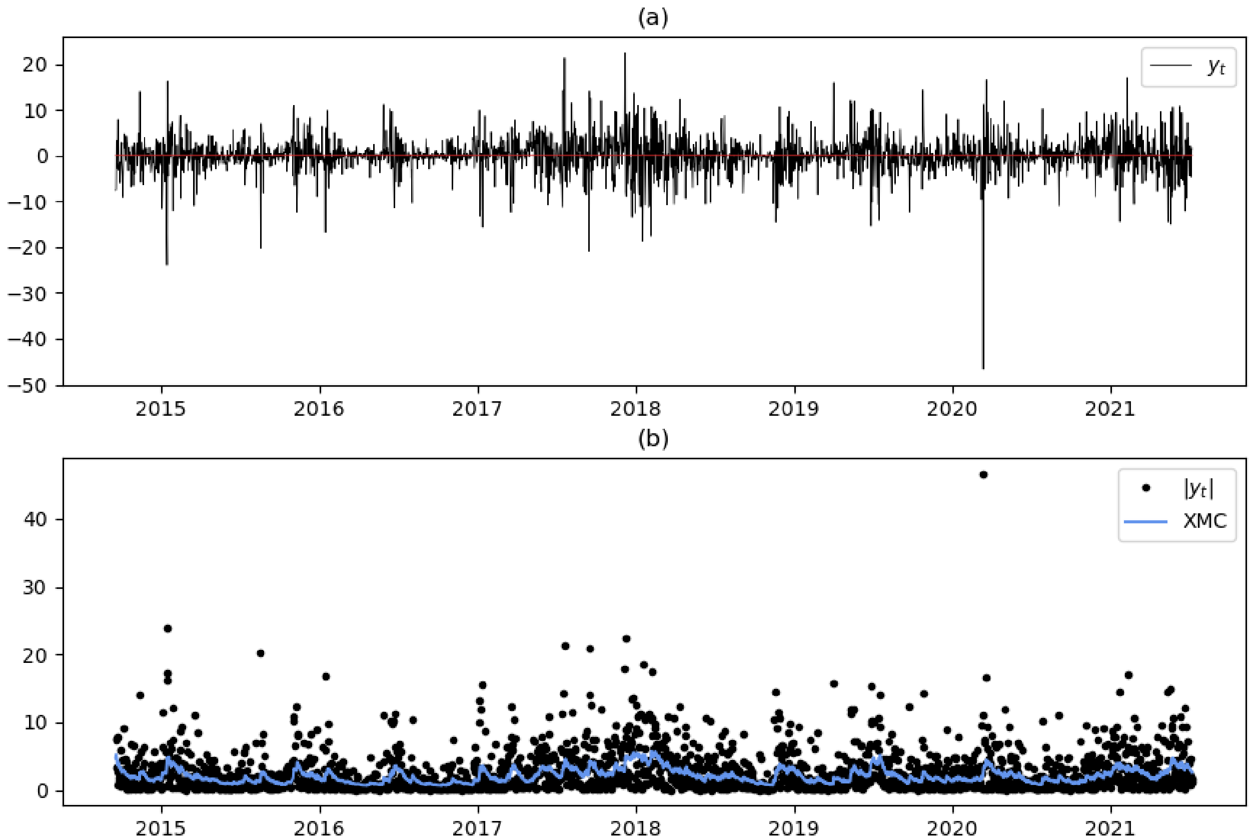

Figure 2: Daily log returns of Bitcoin and filtered estimates of the volatility $v_t = \exp(\alpha_t/2)$: (a) centered log returns times 100; (b) absolute log returns and volatility estimates $\mathbb{E}[v_t|y_{1:t}]$ from the Extremum Monte Carlo filter.

be expected, where we note that the estimate of $\nu = 2.508$ implies that the third moment of y_t does not exist in the auxliary model. The asymmetry in the returns appears to be slightly negative, but to a negligible degree according to both models. As expected, the (log) scale was found to be highly autocorrelated. Overall, the estimates for the structural and auxiliary models seem to be in agreement where direct comparison is possible.

4 Extremum Monte Carlo filtering for the SV model

This section discusses the Extremum Monte Carlo filter (XMC; Blasques et al., 2022, BKM hereafter). The main idea of XMC originates from the "simulation-regression" method of Longstaff and Schwartz (2001) for the valuation of American options in financial trading. Given that we have, say, two stochastic variables X and Y, the method aims at approximating the conditional expectation function $\mathbb{E}(X|Y)$ by simulating N realizations of X and Y from their joint distribution $p(X, Y)$, and performing a least squares regression of X on Y. The estimated regression function $\widehat{f}$ is then evaluated at any point of interest y (e.g. the observed data) to provide the approximation $\widehat{f}(y) \approx \mathbb{E}(X|Y = y)$. The method can be generalized for sets of multiple variables. For example, the XMC method takes X and Y as time series paths instead of scalar variables; in effect, both X and Y are $T \times 1$ vectors where T is the time series length. The regression step is repeated for each time index t, with $t = 1, \ldots, T$. The XMC method can also consider other loss functions such as the absolute error loss. Furthermore, compared with the method from Longstaff and Schwartz (2001), the XMC method improves the approximation by the employment of nonlinear regression methods from the machine learning literature, such as gradient boosting and random forests.

