TI 2023-077/III
Tinbergen Institute Discussion Paper

Asymmetric Stable Stochastic Volatility Models: Estimation, Filtering, and Forecasting

Francisco Blasques[1]
Siem Jan Koopman[2]
Karim Moussa[3]

1 Vrije Universiteit Amsterdam and Tinbergen Institute
2 Vrije Universiteit Amsterdam and Tinbergen Institute
3 Vrije Universiteit Amsterdam and Tinbergen Institute

Asymmetric Stable Stochastic Volatility Models: Estimation, Filtering, and Forecasting

Francisco Blasques, Siem Jan Koopman[*], Karim Moussa

Vrije Universiteit Amsterdam and Tinbergen Institute, the Netherlands

December 1, 2023

Abstract

This paper considers a stochastic volatility model featuring an asymmetric stable error distribution and a novel way of accounting for the leverage effect. We adopt simulation-based methods to address key challenges in parameter estimation, the filtering of time-varying volatility, and volatility forecasting. Specifically, we make use of the indirect inference method to estimate the static parameters, and the extremum Monte Carlo method to extract latent volatility. Both methods can be easily adapted to modifications of the model, such as having other distributions for the errors and other dynamic specifications for the volatility process. Illustrations are presented for a simulated dataset and for an empirical application to a time series of Bitcoin returns.

Some keywords: Filtering, Forecasting, Indirect Inference, Extremum Monte Carlo, Leverage, Bitcoin.

MSC Codes: 37M10, 62M10, 91B84

[*]Corresponding author. E-mail: s.j.koopman@vu.nl. An earlier version of this paper called "Stochastic Volatility with Stable Errors: Estimation, Filtering and Forecasting" was presented at the 2023 Financial Econometrics Conference at Lancaster University. We are grateful for the comments from Torben Andersen, Jean-Marie Dufour, Neil Shephard, Stephen Taylor, and other conference participants. Blasques thanks the Dutch Research Council (VI.Vidi.195.099) for financial support. Koopman acknowledges support from Aarhus University, Denmark, and funding of the Danish National Research Foundation (DNRF78). Work performed in partial fulfillment of the third author's PhD requirements at the Vrije Universiteit Amsterdam.

1 Introduction

In the analysis of financial time series, such as daily fluctuations in stock prices or exchange rates, return series are often leptokurtic and approximately serially uncorrelated, while squared returns or absolute returns are highly correlated (e.g., McNeil, Frey, & Embrechts, 2015, Ch. 3). These empirical findings are well documented, and they have led to a large body of related research. The initial attempt to account for excess kurtosis is due to Mandelbrot (1963), who replaced the (then standard) assumption of normally distributed returns by postulating that they follow the stable distribution instead. In addition to being fat-tailed for almost all of its parameterizations, the stable distribution has several important advantageous theoretical properties. In particular, the distribution is closed under summation, and it plays a pivotal role in the generalized central limit theorem (Gnedenko & Kolmogorov, 1954), which states that any limiting sum of independent and identically distributed (IID) variables must be stable.

Mandelbrot (1963) also noted that "*large changes tend to be followed by large changes, of either sign, and small changes tend to be followed by small changes.*" This phenomenon is referred to as "volatility clustering"; it has led to the (G)ARCH 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 temporal variability in the variance of a financial return series is known as the stochastic volatility (SV) model. The development of SV models has been initiated by the work of Stephen Taylor, in particular in Taylor (1982, 1986). The SV model has a strong theoretical foundation in the finance theory on option pricing (e.g., Hull & White, 1987). It has also been recognized 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 of parameters in SV models has been regarded as an interesting and challenging problem, for which many methods have been explored. For example, some proposed estimation methods are 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

(Gourieroux, Monfort, & Renault, 1993; Lombardi & Calzolari, 2009) and 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; Koopman, Lucas, & Scharth, 2015). A more complete review on SV models, including their estimation methods, are collected in the book of Shephard (2005) and the review chapter of Shephard and Andersen (2009).

In the initial development of the SV model, the observation error was assumed to be normally distributed. However, in later work it has been found that the Gaussian assumption is not sufficient to account for several aspects of the conditional distribution, such as excess kurtosis and skewness; see, for example, Gallant et al. (1997) and Durham (2006). Recently, the SV structure has been combined with the stable distribution for applications that are characterized by extreme movements. Some examples of applications of the stable SV model are the analysis of currency crises (Lombardi & Calzolari, 2009), exchange rates in general (Meintanis & Taufer, 2012), index returns (Sampaio & Morettin, 2020), and the modeling of electricity prices (Müller & Uhl, 2021). Furthermore, Vankov, Guindani, and Ensor (2019) stress the importance of asymmetry in the stable SV model, based on an application to weekly spot prices of propane.

While the stable distribution is rather flexible and has several theoretically appealing properties, its econometric treatment is challenging since its probability density function does not have a closed-form expression in general, and the associated second and higher-order moments do not exist. This precludes the direct use of standard estimators for the static parameters, and the same holds for methods that are used to estimate the time-varying volatility, including MCMC (Kim et al., 1998) and particle filters (Gordon, Salmond, & Smith, 1993; Pitt & Shephard, 1999).

This paper considers the asymmetric stable SV model, and it introduces an extension of the model to account for the leverage effect. A simulation-based approach is adopted to address the challenges of parameter estimation, filtering of the time-varying volatility, and its forecasting. An indirect inference estimator is proposed for the static parameters, while the extremum Monte Carlo method (Blasques, Koopman, & Moussa, 2023) is used to extract the latent volatility. These simulation-based methods can be adapted to different modifications of the model, such as the use of other error distributions or alternative

dynamics of the volatility. Two illustrations are presented. First, we discuss some details of the proposed method for a simulated time series. Second, we carry out an empirical study for a time series of Bitcoin returns.

The remainder of this paper is organized as follows. Section 2 introduces the asymmetric stable stochastic volatility model. Section 3 discusses parameter estimation by the indirect inference method. Section 4 describes how filtering and forecasting can be performed using the extremum Monte Carlo method. Section 5 applies the proposed methods in a study of daily returns of Bitcoin. Section 6 concludes.

2 Asymmetric stable stochastic volatility

2.1 The asymmetric stable stochastic volatility model

For a time series of financial returns $y_{1:T} = (y_1, \ldots, y_T)$, we consider the SV model with stable observation errors (Lombardi & Calzolari, 2009; Vankov et al., 2019) in (1),

$$\begin{aligned} y_t &= \exp(x_t/2)\varepsilon_t^y, & \varepsilon_t^y &\sim \mathrm{S}(\alpha, \beta), \\ x_{t+1} &= \mu_x + \phi_x(x_t - \mu_x) + \sigma_x \varepsilon_t^x, & \varepsilon_t^x &\sim \mathrm{N}(0, 1), \end{aligned} \tag{1}$$

for $t = 1, \ldots, T$, with x_t the unobserved log variance at time t, initialization $x_1 \sim \mathrm{N}(\mu_x, \sigma_x^2/(1-\phi_x^2))$, and static parameters $\mu_x \in \mathbb{R}$, $|\phi_x| < 1$, and $\sigma_x > 0$. Furthermore, $\mathrm{S}(\alpha, \beta)$ denotes the first parameterization of the standard univariate stable distribution as in Nolan (2009), with tail index parameter $\alpha \in (0, 2]$ and asymmetry parameter $\beta \in [-1, 1]$. The density of the stable distribution is known in closed form only in certain special cases, hence the characteristic function is used to describe the distribution:

$$\mathbb{E}[\exp(iu\varepsilon_t^y)] = \exp\left(-|u|^\alpha \left[1 - i\beta \tan(\frac{\pi\alpha}{2})(\operatorname{sgn} u)\right]\right) \qquad \text{if } \alpha \neq 1.$$

In addition to the normal distribution ($\alpha = 2$), the stable distribution contains the Cauchy distribution ($\alpha = 1, \beta = 0$) and the Levy distribution ($\alpha = 1/2, \beta = 1$) as special cases. Several approximations to the stable density exist (e.g., Nolan 1997; Menn and Rachev 2006), but these can become computationally expensive when many evaluations

are needed, and their accuracy varies with the parameter values. To circumvent the issue of limited tractability, we adopt a simulation-based approach to estimation and filtering, which exploits the fact that it is straightforward to draw stable variates via the method of Chambers, Mallows, and Stuck (1976); see also Weron (1996).

In addition to being closed under summation and its pivotal role in the generalized central limit theorem, the stable distribution is characterized by having fat tails for almost all its parameterizations. For $\alpha < 2$, the distribution has bounded moments only of order less than α. As the variance does not generally exist, the term "volatility" will be used to indicate the scale $\exp(x_t/2)$ of y_t. To ensure that the mean of the returns $\{y_t\}$ exists, we impose that $\alpha > 1$, which is a very mild restriction in the context of financial returns. In this case, $\mathbb{E}[\varepsilon_t^y] = 0$, such that the mean is not impacted by the tail and asymmetry parameters. This intuitive property also implies that the returns form a martingale difference sequence, which is consistent with the efficient markets hypothesis.

The above stable SV model is most suitable for applications that are characterized by extreme movements, several examples of which were given in the introduction. In cases thinner tails are more appropriate, one may consider instead a tempered version of the stable distribution (e.g., Barndorff-Nielsen & Shephard, 2001; Schoutens, 2003); alternatively, one may consider an altogether different distribution for the errors. Our proposed estimation and filtering methods are easily adapted to such model adjustments, and the same applies to modifications in the dynamics of the stable SV model. Examples of the latter are the Markov-switching variant of the above model due to Casarin (2004), the version with time-varying autoregressive parameters from Müller and Uhl (2021), and the extension of the model with leverage effect that is considered in the next section.

2.2 Leverage

A common empirical finding in financial returns is the asymmetric relationship between returns and volatility; see Yu (2005) and the references therein. This phenomenon is called the "leverage effect" due to Black (1976), who provided an explanation in terms of financial leverage. It is now widely recognized that other explanations may be more suitable (e.g., Figlewski & Wang, 2000; Hasanhodzic & Lo, 2019), but regardless of the

theoretical justification, there is wide consensus that incorporating a statistical leverage effect is crucial in many applications.

Introducing leverage in the stable SV model poses several challenges. The standard way to introduce leverage in the SV model is by correlating the errors ε_t^x and ε_t^y. However, when $\varepsilon_t^y \sim \mathrm{S}(\alpha, \beta)$ with $\alpha < 2$, the correlation $\rho = \mathrm{Corr}[\varepsilon_t^x, \varepsilon_t^y]$ is undefined because $\mathbb{V}\mathrm{ar}\,[\varepsilon_t^y] = \infty$. In this case, one can still use the equivalent representation of correlation with independent errors, in which the error ε_t^x is replaced by $\widetilde{\varepsilon}_t^x = \sigma_x(\rho\varepsilon_t^y + \sqrt{1-\rho^2}\varepsilon_t^x)$ in the update equation for x_t in (1). However, doing so would introduce fat tails into the log volatility process, which may not be desirable. Moreover, the above specification can be too simplistic in applications where a nonlinear relationship between x_{t+1} and ε_t^y is more suitable. For example, Yu (2012) finds that the use of a spline for the leverage function leads to improved fit and forecasting performance compared with the standard leverage model based on correlation.