The details of the XMC method are provided by Blasques et al. (2022) which can broadly be applied to any dynamic model, including nonlinear and non-Gaussian state space models, from which data can be simulated. For example, the method can be applied to multivariate SV models since it is not liable to the curse of dimensionality. Furthermore, the use of extremum estimation allows for any conditioning set, including data sets with missing entries and unequal spacing. The XMC filtering method places the computational burden predominantly in the off-line phase. The combination of generality and computational modularity entails that it is able to solve many signal extraction problems in real time.

Algorithm 1 presents a version of the XMC filter that is applied to the SV model given by equation (1). The filtering starts with simulating N paths of the log variances α_t and return observations y_t from the joint distribution implied by the SV model (1). Subsequently, we perform regressions of the α_t variates on the corresponding $y_{1:t}$ variates to predict the states. For example, by using the squared error loss $L(u) = u^2$ in Algorithm 1, this procedure allows for estimating the conditional expectation *functions* $\mathbb{E}[\alpha_t|y_{1:t}]$ for $t = 1, \ldots, T$. In practice, the covariates used in the regressions are subsets of the conditioning set, $y_{1:t}$, chosen to prevent overfitting of the training sample. More specifically, we use the W (window size) observations from the conditioning set nearest to time t, which yields the covariate set $y_{\underset{\sim}{t}:\widetilde{t}}$ with $\underset{\sim}{t} = \max\{t - W + 1, 1\}$ and $\widetilde{t} = t$. Together with the other tuning parameters of the chosen regression method, W is determined from a set of candidate values as the minimizer of the average loss in (4) for a separate validation sample of simulated data.

The combination of simulation and regression leads to a flexible filtering method. By

Algorithm 1 XMC filter algorithm for the SV model (1)

1. **Simulate**: Consider SV model (1) to simulate N paths from the joint distribution,

$$\left(\alpha_{1:T}^{(i)}, y_{1:T}^{(i)}\right) \sim p(\alpha_{1:T}, y_{1:T}), \qquad i = 1, \dots, N.$$

2. **Fit**: Perform the following regression for $t = 1, \dots, T$:

$$\widehat{f}_t^N \in \underset{f \in \mathbb{F}_N}{\arg\min} \frac{1}{N} \sum_{i=1}^{N} L\left(\alpha_t^{(i)} - f\left(y_{1:t}^{(i)}\right)\right), \tag{4}$$

   with function space $\mathbb{F}_N$ and loss function L.

3. **Predict**: Evaluate the estimated regression functions $\{\widehat{f}_t^N\}_{t=1}^T$ at the observed data $y_{1:t}$ for $t = 1, \dots, T$ to predict the states:

$$\widehat{\alpha}_t = \widehat{f}_t^N(y_{1:t}).$$

choosing the loss function appropriately, the method can be used to estimate conditional quantiles (tilted absolute error loss), modes (all-or-nothing loss), and other quantities of interest. By adjusting the conditioning set, it can be used to perform k-period forecasting (conditioning set: $y_{1:t-k}, k \in \mathbb{N}$), or smoothing (conditioning set: $y_{1:t+k}$). In addition, unequally-spaced and unbalanced data sets with missing entries can be handled simply by omitting the appropriate covariates. Furthermore, the same algorithm can be used to predict functions of the states (e.g. $v_t = \exp(\alpha_t/2)$) or forecast the observations, simply by changing the dependent variable in the regressions.

To illustrate the XMC filter, we used the Gaussian version of the SV model in (1) to simulate a path of the log variances α_t and observations y_t, and used the latter to filter the former. The parameters were set to $\mu = 0, \phi = 0.96$, and $\sigma_\eta = 0.16$, based on the maximum likelihood estimates of Sandmann and Koopman (1998) for a time series of S&P500 daily log returns. The simulated paths are shown in Figure 3 (a) and (b). The squared error loss was selected to estimate the expectations $\mathbb{E}[\alpha_t | y_{1:t}]$, which we can expect to be nonlinear in the observations $y_{1:t}$ because α_t is a (transformed) scale parameter. A nonlinear regression method is therefore appropriate, hence we used a tree-based gradient boosting method (GB; Friedman, 2001) as in BKM. Figure 3 (b) shows the filtered states alongside the true states (dots). The regularized window size was $W = 40$, which means the covariate set remained of constant size from $t = 40$ onwards. For comparison, we used the bootstrap filter (BF; Gordon et al., 1993) with 10^6 particles, which was also the number of paths for the XMC filter. The estimates of the two filters are seen to be close, both largely following the movements of the true states, which are relatively high at the beginning and lower near the end of the sample. Further discussion of the XMC filter can be found in BKM, with conditions for convergence to an optimal filter as $N \to \infty$ given in their Section 4.

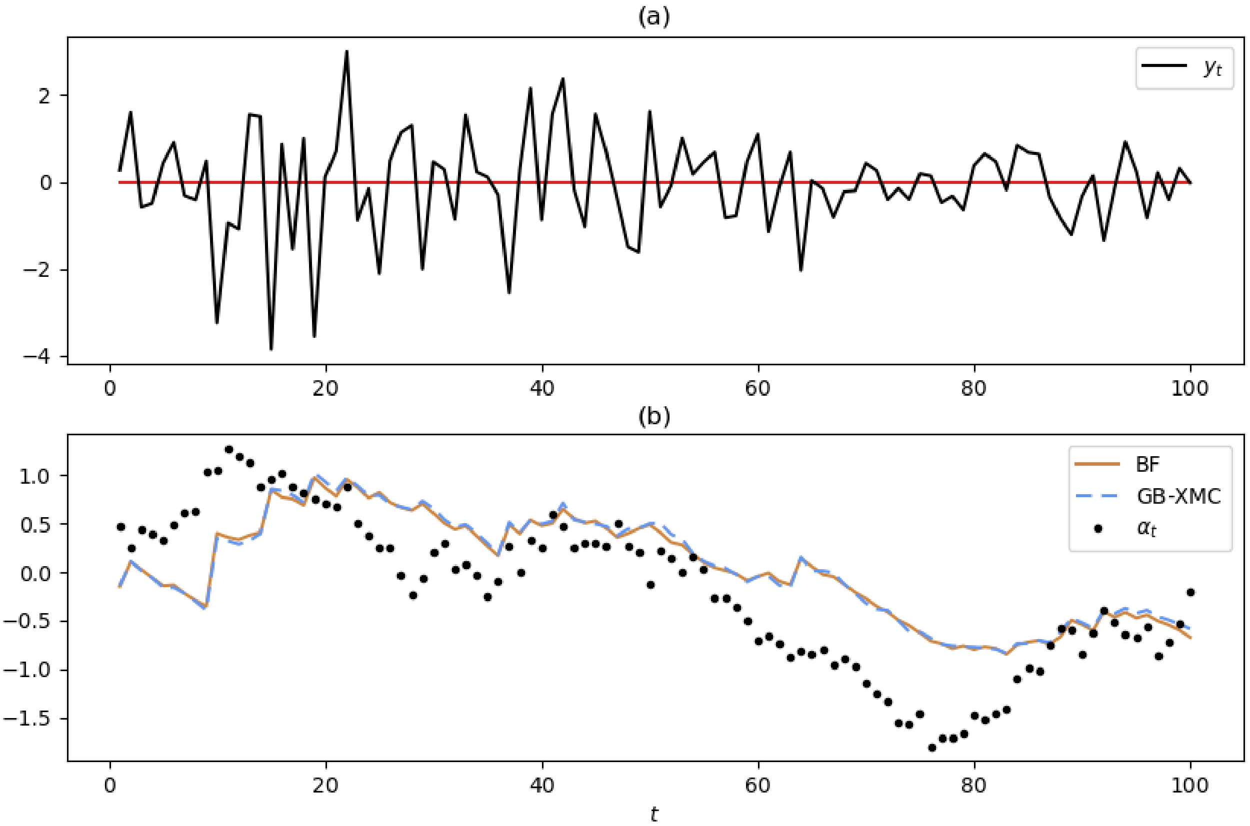

Figure 3: Analysis of a simulated path from the Gaussian SV model: (a) simulated observations; (b) true states and filtered estimates from the BF and GB-XMC filter ($N = 10^6$).

5 Applications

This section provides two pieces of evidence that the XMC filter is able to extract the time-varying volatility for the SV model in (1) with stable errors. We first carry out a simulation study to assess the predictive performance of the XMC filter. We then continue with an empirical illustration where the XMC filter is used to extract the volatility for the daily log returns of Bitcoin. In addition to filtering, we consider 1-period forecasting, and fixed-interval (FI) smoothing, in which the entire sample $y_{1:T}$ is used as conditioning set. More details on this and other variants of smoothing using XMC will be given in the forthcoming work of Moussa, Blasques, and Koopman (2023).