The above issues can be addressed by extending the log volatility update equation to contain a bounded leverage function, $\Lambda(\varepsilon_t^y)$, that can exhibit nonlinearity. This yields the following stable SV model with leverage,

$$\begin{aligned} y_t &= \exp(x_t/2)\varepsilon_t^y, & \varepsilon_t^y &\sim \mathrm{S}(\alpha, \beta), \\ x_{t+1} &= \mu_x + \phi_x(x_t - \mu_x) + \sigma_x\varepsilon_t^x + \Lambda(\varepsilon_t^y), & \varepsilon_t^x &\sim \mathrm{N}(0,1). \end{aligned} \tag{2}$$

We illustrate our approach using the leverage function

$$\Lambda(\varepsilon_t^y) = c\tanh\left(d\varepsilon_t^y\right), \tag{3}$$

with parameters $c \in \mathbb{R}$ and $d > 0$ (for identifiability). The parameter d controls the shape of the function, while the parameter c determines the magnitude of the leverage effect, where we note that $|\Lambda(\varepsilon_t^y)| \leq |c|$. The specification in (3) is flexible as it allows for multiple forms of the leverage effect. This is illustrated by Figure 1, which plots the above leverage function for various values of the parameters c and d. In particular, it generalizes correlation as represented with independent errors, since $c \coloneqq \kappa/d$ yields $\lim_{d\to 0}\Lambda(\varepsilon_t^y) = \kappa\varepsilon_t^y$ for any $\kappa \in \mathbb{R}$. In addition, it contains the simple conditional mean adjustment as a special case, since $\lim_{d\to\infty}\Lambda(\varepsilon_t^y) = c\,\mathrm{sgn}\,(\varepsilon_t^y)$.

With respect to identification, parameter d is not identified when $c = 0$, while as $d \to \infty$, it becomes weakly identified because $\tanh(d\varepsilon_t^y)$ flattens out for any value of ε_t^y. In these cases it is difficult to estimate d precisely, but fortunately, the same does not apply to the leverage function Λ as a whole, which will be the primary interest in most applications. Ideally, the leverage function is chosen such that $\mathbb{E}[\Lambda(\varepsilon_t^y)] = 0$ to disentangle its effect from the unconditional mean, which makes the parameters easier to interpret and prevents related numerical issues. We notice that if $\mathbb{E}[\Lambda(\varepsilon_t^y)] \neq 0$, the unconditional mean $\mathbb{E}[x_t] = \mu_x + \mathbb{E}[\Lambda(\varepsilon_t^y)]/(1 - \phi_x)$ diverges as ϕ_x approaches one. For values of ϕ_x close to one (which are common in applications), numerical overflows could then easily occur in the simulated paths of the log volatility. In practice, we shall therefore work with the centered leverage function $\overline{\Lambda}(\varepsilon_t^y) = \Lambda(\varepsilon_t^y) - \mathbb{E}[\Lambda(\varepsilon_t^y)]$, where the second term can be computed via Monte Carlo integration. Alternatively, for odd functions $\Lambda(\cdot)$ such as (3), one could assume $\beta = 0$ to ensure symmetry of the error density $p(\varepsilon_t^y)$. This approach is legitimate, but it has the drawback that the resulting SV model can either introduce *both* conditional and unconditional asymmetry in the returns or *neither*. In the applications we shall therefore use the centered leverage function based on (3).

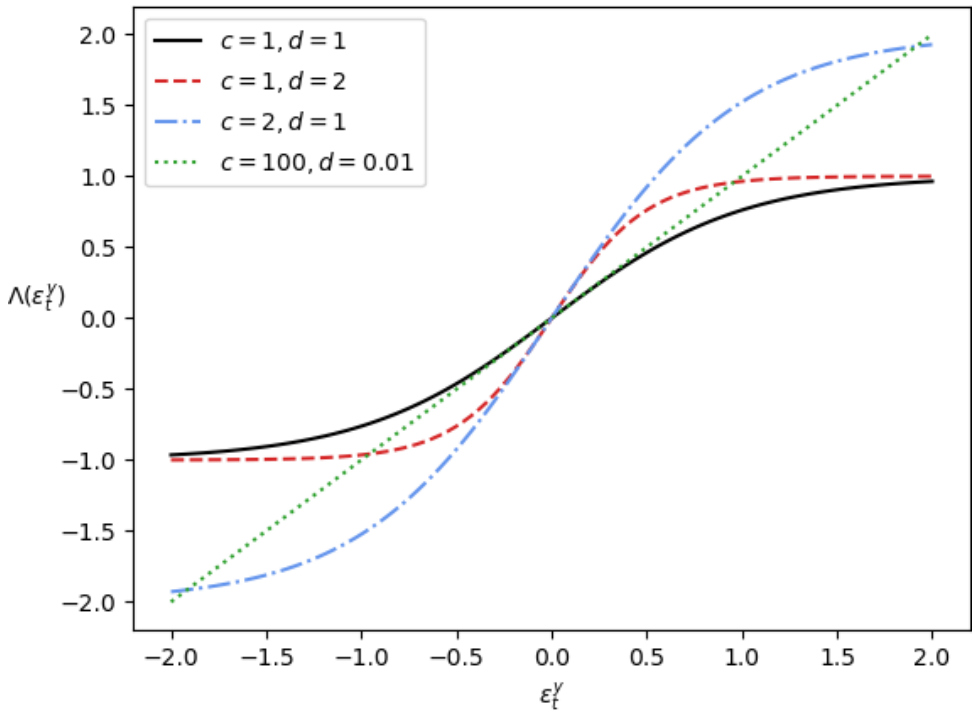

Figure 1: The leverage function $\Lambda(\varepsilon_t^y)$ in (3) for various values of the parameters c and d.

3 Parameter estimation

3.1 Previous literature

Estimation of the static parameters in the stable SV model is challenging due to the fact that probability density function does not have a closed-form expression in general, and its second and higher-order moments do not exist. Moreover, even if (an approximation to) the density can be used, computation of the likelihood remains complicated because it is defined as a high-dimensional integral with respect to the random log volatilities. Several methods have been considered to circumvent the above issues. In the Bayesian setting, Casarin (2004) introduces an auxiliary variable to perform MCMC (see Buckle, 1995), Vankov et al. (2019) use approximate Bayesian computation, and Müller and Uhl (2021) use a Gaussian mixture approximation to the stable distribution. In the classical setting, Meintanis and Taufer (2012) consider a different variant of the stable SV model that allows them to exploit the joint characteristic function of the returns, while Lombardi and Calzolari (2009) and Sampaio and Morettin (2020) use the method of indirect inference (II; Gourieroux et al., 1993) to estimate the parameters of models with symmetric stable errors. Our approach is related to the latter two studies, as we consider an II estimator for the *asymmetric* stable SV models from the previous section.

3.2 Indirect inference

The idea behind the II method is to estimate the parameter vector $\theta \in \Theta$ for some model of interest (the *structural* model) by simulating data for various candidate values of the parameters and choosing the estimate that makes the simulated data most "similar" to the observed data. The similarity is expressed in terms of an auxiliary parameter vector, $\psi \in \Psi$, which must contain at least as many parameters as θ for the purpose of identifiability. The auxiliary parameters form the arguments to a corresponding objective function $Q(\psi; y_{1:T})$ that is easy to evaluate; a common choice for Q is the (average) log likelihood of a similar model. In our case, the structural model corresponds to one of the stable SV models from Section 2, and examples of possible auxiliary models are the GARCH and exponential GARCH (EGARCH Nelson, 1991) models.

More specifically, the II estimator starts with estimating ψ by optimizing the auxiliary objective function based on the observed data:

$$\widehat{\psi} \in \arg\max_{\psi \in \Psi} Q(\psi; y_{1:T}). \tag{4}$$

Let $y_{1:T}^{(j)}(\theta),\ j = 1, \ldots, M$ denote $M \in \mathbb{N}$ paths that are simulated using the structural model with parameter vector θ, and let $\widehat{\psi}_M(\theta) \in \arg\max_{\psi \in \Psi} \sum_{j=1}^{M} Q(\psi; y_{1:T}^{(j)}(\theta))$ be the corresponding estimate of the auxiliary parameters. Then the II estimator is defined as (Gouriéroux & Monfort, 1996, Ch. 4.1.3)

$$\widehat{\theta}_{\text{II}} \in \arg\min_{\theta \in \Theta} \left[\widehat{\psi} - \widehat{\psi}_M(\theta)\right]' \Omega \left[\widehat{\psi} - \widehat{\psi}_M(\theta)\right],$$

where Ω is a positive semi-definite matrix. Because the above estimator corresponds to a nested optimization problem, it can be computationally intensive. We shall therefore use the *score-based* version of the II estimator (or efficient method of moments) due to Gallant and Tauchen (1996). Note first that the score $\partial Q / \partial \psi$ equals zero when evaluated at $\widehat{\psi}$ and $y_{1:T}$. The score-based II estimator attempts to make the score as small as possible when evaluated at $\widehat{\psi}$ and the simulated data. It is defined as

$$\widetilde{\theta}_{\text{II}} \in \arg\min_{\theta \in \Theta} \varsigma(\theta)' \Sigma \varsigma(\theta), \qquad \varsigma(\theta) = \frac{1}{M} \sum_{j=1}^{M} \frac{\partial Q\left(\psi; y_{1:T}^{(j)}(\theta)\right)}{\partial \psi}\bigg|_{\psi = \widehat{\psi}},$$

with Σ a positive semi-definite matrix. Contrary to the regular II estimator $\widehat{\theta}_{\text{II}}$, the score-based estimator $\widetilde{\theta}_{\text{II}}$ does not require nested optimizations. It is particularly attractive when (part of) the score can be computed in closed form.

Both versions of the estimator and their relationship simplify when the number of structural and auxiliary parameters are equal. Because the optimization problems are just-identified, the weighting matrices Ω and Σ become irrelevant and can therefore be replaced by the identity matrix. Moreover, for T sufficiently large it holds that $\widehat{\theta}_{\text{II}} = \widetilde{\theta}_{\text{II}}$ (Gouriéroux & Monfort, 1996, Proposition 4.1). Under appropriate regularity conditions, it can be shown that the II estimator is consistent and asymptotically normal for the true parameter, $\theta_0 \in \Theta$; see Proposition 4.2 in Gouriéroux and Monfort (1996). Two

necessary assumptions for consistency are that the auxiliary objective function converges to a deterministic limit function $Q_\infty(\psi, \theta) = \lim_{T\to\infty} Q(\psi; y_{1:T}(\theta))$, and that the *binding function* $\psi^*\colon \Theta \to \Psi$ defined as

$$\psi^*(\theta) = \arg\max_{\psi\in\Psi} Q_\infty(\psi, \theta)$$

is one-to-one. In practice it is therefore common to choose an auxiliary model that is similar to the structural model (but easier to estimate). It can generally be expected that this strategy results in binding functions that are one-to-one, although this can be difficult to verify rigoursly, since $\psi^*(\theta)$ is usually not available in closed form.

Several studies have applied the II method in the context of the stable distribution. For the stable distribution (including location and scale parameters) as structural model, Lombardi and Calzolari (2008) use the skew-t distribution of Azzalini and Capitanio (2003) as auxiliary model, while Garcia, Renault, and Veredas (2011) propose the skewed-t distribution of Fernández and Steel (1998). In a setting similar to ours, Lombardi and Calzolari (2009) and Sampaio and Morettin (2020) both use the GARCH model with Student's t errors as auxiliary model, which has the advantage that its likelihood function is easy to compute. However, in light of the consistency requirement on the binding function, it is more natural to choose an auxiliary model that, like the SV model, specifies the dynamics of the log volatility instead of the variance. This approach is adopted in the next section.

3.3 The auxiliary model

For the stable SV model in (1) as the structural model, we consider the following variant of the EGARCH model (Nelson, 1991) as the auxiliary model,

$$\begin{aligned} y_t &= \exp(z_t/2)u_t, & u_t &\sim \mathrm{CST}(s, \lambda), \\ z_{t+1} &= \mu_z + \phi_z(z_t - \mu_z) + \sigma_z(|u_t| - \mathbb{E}\,|u_t|), \end{aligned} \tag{5}$$

where $\mathrm{CST}(s, \lambda)$ denotes a centered version of the skew-t distribution from Azzalini and Capitanio (2003) with $\lambda > 0$ degrees of freedom and slant parameter $s \in \mathbb{R}$ which controls

the asymmetry. The auxiliary model excludes the leverage component of the EGARCH model, so that it has the same number of parameters as the structural model with similar interpretation. This is important in light of the required one-to-one property of the binding function for obtaining consistency of the II estimator. In addition, the similarity between parameters makes it straightforward to use the auxiliary estimates to determine reasonable starting values for the structural parameters in the optimization step.

As with the GARCH model, the likelihood of the auxiliary model is easy to compute, hence it can be used as an auxiliary objective function. The log specification for the volatility automatically ensures its positivity and simplifies the parameter restrictions. In line with the constraint $\alpha > 1$ in the structural model and the restrictions used in Lombardi and Calzolari (2009), we impose $\lambda > 1$ and $|\phi_z| < 1$, in addition to $\sigma_z > 0$. The parameter constraints are incorporated by reparameterizing the optimization problem, so that the optimization can be performed over unrestricted transformations of the parameters, such as $\widetilde{\sigma}_z \colon \sigma_z = \exp(\widetilde{\sigma}_z)$.

To discuss the CST distribution, we start by considering the skew-t distribution. Suppose X follows the standard skew normal distribution with slant parameter $s \in \mathbb{R}$, and V follows the chi-squared distribution with λ degrees of freedom, then

$$w = \frac{X}{\sqrt{V}} \sim \mathrm{ST}(s, \lambda)$$

follows the standard skew-t distribution with corresponding parameters s and λ as above (Azzalini & Capitanio, 2014, Ch. 4.3). In this case, w has density

$$p_{\mathrm{st}}(w; s, \lambda) = 2p_{\mathrm{t}}(w; \lambda) F_{\mathrm{t}}\left(s \cdot w \cdot \sqrt{\frac{\lambda + 1}{\lambda + w^2}}; \lambda + 1\right), \tag{6}$$

where p_{t} and F_{t} denote the PDF and CDF of the Student's t distribution, respectively. For $s = 0$ the ST distribution reduces to the Student's t distribution, while for $\lambda \to \infty$ it yields the skew normal distribution. An important advantage of the skew-t distribution is that its score exists and is available almost entirely in closed form (Azzalini & Capitanio, 2014, Ch. 4.3.3), which does not hold for the well-known asymmetric t distribution of Fernández and Steel (1998). As is common with asymmetric distributions, the shape

parameters s and λ impact the mean,

$$\mathbb{E}[w] \equiv m(s,\lambda) = \mu_s \cdot \mu_\lambda, \tag{7}$$

where

$$\mu_s = \frac{s}{\sqrt{1+s^2}} \qquad \text{and} \qquad \mu_\lambda = \frac{\sqrt{\lambda}\Gamma\left(\frac{1}{2}(\lambda-1)\right)}{\sqrt{\pi}\Gamma\left(\frac{1}{2}\lambda\right)}. \tag{8}$$

It follows that $m(s,\lambda) \neq 0$ if $s \neq 0$ for $\lambda > 1$, which is undesirable as it complicates interpretation of the shape parameters. Moreover, use of ST errors in the auxiliary model in (5) would imply that $\{y_t\}$ is not a martingale difference series, which is inconsistent with the efficient market hypothesis. Instead, we therefore work with the CST distribution, which is defined by the centered variates

$$u = w - m(s,\lambda) \sim \text{CST}(s,\lambda).$$

Because the auxiliary model in (5) has IID errors $\{u_t\}$ and $\{z_t\}$ is a Markov process, the joint density of the observations can be decomposed as

$$p(y_{1:T}) = \prod_{t=1}^{T} p(y_t|y_{1:t-1}) = \prod_{t=1}^{T} p(y_t|\widehat{z}_t),$$

where $y_{1:0} = \emptyset$ and the $\widehat{z}_t$ are the filtered log scales given by the recursion

$$\widehat{z}_{t+1} = \mu_z + \phi_z(\widehat{z}_t - \mu_z) + \sigma_z(|\widehat{u}_t| - \mathbb{E}\,|u_t|), \qquad \widehat{z}_1 = \mu_z, \tag{9}$$

with residuals $\widehat{u}_t = \exp(-\widehat{z}_t/2)y_t$. Since $u_t \sim \text{CST}(s,\lambda)$, the log likelihood for parameter vector $\psi = (\mu_z, \phi_z, \sigma_z, s, \lambda)'$ is

$$\mathcal{L}(\psi; y_{1:T}) = \sum_{t=1}^{T} \ell(\psi; y_t), \qquad \ell(\psi; y_t) = \ell_{\text{st}}(-\widehat{\omega}_t m(s,\lambda), \widehat{\omega}_t, s, \lambda; y_t), \tag{10}$$

with filtered volatilities $\widehat{\omega}_t = \exp(\widehat{z}_t/2)$, and with $\ell_{\text{st}}(\xi, \omega, s, \lambda; y)$ the log likelihood of the skew-t distribution with location ξ and scale ω for a single observation y, which can be found in Lemma 2 of Appendix A.

The above log likelihood can be used as auxiliary objective function in the regular II estimator, but as discussed in the previous section, it is computationally preferable to use the score-based variant of II. To show that the score of the auxiliary model exists, we consider the two potentially problematic terms with absolute values in (9). First,

$$|\widehat{u}_t| = |\exp(-\widehat{z}_t/2)y_t| = \exp(-\widehat{z}_t/2) \cdot |y_t|$$

by absolute homogeneity, which is differentiable in $\widehat{z}_t$. Second, the existence of $d\,\mathbb{E}\,|u_t|/d\psi_i$ is established in the following result, which provides a corresponding expression.

Lemma 1 (Derivatives of absolute mean for CST distribution)**.** *Let $u \sim \mathrm{CST}(s,\lambda)$, with $s \in (\underline{s},\bar{s})$ for $-\infty < \underline{s} < \bar{s} < \infty$ and $\lambda \in (\underline{\lambda},\bar{\lambda})$ for $1 < \underline{\lambda} < \bar{\lambda} < \infty$. Then*

$$\frac{d}{d\psi_i}\mathbb{E}\,|u| = \begin{cases} \mathbb{E}\left[|w-m| \cdot \frac{\partial \ell_{\mathrm{st}}}{\partial \psi_i}\right] + 2(F_{\mathrm{st}}(m) - 1/2)\frac{\partial m}{\partial \psi_i} & \text{if } \psi_i \in \{s,\lambda\}, \\ 0 & \text{otherwise,} \end{cases}$$

with $w \sim \mathrm{ST}(s,\lambda)$, $\ell_{\mathrm{st}} = \ell_{\mathrm{st}}(0,1,s,\lambda;w)$, F_{st} is the CDF of the standard skew-t distribution, $m = m(s,\lambda)$ as in (7), and

$$\begin{aligned} \frac{\partial m}{\partial s} &= (1+s^2)^{-3/2}\mu_\lambda, \\ \frac{\partial m}{\partial \lambda} &= \mu_s \frac{\Gamma(\frac{\lambda-1}{2})\left(\lambda\Psi_0(\frac{\lambda-1}{2}) - \lambda\Psi_0(\frac{\lambda}{2}) + 1\right)}{2\sqrt{\pi}\sqrt{\lambda}\Gamma(\frac{\lambda}{2})}, \end{aligned} \tag{11}$$

where μ_λ and μ_s are given in (8) and Ψ_0 is the digamma function.

Proof. See Appendix B. □

In the above result, the bounds on the parameters are arbitrary, apart from the restrictions $-\infty < \underline{s} < \bar{s} < \infty$ and $1 < \underline{\lambda} < \bar{\lambda} < \infty$. The assumption that $s \in (\underline{s},\bar{s})$ and $\lambda \in (\underline{\lambda},\bar{\lambda})$ is therefore of little impact in practice. The result allows us to derive the score of the auxiliary model, which is given in Appendix C. For computational purposes, we shall use the corresponding score-based variant of II as estimator of the structural parameters in the applications.

3.4 Adapting the auxiliary model

The auxiliary model in (5) is easily adapted to adjustments to the structural model. For example, if a different distribution is assumed for the errors ε_t^y, and the density $p(\varepsilon_t^y)$ is tractable (e.g., Student's t), one can assume

$$u_t \sim p(\varepsilon_t^y).$$

The same applies for modifications to the dynamics of the structural model. For the SV model with leverage in (2), we can use the following auxiliary model,

$$\begin{aligned} y_t &= \exp(z_t/2)u_t, & u_t &\sim \mathrm{CST}(s,\lambda), \\ z_{t+1} &= \mu_z + \phi_z(z_t - \mu_z) + \sigma_z(|u_t| - \mathbb{E}|u_t|) + \Lambda(u_t), \end{aligned} \tag{12}$$

which automatically accommodates the chosen leverage function Λ in the structural model. This is an important advantage over the GARCH-t model as auxiliary model, which would require switching to a suitable specification with leverage, the choice of which is not immediate. Moreover, the resulting model would likely not have the same degree of similarity with the structural model as (12).

4 Filtering and forecasting

Without a closed-form density for the observation errors, most standard methods for extracting the time-varying volatility cannot be used directly. For example, both MCMC methods and particle filters need to evaluate $p(y_t|x_t) = p(\varepsilon_t^y)$. In some cases it is possible to circumvent this issue by introducing an auxiliary variable. For example, if we limit ourselves to observation errors from the symmetric stable distribution, $\varepsilon_t^y \sim \mathrm{S}(\alpha, 0)$, the well-known stochastic representation $\varepsilon_t^y = ZV$ with normal variate $Z \sim \mathrm{S}(2, 0)$ and $V \sim \mathrm{S}(\alpha/2, 0)$ allows for exploiting the fact that ε_t^y is Gaussian conditional on V. Lombardi and Godsill (2006) use this approach to apply particle filtering methods to extract the time-varying parameters in a model with stable measurement noise. To extract the latent volatility in the more general asymmetric case, we propose to use the recently developed extremum Monte Carlo (XMC) method (Blasques et al. 2023; BKM

hereafter), which is explained in the following section.