5.1 Simulation study

To assess the predictive performance of the XMC filter for the stable SV model in (1), we performed a simulation study using the following parameters from Vankov et al. (2019): $\mu = -0.2, \phi = 0.95, \sigma_\eta = 0.2, a = 1.75$ and $b = 0.1$. The path length was set to $T = 100$, and the squared error loss was used to train the XMC filter and evaluate the performance of predicting the log variances α_t. The latter was done based on a separate test sample of 10^5 paths, and an equal number of paths was used to train the the XMC filter. For comparison, we used the QML filter from Harvey et al. (1994), which remains applicable without a tractable observation density. The QML filter starts by transforming the observations via $\widetilde{y}_t = \log y_t^2$ to cast the SV model into a linear state space form, that is

$$\begin{aligned}\widetilde{y}_t &= \alpha_t + 2\widetilde{\varepsilon}_t,\\ \alpha_{t+1} &= (1-\phi)\mu + \phi\alpha_t + \sigma_\eta\eta_t,\end{aligned}$$

with $\widetilde{\varepsilon}_t = \log|\varepsilon_t|$. Although $\widetilde{\varepsilon}_t$ is not normally distributed, one can assume it is, so that the Kalman filter (Kalman, 1960) can be used to act as an approximate filter for α_t. This approach uses the mean and variance of $\widetilde{\varepsilon}_t$, which are given by Lemma 3.19 of Nolan (2009) for $a \neq 1$:

$$\begin{aligned}\mathbb{E}[\widetilde{\varepsilon}_t] &= \gamma_{Euler}(1/a - 1) - \frac{1}{a}\log\left(\cos[a \cdot c(a,b)]\right),\\ \mathbb{V}\text{ar}[\widetilde{\varepsilon}_t] &= \frac{\pi^2\left(1 + 2/a^2\right)}{12} - c(a,b)^2,\end{aligned}$$

where $\gamma_{Euler} \approx 0.577$ is Euler's constant, and

$$c(a,b) = a^{-1}\arctan\left(b\tan\left(\frac{\pi a}{2}\right)\right).$$

Figure 4 shows the root mean squared error (RMSE) against time for both methods corresponding to 1-period forecasting (a), filtering (b), and FI smoothing (c). As expected, the RMSEs of both methods decrease as the conditioning set grows, which can be seen both as t increases in plots (a) and (b), and by comparing the plots vertically. Table 2 contains the overall RMSE for both methods. Similarly to the figure, it can be seen that the XMC filter outperforms the QML filter for each choice of conditioning set. Perhaps most conclusive on the difference in predictive quality is the fact that the 1-period forecasts of the XMC filter outperform the filtered QML predictions, even though the former are based on a smaller conditioning set that excludes the most informative element y_t.