4.1 Extremum Monte Carlo filtering

The XMC method combines simulation and regression techniques to estimate the time-varying conditional means, quantiles, modes, and other features of interest. The method relies on a technique that originates from the least squares Monte Carlo method (Longstaff & Schwartz, 2001) that was developed for the valuation of American-style options and related financial derivatives. Given two random variables, say, X and Y, the latter method estimates the conditional expectation *function* $\mathbb{E}[X|Y]$ by drawing N variates of X and 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., observed data) to provide the approximation $\widehat{f}(y) \approx \mathbb{E}[X|Y = y]$. The XMC method exploits this technique to filter unobserved signals, by applying the above for $t = 1, \ldots, T$ with $X = x_t$ and $Y = \widetilde{Y}_t \subseteq y_{1:t}$, with $\widetilde{Y}_t$ a suitable subset of the available observations.

Algorithm 1 presents the XMC filtering method for extracting the log volatility x_t for the SV models in (1) and (2). The method starts by using the SV model of choice to simulate N paths of the log volatility and observations. Next, at every time $t = 1, \ldots, T$, the log volatility variates are regressed on a subset of the available observations. This subset is used, rather than all available observations, to prevent overfitting of the training sample. More specifically, for a given "window size" parameter $W \in \mathbb{N}$, we define the set of covariates by

$$\widetilde{Y}_t^{(i)} = y_{\underset{\sim}{t}:t}^{(i)}, \qquad \text{with} \qquad \underset{\sim}{t} = \max\{t - W + 1, 1\},$$

such that it consists of the W observations nearest to time t. Together with the other tuning parameters of the chosen regression method, the window size is selected from a set of candidate values as the minimizer of the average loss in (13) incurred on a separate validation sample of simulated data. Lastly, the log volatilities are predicted by evaluating the estimated regression functions at the observed data $\widetilde{Y}_t$ (distinguishable from the simulated data by the lack of superscript) for $t = 1, \ldots, T$.

The algorithm requires a choice of function space $\mathbb{F}_N$ and loss function L, the combina-

Algorithm 1 Extremum Monte Carlo filtering method for stochastic volatility models.

1. **Simulate**: Simulate N paths of the log volatility x_t and observations y_t,

$$x_{1:T}^{(i)},\ y_{1:T}^{(i)}, \qquad i = 1, \ldots, N,$$

with $T \in \mathbb{N}$ the length of the observed data $y_{1:T}$.

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

$$\widehat{f}_t^N \in \underset{f \in \mathbb{F}_N}{\arg\min} \frac{1}{N} \sum_{i=1}^{N} L\left[x_t^{(i)} - f\left(\widetilde{Y}_t^{(i)}\right)\right], \tag{13}$$

where the covariates $\widetilde{Y}_t^{(i)} \subseteq y_{1:t}^{(i)}$ are a subset of the observations available at time t, and with $\mathbb{F}_N$ denoting a function space and L a loss function of choice.

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

$$\widehat{x}_t = \widehat{f}_t^N(\widetilde{Y}_t).$$

tion of which characterizes the adopted regression method. The function space represents a tradeoff between estimation error (a "large" $\mathbb{F}_N$) and misspecification error (a "small" $\mathbb{F}_N$). The loss function, on the other hand, determines the desired estimate of the log volatility. For example, it is well known that the mean squared error loss is minimized by the conditional expectation function; in our setting, this yields the filtering means $\mathbb{E}[x_t|y_{1:t}]$ for $t = 1, \ldots, T$. Moreover, by using other loss functions, the XMC method allows for estimating different aspects of the conditional distributions of interest, such as the quantiles (tilted absolute error loss) or the modes (all-or-nothing loss).

The combination of simulation and regression results offers a large degree of flexibility, which makes it easy to extend Algorithm 1 in several ways. In addition to filtering the log volatility, the XMC method can be used to directly extract the volatility itself,

$$\sigma_t = \exp(x_t/2),$$

just by using the latter as a dependent variable in the regressions. Moreover, by choosing a conditioning set different from $y_{1:t}$, it can be used to perform fixed-interval smoothing (conditioning set: $y_{1:T}$) or k-period forecasting (conditioning set: $y_{1:t-k}$, $k \in \mathbb{N}$). In

particular, by using y_t as dependent variable, it is possible to forecast the observations and determine corresponding probability intervals. The common issue of missing data can also be handled by adjusting the conditioning set, namely by omitting the appropriate covariates. Furthermore, multivariate SV models, such as the ones proposed by Harvey et al. (1994), are directly accommodated by noting that if the volatility $x_t = (x_{1,t}, \ldots, x_{N_x,t})'$ is a vector, Algorithm 1 can be performed separately for each element $x_{j,t},\ j = 1, \ldots, N_x$. Lastly, for lengthy time series, that is large T, which are common in financial studies, substantial computational savings can be obtained by re-using the estimated regression functions at other time points. This "steady state" approach ensures that the regression step only has to be performed for part of the times $t = 1, \ldots, T$. Further discussion of these issues and other aspects of the XMC filtering method can be found in BKM.

4.2 Filtering illustrations

This section makes use of simulated data to illustrate the performance of the XMC filter for estimating the filtering expectations $\mathbb{E}[x_t|y_{1:t}]$ for $t = 1, \ldots, T$. The simulated data is generated from the stable SV model in (1) for two different parameter choices. The first choice corresponds to the Gaussian SV model ($\alpha = 2, \beta = 0$), with the other parameters set to $\mu_x = 0,\ \phi_x = 0.96$, and $\sigma_x = 0.16$. These parameters are based on the maximum likelihood estimates obtained by Sandmann and Koopman (1998) for a time series of S&P500 daily log returns. The resulting SV model was used to simulate a path of the states $x_{1:T}$ and observations $y_{1:T}$ of length $T = 100$, which is shown in Figure 2. The Gaussian case is convenient because the density is available in closed form, so that we can directly apply a particle filter for comparison. Figure 2 (b) shows the filtered states from the bootstrap filter (BF; Gordon et al., 1993), a popular version of the particle filter, represented by the orange, solid line.

To select the appropriate XMC filter, note that because x_t is a transformation of the scale of y_t, the filtering expectations are nonlinear in the observations. A nonlinear regression method is therefore appropriate, hence we consider an XMC filter based on the tree-based gradient boosting method (GB; Friedman, 2001) combined with the squared error loss as in BKM. To ensure that the estimates of the filtering expectations are highly

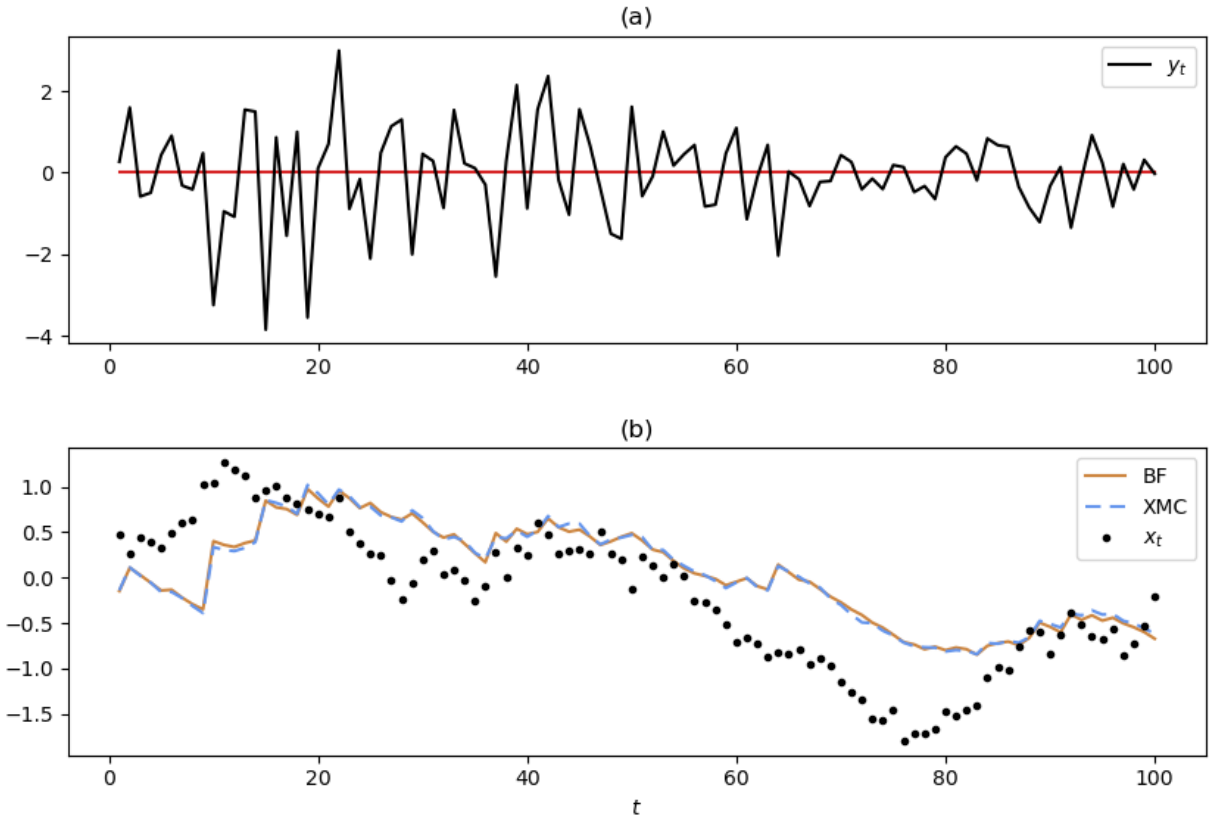

Figure 2: Analysis of a simulated path from the Gaussian version ($\alpha = 2, \beta = 0$) of the stable SV model in (1), with the other parameters set to $\mu_x = 0$, $\phi_x = 0.96$, and $\sigma_x = 0.16$ as in Sandmann and Koopman (1998): (a) simulated observations; (b) true states and filtered states by the bootstrap filter (BF) and gradient boosting (GB) version of the XMC filter. Both filters are based on 10^6 draws (i.e., particles and simulated paths).

accurate, both the BF and the XMC filter are based on 10^6 draws (i.e., particles and simulated paths). Figure 2 (b) shows the filtered states from the XMC filter (the blue, dashed line). As expected, the estimates of the two filters are seen to be close, and both roughly follow the movements of the true states, which are relatively high at the beginning and lower near the end of the sample.

As a second illustration, we repeat the above using the following parameters from Vankov et al. (2019): $\mu_x = -0.2$, $\phi_x = 0.95$, $\sigma_x = 0.2$, $\alpha = 1.75$, and $\beta = 0.1$. The simulated paths of the observations and states are shown in Figure 3. In this case, $\beta \neq 0$, which means that the particle filter cannot be used directly, nor via the approach of Lombardi and Godsill (2006). By contrast, the XMC filter remains applicable because the only model-specific knowledge it requires are simulated paths of the states and observations. The filtered states by the XMC filter (blue, dashed line) are shown in Figure 3

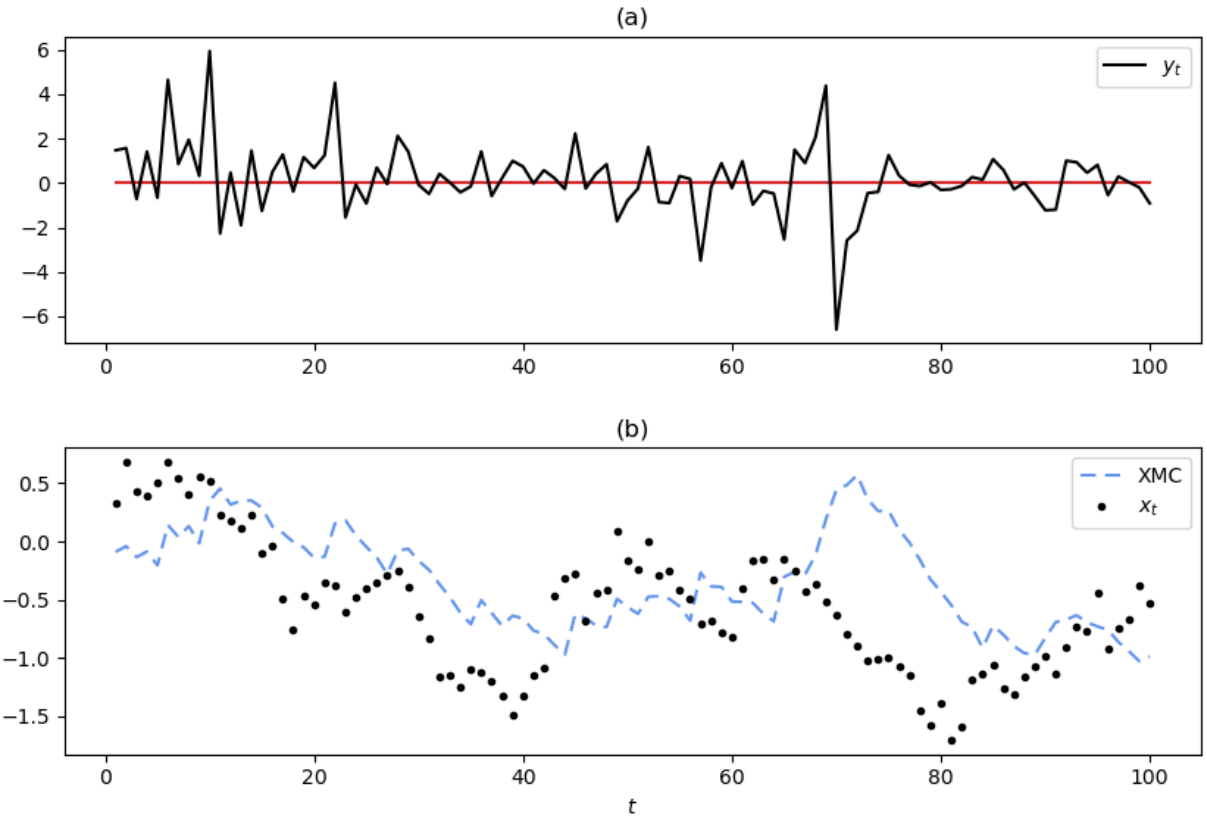

Figure 3: Analysis of a simulated path from the stable SV model in (1) with $\mu_x = -0.2$, $\phi_x = 0.95$, $\sigma_x = 0.2$, $\alpha = 1.75$ and $\beta = 0.1$ as in Vankov et al. (2019): (a) simulated observations; (b) true states and filtered states by the gradient boosting (GB) version of the XMC filter based on 10^6 draws.

(b). As with the Gaussian illustration, the filtered states follow a similar pattern as the true states, except for the period around $t = 70$ where several extreme observations are concentrated. However, it is noticeable that the impact of extreme observations on the filter are limited, a property that can be attributed to the assumed stable distribution for the observation errors.

Lastly, we consider a simulation study based on the stable SV model in (1) with the second parameter choice. This simulation study extends the one in BKM by considering 1-period forecasting and smoothing. The SV model is used to generate a test sample of 10^5 paths of length $T = 100$. For each path, the GB-XMC filter ($N = 10^5$) uses the observations to filter the states x_t. In addition to filtering, we also consider 1-period forecasting of the states with conditioning sets $y_{1:t-1}$, as well as (fixed-interval) smoothing with conditioning sets $y_{1:T}$. For comparison, we consider the QML filter from Harvey et al. (1994), which remains applicable without a tractable observation density. The QML

Table 1: Results from simulation study with the stable SV model in (1): overall root mean squared error (RMSE) for estimates of x_t by the QML filter and the gradient boosting version of the XMC filter ($N = 10^5$) using 1-period forecasting, filtering, and (fixed-interval) smoothing. The results are based on a sample of 10^5 simulated paths from the stable SV model using the following parameters from Vankov et al. (2019): $\mu_x = -0.2$, $\phi_x = 0.95$, $\sigma_x = 0.2$, $\alpha = 1.75$, and $\beta = 0.1$.

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

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,

$$\begin{aligned} \widetilde{y}_t &= x_t + 2\widetilde{\varepsilon}_t^y, \\ x_{t+1} &= (1-\phi_x)\mu_x + \phi_x x_t + \sigma_x \varepsilon_t^x, \end{aligned}$$

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 x_t. This approach uses the mean and variance of $\widetilde{\varepsilon}_t$, which are given by Lemma 3.19 of Nolan (2009) for $\alpha \neq 1$:

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

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

$$c(\alpha,\beta) = \alpha^{-1}\arctan\left[\beta\tan\left(\frac{\pi\alpha}{2}\right)\right].$$

Table 1 presents the overall root mean squared error (RMSE) for both methods. The XMC filter outperforms the QML filter for each choice of conditioning set. The difference in predictive performance is underlined by the fact that the 1-period forecasts of the XMC filter outperform the filtered QML predictions, although the former are based on a smaller conditioning set that excludes the most informative element, y_t.

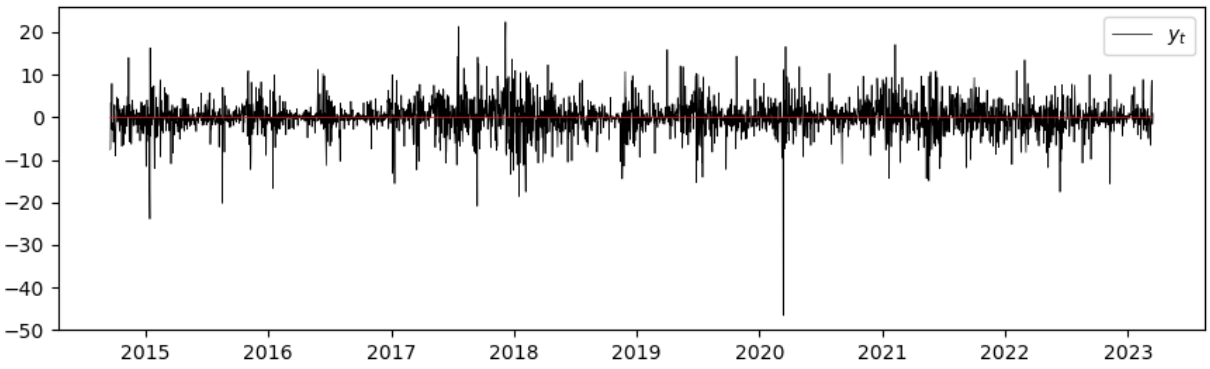

Figure 4: Centered daily log returns of the BTCUSD exchange rate (Bitcoin) multiplied by 100, starting from September 18$^{\text{th}}$, 2014, to March 16$^{\text{th}}$, 2023 ($T = 3102$). Source: Yahoo Finance.

5 Empirical study of Bitcoin returns

In this section, we apply the stable SV models and adopt the methods from previous sections to study a time series of daily Bitcoin returns (obtained from *Yahoo Finance*). As for other cryptocurrencies, Bitcoin has recently gained much interest from the scientific community (e.g., Hafner 2020; Makarov and Schoar 2020; Liu, Tsyvinski, and Wu 2022; Biais, Bisiere, Bouvard, Casamatta, and Menkveld 2023). From a financial point of view, cryptocurrencies are interesting because data on every transaction is publicly available, while from an econometric perspective they are interesting due to the frequent occurrence of extreme movements (Härdle, Harvey, & Reule, 2020).

Figure 4 shows the centered daily log returns of Bitcoin (multiplied by 100), starting from September 18$^{\text{th}}$, 2014, to March 16$^{\text{th}}$, 2023 ($T = 3102$). The data show clear signs of volatility clustering, and absolute returns exceeding 10 percent are often observed. It is noticeable that such movements are also common in the relatively calm periods. The combination of these features makes this data set an interesting application for the stable SV models from Section 2.

5.1 Parameter estimation

We apply the proposed II estimator to the log returns of Bitcoin shown in Figure 4. Table 2 shows the parameter estimates for the stable SV model in (1) and the auxiliary EGARCH-ST model in (5), where the latter are obtained by the method of maximum

likelihood. As expected, the estimates of the structural and auxiliary models are in close agreement overall. The estimates of the asymmetry parameters β and s are slightly negative, but in both cases they are not significant at the 10% level. The autocorrelation coefficients ϕ_x and ϕ_z for the log scale are both close to one, which indicates high persistence in the log volatility. The estimates for the unconditional mean of the log volatility are also similar, but contrary to μ_z, the estimate of μ_x is not significant. This can be explained by noting that, in addition to the models being different, the maximum likelihood estimator is generally expected to be more efficient than the II estimator, which results in a smaller standard error.

The estimated tail parameters $\alpha = 1.766$ and $\lambda = 2.602$ reflect the frequent extreme movements that are shown in Figure 4, where we note that as both the stable and skew-t distributions are power laws, their densities satisfy $\lim_{|v|\to\infty} p(v) \propto v^{-(k+1)}$ with corresponding tail index $k \in \{\alpha, \lambda\}$. However, the difference in these estimates is not surprising, as the above limiting behavior is but one aspect of the density that is controlled by these shape parameters. In this light, it should be noted that any estimate is based on data that ranges over a *finite* domain, for which the above limiting behavior may not be relevant; see also Fofack and Nolan (1999), who find that the point at which the limiting formula becomes a good approximation to the stable density depends heavily on the parameter values and can be very large for values of α close to two. Moreover, the larger estimate for the skew-t parameter was also to be expected because for a substantial part of the support, the $t(2)$ density is much heavier tailed than the corresponding stable density for α near 2; see the corresponding discussion in Lombardi and Calzolari (2008,

Table 2: Parameter estimates for the stable SV model in (1) and the auxiliary EGARCH-ST model in (5) based on the daily log returns of Bitcoin from Figure 4. The estimates are obtained by the method of II for the SV model, and by maximum likelihood for the EGARCH-ST model. Asterisks indicate significance relative to zero at the 10% (*), 5% (**), and 1%(***) level, respectively.