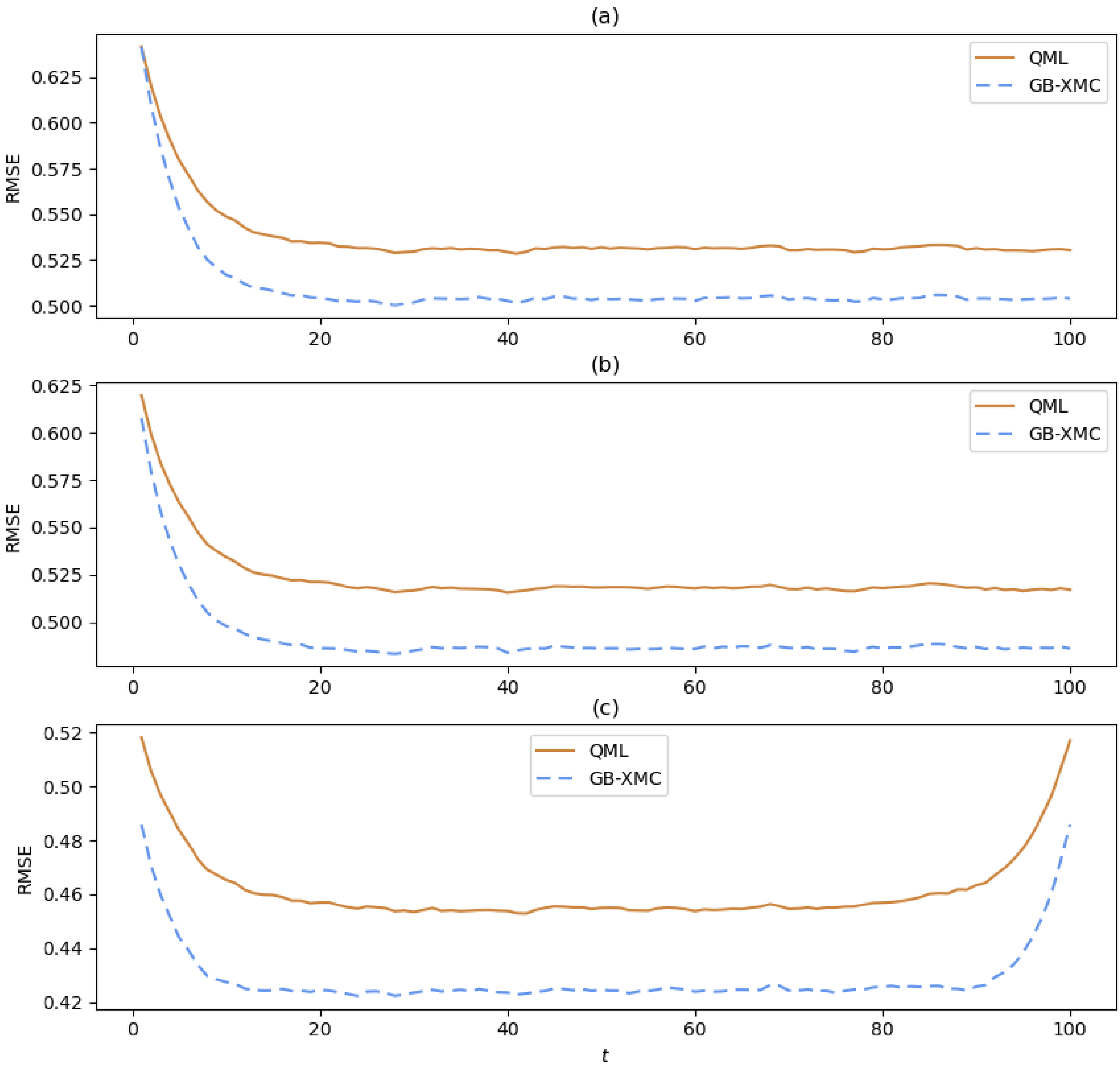

Figure 4: Results from simulation study based on model (1), showing the RMSE over time for the QML and GB-XMC predictions: (a) 1-period forecasting; (b) filtering; (c) FI smoothing.

Table 2: Results from simulation study based on model (1): overall RMSE of the QML and GB-XMC filters for 1-period forecasting, filtering and FI smoothing.

	Method	Forecasting	Filtering	FI smoothing
RMSE	QML	0.537	0.524	0.462
	XMC	**0.510**	**0.492**	**0.429**

5.2 Extracting Bitcoin volatility

We consider several applications of the XMC filter to extract the time-varying volatility from the daily log returns of Bitcoin shown in Figure 2 (a), using the II parameter estimates from Table 1. For long time series such as the one used ($T = 2481$), large computational savings can be obtained with the XMC filter by re-using the estimated regression functions in future time periods. This idea is called a steady state (SS; BKM Section 5), and its workings are particularly simple for filtering, where no new functions are estimated after some time $t_{SS} \geq W$ (where the covariate set remains of constant size). In this case, an SS has been reached, and all subsequent prediction is done using the function estimate $\widehat{f}^N_{t_{ss}}$. The SS approach was used for all applications of the XMC filter in this section.

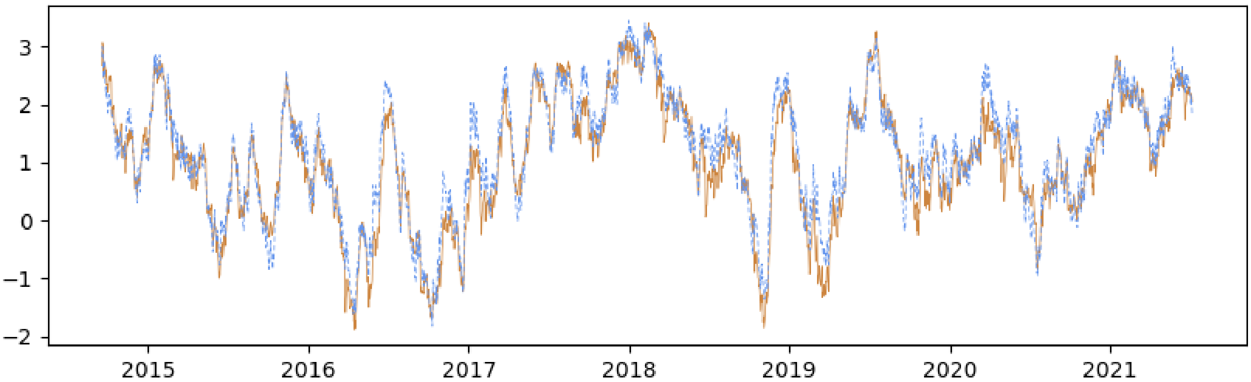

Figure 5: Filtered log variances α_t for the log returns of Bitcoin from the QML (brown) and GB-XMC (blue; $N = 10^5$) filters.

Figure 5 (a) shows the estimates of $\mathbb{E}[\alpha_t|y_{1:t}]$ from the QML and GB-XMC filters ($N = 10^5$), with the SS estimate corresponding to $t = 37$ ($W = 37$). In this case, the SS estimate has reduced the number of regressions by 98.5%. Furthermore, it can be seen that the estimates of both methods are close and tend to move in the same direction as new observations come available. The GB-XMC filter responds somewhat more strongly to the large negative log return of -46.65 at March 12th, 2020, but it is striking that the impact is limited on both filters: the use of stable observation errors does not require a large value of the scale to make extreme observations likely, which explains the robustness of both filters.