		Log scale		*Shape*	
Model	Mean	Autocorr.	Scale	Tail	Asymmetry
EGARCH-ST	μ_z	ϕ_z	σ_z	λ	s
	1.780^{***}	0.987^{***}	0.142^{***}	2.602^{***}	-0.052
Stable SV	μ_x	ϕ_x	σ_x	α	β
	2.092	0.992^{***}	0.288^{**}	1.766^{***}	-0.064

pp. 201-202),

Table 3 shows the parameter estimates for the stable SV model with leverage in (2) and the corresponding EGARCH-ST model with leverage in (12). Overall, the estimates of the structural and auxiliary models are again in close agreement. The results are also in line with those for the models without leverage in Table 2, with a somewhat larger estimate of μ_x and a smaller (but still insignificant) estimate of β being the most salient changes. The estimates of the shape parameter d are both large, which implies that the leverage effect is best approximated by a step function. Both estimates are insignicant at the 10% level; as noted in Section 2.2, it can be difficult to obtain precise estimates of d because the leverage function in (3) quickly becomes similar for large values of this parameter. However, this issue is of limited concern because our primary interest is the estimation of the leverage effect, $\Lambda(\cdot)$, which is not liable to the same issue. For the parameter c, both estimates are positive and relatively small, as the magnitude amounts to less than 1.9% (stable SV) and 4.2% (EGARCH-ST) of the estimated unconditional mean. The positivity of these estimates indicates that volatility tends to increase slightly after a rise in Bitcoin, although the statistical insignificance of the stable SV estimate makes it difficult to draw this inference convincingly.

5.2 Volatility filtering

We consider several applications of the XMC filter to extract the time-varying volatility from the daily log returns of Bitcoin shown in Figure 4. Here we focus on the stable SV

Table 3: Parameter estimates for the stable SV model with leverage in (2) and the auxiliary EGARCH-ST model with leverage in (12) based on the daily log returns of Bitcoin from Figure 4. The estimates are obtained by the method of II for the SV model, and by maximum likelihood for the EGARCH-ST model. Asterisks indicate significance relative to zero at the 10% (*), 5% (**), and 1%(***) level, respectively.

		Log scale		*Shape*		*Leverage*	
Model	Mean	Autocorr.	Scale	Tail	Asymmetry	Magnitude	Shape
EGARCH-ST	μ_z	ϕ_z	σ_z	λ	s	c	d
	1.792^{***}	0.989^{***}	0.122^{***}	2.596^{***}	-0.074	0.076^{***}	22.377
Stable SV	μ_x	ϕ_x	σ_x	α	β	c	d
	3.234^{*}	0.992^{***}	0.292^{**}	1.734^{***}	-0.194	0.061	1935.8

model without the leverage component. The static parameters are set to the estimates in Table 2.

Figure 5 presents the log volatility that is filtered from the returns by the QML filter (brown, solid line) and the GB-XMC filter (blue, dashed line) with $N = 10^5$. The log volatility is extracted via estimates of the filtering means $\mathbb{E}[x_t|y_{1:t}]$ for $t = 1, \ldots, T$. The estimates of both methods are close and tend to move in the same direction as new observations come available. The 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. This can be explained by the fact that stable observation errors do not require a large value of the scale to make extreme observations likely.

The XMC filter can also be used to directly extract the volatility, $\sigma_t = \exp(x_t/2)$, by using it as dependent variable in the regressions. For comparison, the EGARCH-ST filter in (9) was used to extract the corresponding volatility $\exp(z_t/2)$, which is shown in Figure 6 (a). It reveals that the extreme observation at the start of 2020 resulted in a drastic increase of the filtered volatility. Figure 6 (b) shows the 1-period forecasts of σ_t from the XMC filter, which are based on the past observations $y_{1:t-1}$ like the EGARCH filter. Apart from the difference in initialization, the volatility estimates by the two filters have very similar movements. However, the impact of the extreme observation on the XMC filter is limited, which can be explained by the fact that estimated stable distribution ($\alpha = 1.766$) has thicker tails than the estimated skew-t distribution ($\lambda = 2.602$). In addition, Figure 6

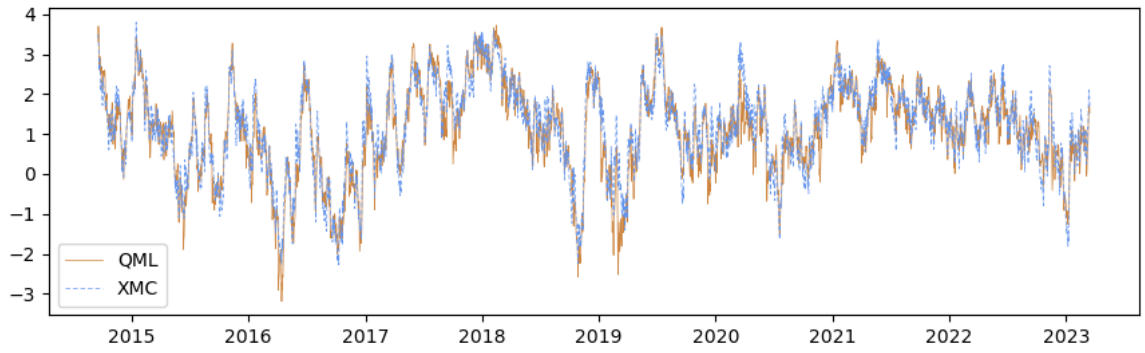

Figure 5: Log volatility filtered from the daily Bitcoin returns by the QML filter (brown, solid line) and the GB-XMC filter (blue, dashed line) based on $N = 10^5$ draws. The log volatility is extracted via estimates of $\mathbb{E}[x_t|y_{1:t}]$ for $t = 1, \ldots, T$.

(c) shows the smoothed volatilities by the XMC filter, which are based on the smoothing means $\mathbb{E}[\sigma_t|y_{1:T}]$ for $t = 1, \ldots, T$. When compared with the forecasts presented in Figure 6 (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 noticeable that the impact of the extreme observation is even smaller

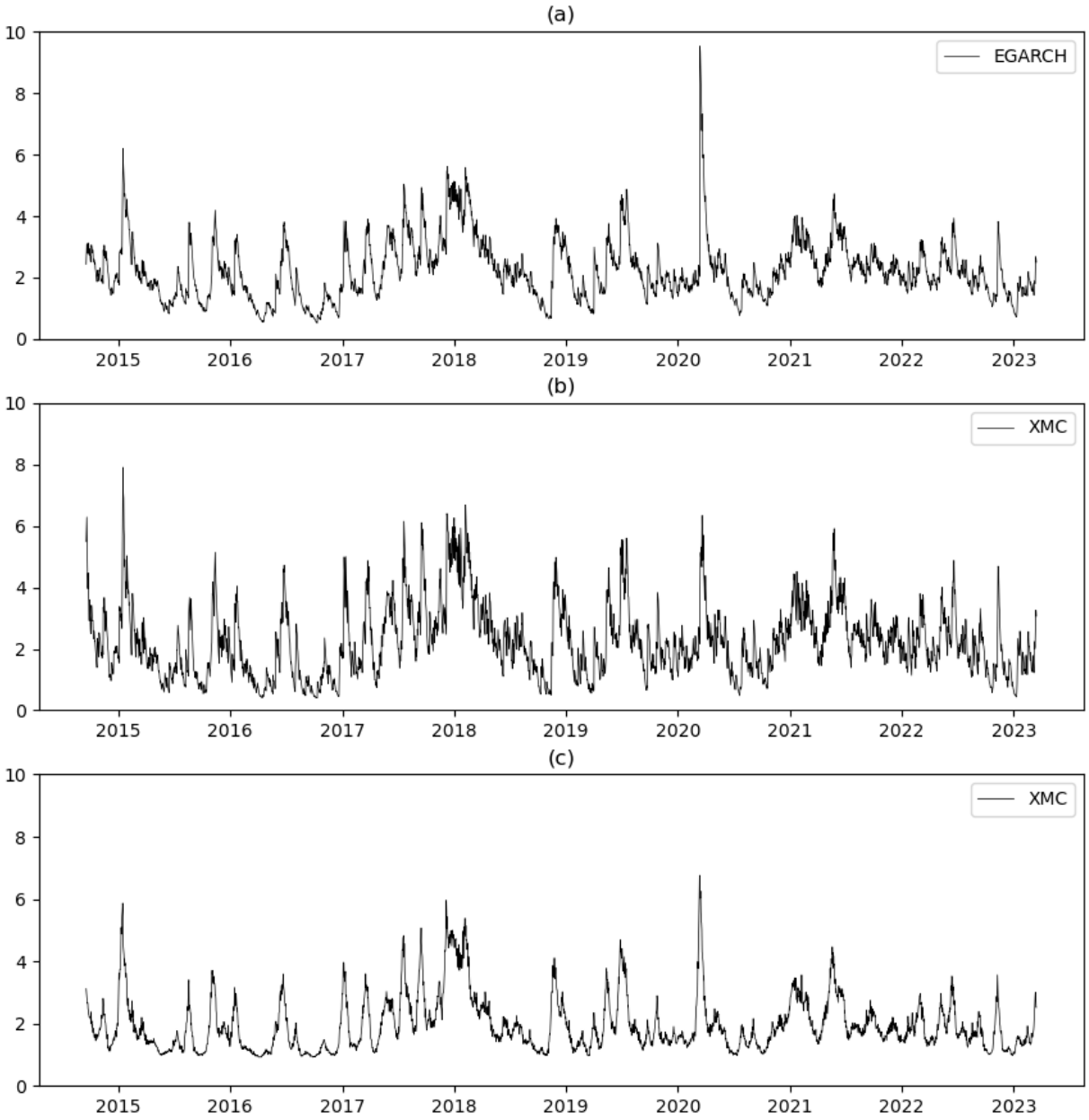

Figure 6: Estimated Bitcoin volatility $\sigma_t = \exp(x_t/2)$ for the EGARCH-ST and GB-XMC filters: (a) filtered EGARCH volatility $\exp(\widehat{z}_t/2)$ based on the corresponding filter for z_t in (9); (b) XMC 1-period forecasts $\mathbb{E}[\sigma_t|y_{1:t-1}]$ ($N = 10^6$); (c) XMC smoothed volatilities $\mathbb{E}[\sigma_t|y_{1:T}]$ ($N = 10^5$).

than before. The incorporation of future returns leads to a lower estimate of the volatility, which is as one would expect based on Figure 4.

6 Conclusion

In this paper, we have considered the stochastic volatility model with an asymmetric stable error distribution and considered an extension of the model to account for potential leverage effects. An indirect inference estimator is proposed for estimating the static parameters, while the extremum Monte Carlo method is used for filtering and forecasting. Both methods are easily adapted to modifications of the model, such as the use of other error distributions and alternative dynamics of the volatility. Illustrations are presented for simulated data and for an empirical application to a time series of Bitcoin returns. The application to Bitcoin illustrates that the extracted volatility based on the stable SV model is particularly robust to extreme observations, especially when compared to the EGARCH-ST filter. In this context, an important advantage of the XMC method is the ease with which it allows for the incorporation of additional observations. It is also indicated in the simulation study of Section 4.2 that the benefits of doing so can be substantial. The estimation results for the Bitcoin data indicate that, if there is asymmetry in the process for generating the daily returns of Bitcoin, it appears to be limited to a small positive leverage effect. The estimated leverage function corresponds to a mean adjustment, which implies that the volatility tends to go up whenever Bitcoin rises. Given that the evidence for a leverage effect is weak, further study is needed to obtain more conclusive results. It would be of interest to consider other cryptocurrencies to determine whether similar findings are obtained.

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.

Azzalini, A., & Capitanio, A. (2014). *The skew-normal and related families* (Vol. 3). Cambridge University Press.