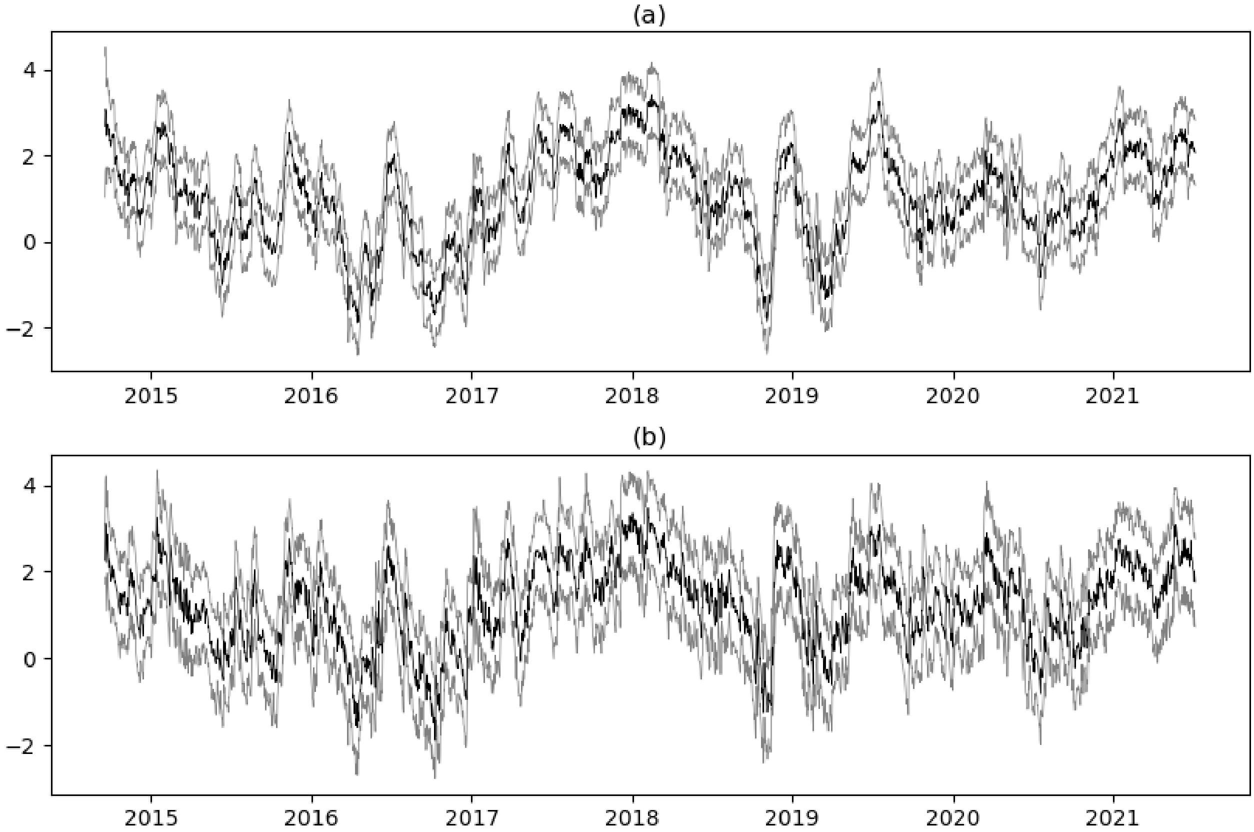

Figure 6: Filtered $10\%, 50\%$, and 90% quantiles of the log variance α_t for the log returns of Bitcoin: (a) QML filter; (b) QRF-XMC ($N = 10^5$).

For many applications it is of interest to estimate conditional quantiles rather than the expectation of the states or future observations, such as in the computation of value at risk and expected shortfall in finance (McNeil, Frey, & Embrechts, 2015). As in BKM we use the quantile regression forest (QRF; Meinshausen, 2006) for this purpose, which extracts information of the conditional distribution from the weights of a random forest. Figure 6 (a) and (b) show the $10\%, 50\%$, and 90% quantile estimates of α_t from the QML and QRF-XMC filters, respectively. The QRF-XMC estimates are based on $N = 10^5$, and an SS estimate corresponding to $t = 170$ ($W = 35$), which has reduced the number of regressions by 93.1%. The differences between the two methods are more pronounced than when the mean was estimated. Considering again the period shortly after the extreme negative log return at the start of 2020, the 90% quantile was estimated close to four by the QRF-XMC filter, which differs substantially from the corresponding QML estimate of around two. In general, the

Gaussian approximation by the QML filter is less suitable when the analysis extends beyond the first two moments. In such situations, the XMC filter may provide an adequate solution.

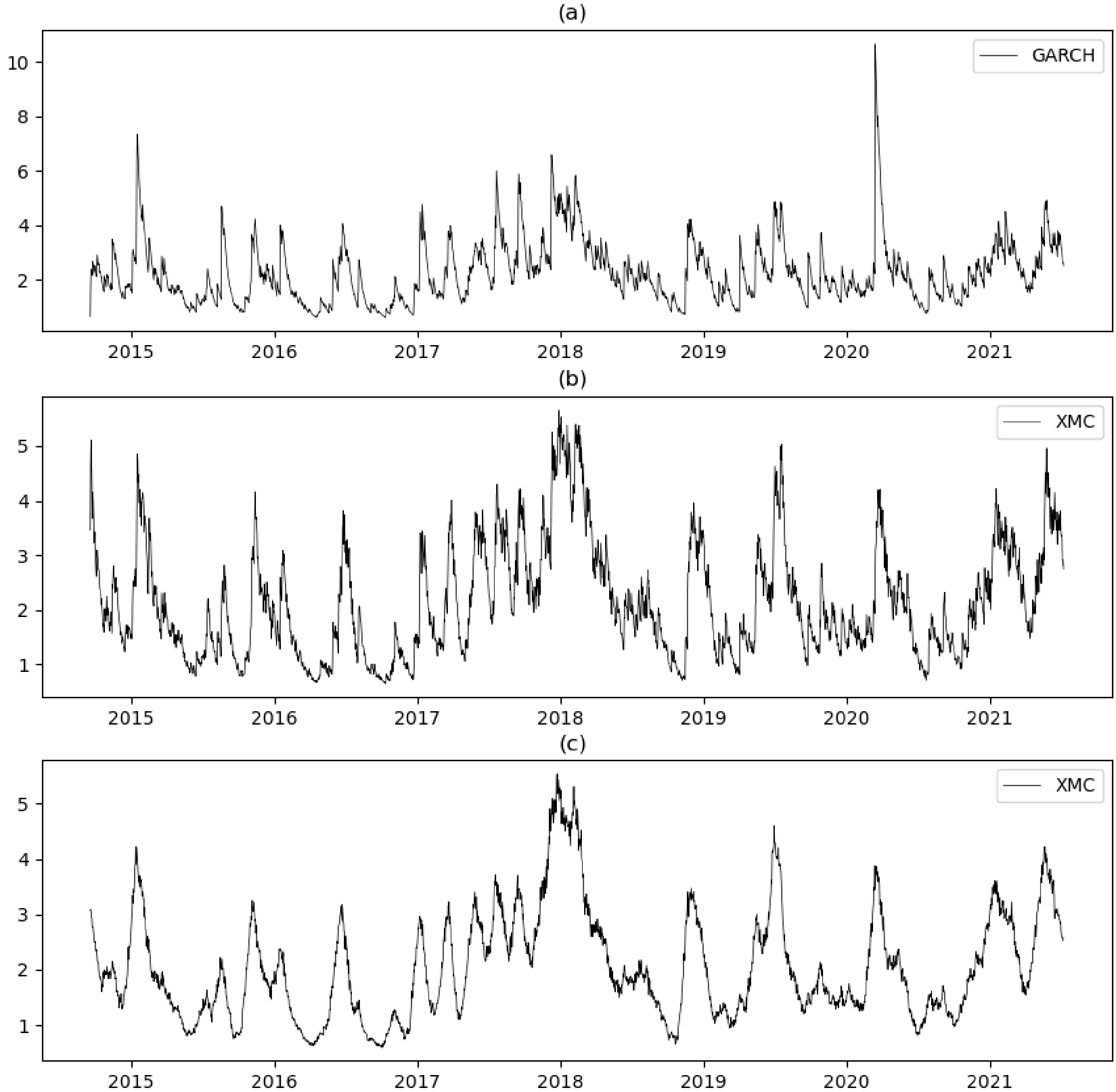

Figure 7: Estimated Bitcoin volatility $v_t = \exp(\alpha_t/2)$ for the GARCH and GB-XMC ($N = 10^5$) filters: (a) filtered GARCH volatility based on the estimated auxiliary model from Section 3; (b) XMC 1-period forecasts; (c) XMC FI smoothing.

As was shown in Figure 2, the XMC filter can be used to directly extract the volatility $v_t = \exp(\alpha_t/2)$ instead of the states by using the former as dependent variables in the regressions. For comparison, we used the estimated auxiliary GARCH model from Section 3 to filter the volatility using the update equation in (3). The GARCH predictions are shown in Figure 7 (a). It can be seen that the outlier at the start of 2020 caused the filtered volatlity to increase drastically. Figure 7 (b) shows the 1-period forecasts of v_t from the GB-XMC filter, which uses the same observations as the GARCH filter. The volatility estimates of the two filters are seen to have very similar movements, except that the impact of the outlier is limited on the XMC filter. This difference in robustness can be explained by the fact that the estimated stable distribution ($a = 1.677$) has fatter tails than the estimated t distribution ($\nu = 2.508$). Figure 7 (c) shows the FI smoothed estimates based on the GB-XMC filter. Compared with the forecasts in Part (b), most of the jagged movements have disappeared because the same conditioning set is used at all times. The enlarged conditioning set should generally result in improved estimates, and it is interesting to note that the impact of the outlier was estimated to be even more limited than before.