Barndorff-Nielsen, O. E., & Shephard, N. (2001). *Normal modified stable processes*. Citeseer.

Biais, B., Bisiere, C., Bouvard, M., Casamatta, C., & Menkveld, A. J. (2023). Equilibrium bitcoin pricing. *The Journal of Finance*, *78*(2), 967–1014.

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

Blasques, F., Koopman, S. J., & Moussa, K. (2023). Extremum Monte Carlo filters: Real-time signal extraction via simulation and regression. *Discussion Paper Tinbergen Institute TI 2023-016/III*. Retrieved from https://papers.tinbergen.nl/23016.pdf

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

Buckle, D. (1995). Bayesian inference for stable distributions. *Journal of the American Statistical Association*, *90*(430), 605–613.

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

Chambers, J. M., Mallows, C. L., & Stuck, B. (1976). A method for simulating stable random variables. *Journal of the American Statistical Association*, *71*(354), 340–344.

Cheng, S. (2006). Differentiation under the integral sign with weak derivatives. *tech. report*.

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.

Fernández, C., & Steel, M. F. (1998). On bayesian modeling of fat tails and skewness. *Journal of the american statistical association*, *93*(441), 359–371.

Figlewski, S., & Wang, X. (2000). Is the'leverage effect'a leverage effect? *Available at SSRN 256109*.

Fofack, H., & Nolan, J. P. (1999). Tail behavior, modes and other characteristics of stable distributions. *Extremes*, *2*, 39–58.

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. (1996). Which moments to match? *Econometric theory*, *12*(4), 657–681.

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

Gnedenko, B. V., & Kolmogorov, A. N. (1954). *Limit distributions for sums of independent random variables*. Addison-Wesley Reading, MA.

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

Gouriéroux, C., & Monfort, A. (1996). *Simulation-based econometric methods*. Oxford University Press.

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

Hafner, C. M. (2020). Testing for bubbles in cryptocurrencies with time-varying volatility.

Journal of Financial Econometrics, *18*(2), 233–249.

Härdle, W. K., Harvey, C. R., & Reule, R. C. (2020). *Understanding cryptocurrencies* (Vol. 18) (No. 2). Oxford University Press.

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

Hasanhodzic, J., & Lo, A. W. (2019). On black's leverage effect in firms with no leverage. *The Journal of Portfolio Management*, *46*(1), 106–122.

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. 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.

Koenker, R. (2005). *Quantile regression* (Vol. 38). Cambridge university press.

Koopman, S. J., Lucas, A., & Scharth, M. (2015). Numerically accelerated importance sampling for nonlinear non-Gaussian state space models. *Journal of Business and Economic Statistics*, *33*(2), 114–127.

Liu, Y., Tsyvinski, A., & Wu, X. (2022). Common risk factors in cryptocurrency. *The Journal of Finance*, *77*(2), 1133–1177.

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.

Lombardi, M. J., & Godsill, S. J. (2006). On-line bayesian estimation of signals in symmetric/spl alpha/-stable noise. *IEEE Transactions on Signal Processing*, *54*(2), 775–779.

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.

Makarov, I., & Schoar, A. (2020). Trading and arbitrage in cryptocurrency markets.

Journal of Financial Economics, *135*(2), 293–319.

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

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

Meintanis, S. G., & Taufer, E. (2012). Inference procedures for stable-paretian stochastic volatility models. *Mathematical and Computer Modelling*, *55*(3-4), 1199–1212.

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

Menn, C., & Rachev, S. T. (2006). Calibrated fft-based density approximations for α-stable distributions. *Computational statistics & data analysis*, *50*(8), 1891–1904.

Müller, G., & Uhl, S. (2021). Estimation of time-varying autoregressive stochastic volatility models with stable innovations. *Statistics and Computing*, *31*(3), 36.

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

Nolan, J. P. (1997). Numerical calculation of stable densities and distribution functions. *Communications in statistics. Stochastic models*, *13*(4), 759–774.

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.

Sampaio, J. M., & Morettin, P. A. (2020). Stable randomized generalized autoregressive conditional heteroskedastic models. *Econometrics and Statistics*, *15*, 67–83.

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

Schoutens, W. (2003). *Lévy processes in finance: pricing financial derivatives*. Wiley Online Library.

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.

Weron, R. (1996). On the chambers-mallows-stuck method for simulating skewed stable random variables. *Statistics & probability letters*, *28*(2), 165–171.

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

Yu, J. (2012). A semiparametric stochastic volatility model. *Journal of Econometrics*, *167*(2), 473–482.

Appendix A Results for the skew-t distribution

The derivations in the following sections make use of the log likelihood and score of the ST distribution, which is given in Section 4.3.3 of Azzalini and Capitanio (2014). We collect this auxiliary result in the lemma below. Note that this result contains both *total* and *partial* derivatives. For a function $f\big(x, y(x)\big)$ with y a function of x, the total derivative is $df/dx = \partial f/\partial x + \partial f/\partial y \cdot \partial y/\partial x$, while the notation $\partial f/\partial x$ will be used to indicate a partial (or ordinary) derivative.

Lemma 2 (Log likelihood and score of the skew-t distribution)**.** *The log likelihood for a single observation y from the skew-t distribution with location ξ, scale ω, slant s, and degrees of freedom λ is*

$$\begin{aligned}\ell_{\mathrm{st}}(\xi, \omega, s, \lambda; y) = \text{constant} - \log\omega - \frac{1}{2}\log\lambda + \log\Gamma\left(\frac{1}{2}(\lambda+1)\right) \\ - \log\Gamma\left(\frac{1}{2}\lambda\right) - \frac{1}{2}(\lambda+1)\log\left(1 + \frac{\widetilde{y}^2}{\lambda}\right) + \log F_{\mathrm{t}}(w; \lambda+1),\end{aligned} \tag{14}$$

with

$$\widetilde{y} = \frac{y-\xi}{\omega}, \quad q = q(\widetilde{y}) = s\widetilde{y}r, \quad r = r(\widetilde{y}, \lambda) = \sqrt{\frac{\lambda+1}{\lambda+\widetilde{y}^2}}. \tag{15}$$

The corresponding score is

$$\begin{aligned}\frac{\partial\ell_{\mathrm{st}}}{\partial\xi} &= \frac{1}{\omega}\left(\widetilde{y}r^2 - \frac{s\lambda r h(q)}{\lambda+\widetilde{y}^2}\right), \\ \frac{\partial\ell_{\mathrm{st}}}{\partial\omega} &= \frac{1}{\omega}\left(-1 + (\widetilde{y}r)^2 - \frac{\lambda q h(q)}{\lambda+\widetilde{y}^2}\right), \\ \frac{\partial\ell_{\mathrm{st}}}{\partial s} &= \widetilde{y}r h(q), \\ \frac{\partial\ell_{\mathrm{st}}}{\partial\lambda} &= \frac{1}{2}\Bigg(\Psi_0\left(\frac{1}{2}\lambda+1\right) - \Psi_0\left(\frac{1}{2}\lambda\right) - \frac{2\lambda+1}{\lambda(\lambda+1)} - \log\left(1+\frac{\widetilde{y}^2}{\lambda}\right) \\ &\quad + \frac{(\widetilde{y}r)^2}{\lambda} + \frac{s\widetilde{y}(\widetilde{y}^2-1)h(q)}{(\lambda+\widetilde{y}^2)^2 r}\frac{g(\lambda)}{F_{\mathrm{t}}(q;\lambda+1)}\Bigg).\end{aligned} \tag{16}$$

with

$$h(q) = \frac{p_{\mathrm{t}}(q;\lambda+1)}{F_{\mathrm{t}}(q;\lambda+1)},$$

and

$$g(\lambda) = \frac{dF_{\mathrm{t}}(q(s\widetilde{y}r(\widetilde{y},\lambda));\lambda+1)}{d\lambda}$$
$$= \int_{-\infty}^{q} \left(\frac{(\lambda+2)x^2}{(\lambda+1)(\lambda+1+x^2)} - \log\left(1+\frac{x^2}{\lambda+1}\right)\right) p_{\mathrm{t}}(x;\lambda+1)dx. \tag{17}$$

Appendix B Proof of Lemma 1

We start by writing

$$\mathbb{E}\,|u| = \mathbb{E}\,|w-m| = \int_{-\infty}^{\infty} |w-m| p_{\mathrm{st}}(w;s,\lambda)dw =: K(m,s,\lambda),$$

where $m = m(s,\lambda)$ is given in (7), from which it is immediate that $d\,\mathbb{E}\,|u|/d\psi_i = 0$ for $\psi_i \notin \{s,\lambda\}$. For $\psi_i \in \{s,\lambda\}$,

$$\frac{d}{d\psi_i}\mathbb{E}\,|u| = \frac{\partial K}{\partial \psi_i} + \frac{\partial K}{\partial m}\frac{\partial m}{\partial \psi_i}.$$

Since K denotes the mean absolute error between w and m, a standard result from the quantile regression literature (e.g., Eq. (1.11) in Koenker, 2005) yields

$$\frac{\partial K}{\partial m} = 2(F_{\mathrm{st}}(m) - 1/2),$$

where the multiplication by 2 is needed to make the quantile loss equivalent to the absolute error loss. Furthermore, (11) follows from applying the chain and quotient rules to (7).

For $\partial K/\partial\psi_i$, write $K(m,s,\lambda) = \int_{-\infty}^{\infty}\kappa(w;s,\lambda)dw$ with integrand

$$\kappa(w;s,\lambda) = |w-m|\cdot p_{\mathrm{st}}(w;s,\lambda),$$

in which m is treated as a constant with respect to s and λ because it is an argument to K and we consider the *partial* derivative $\partial K/\partial\psi_i$. We will show that the derivative and integral can be interchanged,

$$\partial K/\partial\psi_i = \int_{-\infty}^{\infty}\frac{\partial}{\partial\psi_i}\kappa(w;s,\lambda)dw = \int_{-\infty}^{\infty}|w-m|\cdot\frac{\partial}{\partial\psi_i}p_{\mathrm{st}}(w;s,\lambda)dw,$$

and that the right-hand side is finite. A sufficient condition for the above interchange is (Cheng, 2006, Theorem A.1) that the following holds for all $s \in (\underline{s}, \bar{s})$ and $\lambda \in (\underline{\lambda}, \bar{\lambda})$:

1. $\mathbb{E}\,|u| = \int_{-\infty}^{\infty} \kappa(w; s, \lambda) dw$ exists.

2. The derivative $\partial\kappa(w; s, \lambda)/\partial\psi_i$ exists almost surely.

3. There exists an integrable function $H(w)$ such that

$$\left| \frac{\partial}{\partial \psi_i} \kappa(w; s, \lambda) \right| \leq H(w). \tag{18}$$

The first condition is satisfied because $\lambda > \underline{\lambda} > 1$. The second condition is satisfied because it follows from the product rule of differentation that

$$\frac{\partial \log p_{\text{st}}}{\partial \psi_i} = \frac{1}{p_{\text{st}}} \cdot \frac{\partial p_{\text{st}}}{\partial \psi_i} \qquad \Longleftrightarrow \qquad \frac{\partial p_{\text{st}}}{\partial \psi_i} = \frac{\partial \log p_{\text{st}}}{\partial \psi_i} \cdot p_{\text{st}}, \tag{19}$$

where the scores $\partial \log p_{\text{st}}/\partial\psi_i$ are given by (16), and the density p_{st} is given by (6).