References

Azzalini, A., & Capitanio, A. (2003). Distributions generated by perturbation of symmetry with emphasis on a multivariate skew t-distribution. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *65*(2), 367–389.

Blasques, F., Koopman, S. J., & Moussa, K. (2022). Extremum Monte Carlo Filters: Signal Extraction by Simulation and Regression. *Working paper*.

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

Casarin, R. (2004). Bayesian inference for generalised markov switching stochastic volatility models. *CEREMADE Journal Working Paper*(0414).

Danielsson, J. (1994). Stochastic volatility in asset prices: estimation with simulated maximum likelihood. *Journal of Econometrics*, *61*, 375–400.

Durham, G. B. (2006). Monte carlo methods for estimating, smoothing, and filtering one-and two-factor stochastic volatility models. *Journal of Econometrics*, *133*(1), 273–305.

Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica*, *50*, 987–1007.

Fama, E. F. (1963). Mandelbrot and the stable paretian hypothesis. *The Journal of Business*, *36*(4), 420–429.

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

Friedman, J. H. (2001). Greedy function approximation: a gradient boosting machine. *Annals of Statistics*, 1189–1232.

Gallant, A. R., Hsieh, D., & Tauchen, G. (1997). Estimation of stochastic volatility models with diagnostics. *Journal of Econometrics*, *81*(1), 159–192.

Gallant, A. R., & Tauchen, G. (1998). Reprojecting partially observed systems with application to interest rate diffusions. *Journal of the American Statistical Association*, *93*(441), 10–24.

Garcia, R., Renault, E., & Veredas, D. (2011). Estimation of stable distributions by indirect inference. *Journal of Econometrics*, *161*(2), 325–337.

Gnedenko, B., & Kolmogorov, A. (1954). Limit distributions for sums of independent. *Am. J. Math*, *105*.

Gordon, N. J., Salmond, D. J., & Smith, A. F. (1993). Novel approach to nonlinear/non-gaussian bayesian state estimation. In *Iee proceedings f (radar and signal processing)* (Vol. 140, pp. 107–113).

Gourieroux, C., Monfort, A., & Renault, E. (1993). Indirect inference. *Journal of applied econometrics*, *8*(S1), S85–S118.

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

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

Kalman, R. E. (1960). A new approach to linear filtering and prediction problems. *Transactions of the ASME–Journal of Basic Engineering*, *82*(Series D), 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.

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

Lombardi, M. J., & Calzolari, G. (2008). Indirect estimation of α-stable distributions and processes. *The Econometrics Journal*, *11*(1), 193–208.

Lombardi, M. J., & Calzolari, G. (2009). Indirect estimation of α-stable stochastic volatility models. *Computational Statistics & Data Analysis*, *53*(6), 2298–2308.

Longstaff, F. A., & Schwartz, E. S. (2001). Valuing american options by simulation: a simple least-squares approach. *The Review of Financial Studies*, *14*(1), 113–147.

Mandelbrot, B. (1963a). New methods in statistical economics. *Journal of Political Economy*, *71*(5), 421–440.

Mandelbrot, B. (1963b). The variation of certain speculative prices. *The Journal of Business*, *36*(4), 394–419.

Mandelbrot, B. (1967). The variation of some other speculative prices. *The Journal of Business*, *40*(4), 393–413.

McNeil, A. J., Frey, R., & Embrechts, P. (2015). *Quantitative risk management: concepts, techniques and tools-revised edition.* Princeton university press.

Meinshausen, N. (2006). Quantile regression forests. *Journal of Machine Learning Research*, *7*(6).

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

Moussa, K., Blasques, F., & Koopman, S. J. (2023). Extremum Monte Carlo Smoothing. *(Forthcoming)*.

Nolan, J. P. (2009). Univariate stable distributions. *Stable Distributions: Models for Heavy Tailed Data*, *22*(1), 79–86.

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

Sandmann, G., & Koopman, S. J. (1998). Estimation of stochastic volatility models via monte carlo maximum likelihood. *Journal of Econometrics*, *87*(2), 271–301.

Shephard, N. (2005). *Stochastic volatility: Selected readings.* Oxford University Press.

Shephard, N., & Andersen, T. G. (2009). Stochastic volatility: origins and overview. In *Handbook of financial time series* (pp. 233–254). Springer.

Taylor, S. J. (1982). Financial returns modelled by the product of two stochastic processes-a study of the daily sugar prices 1961-75. *Time Series Analysis: Theory and Practice*, *1*, 203–226.

Taylor, S. J. (1986). *Modeling financial time series.* Chichester: Wiley.

Vankov, E. R., Guindani, M., & Ensor, K. B. (2019). Filtering and estimation for a class of stochastic volatility models with intractable likelihoods. *Bayesian Analysis*, *14*(1), 29–52.