To derive an integrable upper bound $H(w)$ as in (18) for the third condition, it will be convenient to consider the closures of the parameter domains,

$$C_s = [\underline{s}, \bar{s}] \qquad \text{and} \qquad C_\lambda = [\underline{\lambda}, \bar{\lambda}],$$

such that we can exploit the fact that continuous functions are bounded on compact sets. The bounds derived in this way will then also hold for the actual domains $(\underline{s}, \bar{s}) \subset C_s$ and $(\underline{\lambda}, \bar{\lambda}) \subset C_\lambda$. We consider the components on the right-hand side of (19) separately, starting with p_{st}:

$$\begin{aligned} p_{\text{st}}(w; s, \lambda) &= 2p_{\text{t}}(w; \lambda) F_{\text{t}}\left(sw\sqrt{\frac{\lambda+1}{\lambda+w^2}}; \lambda+1 \right) \\ &\leq 2p_{\text{t}}(w; \lambda) = 2\frac{\Gamma(\frac{\lambda+1}{2})}{\sqrt{\lambda\pi}\Gamma(\frac{\lambda}{2})} \left(1 + \frac{w^2}{\lambda}\right)^{-\frac{\lambda+1}{2}} \\ &\leq B_1 \left(1 + \frac{w^2}{\bar{\lambda}}\right)^{-\frac{\lambda+1}{2}} =: \bar{p}(w) = O\left(|w|^{-(\underline{\lambda}+1)}\right), \end{aligned}$$

with

$$B_1 = \max_{\lambda \in C_\lambda} \left(2 \frac{\Gamma(\frac{\lambda+1}{2})}{\sqrt{\lambda\pi}\Gamma(\frac{\lambda}{2})} \right) \in \mathbb{R}_+.$$

For the scores $\partial \ell_{\text{st}} / \partial \psi_i = \partial \log p_{\text{st}} / \partial \psi_i$, note that the terms $\widetilde{y}r$ and q are bounded for given values of s and λ, since by (15),

$$\begin{aligned} |\widetilde{y}r| &= \left| \widetilde{y} \sqrt{\frac{\lambda+1}{\lambda+\widetilde{y}^2}} \right| = \sqrt{\lambda+1} \left| \sqrt{\frac{\widetilde{y}^2}{\lambda+\widetilde{y}^2}} \right| \leq \sqrt{\lambda+1}, \\ |q| &= |s\widetilde{y}r| = |s| \cdot |\widetilde{y}r| \leq |s| \cdot \sqrt{\lambda+1}. \end{aligned} \tag{20}$$

Therefore,

$$|h(q)| \leq \bar{h} \coloneqq \max_{(s,\lambda) \in [\underline{s}, \bar{s}] \times [\underline{\lambda}, \bar{\lambda}]} h(q) < \infty,$$

where the upper bound $\bar{h}$ is finite because h is continuous in λ and q with

$$|q| \leq |\bar{s}| \cdot \sqrt{\bar{\lambda}+1}$$

by (20). It follows that

$$\frac{\partial \ell_{\text{st}}}{\partial s} = |\widetilde{y}r h(q)| = |\widetilde{y}r| \cdot |h(q)| \leq \sqrt{\lambda+1} \cdot |\bar{h}| \leq \sqrt{\bar{\lambda}+1} \cdot |\bar{h}| \in \mathbb{R}_+$$

where the bound for $\widetilde{y}r$ follows from (20). We conclude that (18) holds for $\psi_i = s$ with upper bound

$$H(w) = |w-m| \sqrt{\bar{\lambda}+1} \cdot |\bar{h}| \cdot \bar{p}(w) = O(|w|) \cdot O(1) \cdot O\big(|w|^{-(\underline{\lambda}+1)}\big) = O(|w|^{-\underline{\lambda}}),$$

which is integrable because $\underline{\lambda} > 1$.

For $\partial \ell_{\mathrm{st}}/\partial \lambda$, we have that

$$\begin{aligned}\left|\frac{\partial \ell_{\mathrm{st}}}{\partial \lambda}\right| &= \frac{1}{2}\left|\Psi_0\left(\frac{1}{2}\lambda+1\right)-\Psi_0\left(\frac{1}{2}\lambda\right)-\frac{2\lambda+1}{\lambda(\lambda+1)}-\log\left(1+\frac{\widetilde{y}^2}{\lambda}\right)\right.\\ &\quad\left.+\frac{(\widetilde{y}r)^2}{\lambda}+\frac{s\widetilde{y}(\widetilde{y}^2-1)h(q)}{(\lambda+\widetilde{y}^2)^2 r}\cdot\frac{g(\lambda)}{F_{\mathrm{t}}(q;\lambda+1)}\right|\\ &\leq \left|\Psi_0\left(\frac{1}{2}\lambda+1\right)-\Psi_0\left(\frac{1}{2}\lambda\right)-\frac{2\lambda+1}{\lambda(\lambda+1)}\right|+\left|\log\left(1+\frac{\widetilde{y}^2}{\lambda}\right)\right|\\ &\quad+\left|\frac{(\widetilde{y}r)^2}{\lambda}\right|+\left|\frac{s\widetilde{y}(\widetilde{y}^2-1)h(q)}{(\lambda+\widetilde{y}^2)^2 r}\right|\cdot\left|\frac{g(\lambda)}{F_{\mathrm{t}}(q;\lambda+1)}\right|\\ &\leq B_2+\log\left(1+\frac{\widetilde{y}^2}{\underline{\lambda}}\right)+2+\left|\frac{\bar{s}\widetilde{y}(\widetilde{y}^2-1)\bar{h}}{(\underline{\lambda}+\widetilde{y}^2)^{3/2}\sqrt{\underline{\lambda}+1}}\right| B_3 =: \widetilde{H}(\widetilde{y}),\end{aligned}$$

with bounds

$$\begin{aligned}B_2 &= \max_{\lambda\in[\underline{\lambda},\bar{\lambda}]}\left|\Psi_0\left(\frac{1}{2}\lambda+1\right)-\Psi_0\left(\frac{1}{2}\lambda\right)-\frac{2\lambda+1}{\lambda(\lambda+1)}\right|<\infty,\\ B_3 &= \max_{q\in\left[-\sqrt{\lambda+1},\sqrt{\lambda+1}\right]}\left|\frac{g(\lambda)}{F_{\mathrm{t}}(q;\lambda+1)}\right|<\infty,\end{aligned}$$

which are finite by the continuity of the functions that are maximized. Furthermore, it holds by (20) that for $\lambda > 1$,

$$\left|\frac{(\widetilde{y}r)^2}{\lambda}\right| \leq \frac{\lambda+1}{\lambda} \leq 2,$$

and

$$\left|\frac{s\widetilde{y}(\widetilde{y}^2-1)h(q)}{(\lambda+\widetilde{y}^2)^2 r}\right| = \left|\frac{s\widetilde{y}(\widetilde{y}^2-1)h(q)}{(\lambda+\widetilde{y}^2)^{3/2}\sqrt{\lambda+1}}\right| \leq \left|\frac{\bar{s}\widetilde{y}(\widetilde{y}^2-1)\bar{h}}{(\underline{\lambda}+\widetilde{y}^2)^{3/2}\sqrt{\underline{\lambda}+1}}\right| = \frac{O(\widetilde{y}^3)}{O(\widetilde{y}^3)} = O(1).$$

The complexity of the upper bound $\widetilde{H}$ follows as

$$\widetilde{H}(\widetilde{y}) = O(1)+O\big(\log(|\widetilde{y}|)\big)+O(1)+O(1)\cdot O(1) = O\big(\log(|\widetilde{y}|)\big),$$

and since $\widetilde{y}$ represents a standard skew-t variate by (15), the results above hold for $w = \widetilde{y}$. We conclude that (18) holds for $\psi_i = \lambda$ with upper bound

$$H(w) = |w-m|\widetilde{H}(w)\cdot\bar{p}(w),$$

which is integrable because $\widetilde{H}(w) = O\big(\log(|w|)\big) = O(|w|^{\delta})$ for any $\delta > 0$, such that

$$H(w) = O(|w|) \cdot O(|w|^{\delta}) \cdot O(|w|^{-(\underline{\lambda}+1)}) = O(|w|^{\delta-\underline{\lambda}}),$$

where $\delta - \underline{\lambda} < -1$ since $\underline{\lambda} > 1$ and $\delta > 0$ is arbitrarily small.

Appendix C Score of the auxiliary model

The score of the auxiliary model in (5) is given by

$$\frac{d}{d\psi}\mathcal{L}(\psi; y_{1:T}) = \sum_{t=1}^{T} \frac{d}{d\psi}\ell(\psi; y_t),$$
$$\frac{d\ell(\psi; y_t)}{d\psi_i} = \frac{\partial \ell_{\text{st}}}{\partial \psi_i} + \frac{\partial \ell_{\text{st}}}{\partial \xi}\Big|_{\xi=\xi_t} \cdot \frac{\partial \xi_t}{\partial \psi_i} + \frac{\partial \ell_{\text{st}}}{\partial \omega}\Big|_{\omega=\widehat{\omega}_t} \cdot \frac{\partial \widehat{\omega}_t}{\partial \psi_i},$$

with

$$\xi_t = -\widehat{\omega}_t m(s, \lambda) \qquad \text{and} \qquad \widehat{\omega}_t = \exp(\widehat{z}_t/2),$$

where $\widehat{z}_t$ is given by the recursion in (9). The term $\partial \ell_{\text{st}}/\partial \psi_i$ is given by (16) if $\psi_i \in \{s, \lambda\}$ and it equals zero if $\psi_i \in \{\mu_z, \phi_z, \sigma_z\}$. Furthermore,

$$\frac{\partial \xi_t}{\partial \psi_i} = \begin{cases} -\widehat{\omega}_t \frac{\partial m(s,\lambda)}{\partial \psi_i} - \frac{\partial \widehat{\omega}_t}{\partial \psi_i} m(s, \lambda) & \text{if } \psi_i \in \{s, \lambda\}, \\ -\frac{\partial \widehat{\omega}_t}{\partial \psi_i} m(s, \lambda) & \text{otherwise,} \end{cases}$$

where $\partial m/\partial \psi_i$ is given by (11) if $\psi_i \in \{s, \lambda\}$ and it equals zero if $\psi_i \in \{\mu_z, \phi_z, \sigma_z\}$. In addition,

$$\frac{\partial \widehat{\omega}_t}{\partial \psi_i} = \frac{\widehat{\omega}_t}{2} \cdot \frac{\partial \widehat{z}_t}{\partial \psi_i},$$

where it follows from (9) that the second component is given by the recursion

$$\frac{\partial \widehat{z}_t}{\partial \psi_i} = \frac{\partial \gamma(\psi)}{\partial \psi_i} + \frac{\partial \phi_z \widehat{z}_{t-1}}{\partial \psi_i} + \frac{\partial \sigma_z \exp\left(-\widehat{z}_{t-1}/2\right)}{\partial \psi_i}|y_{t-1}|, \tag{21}$$

with $\gamma(\psi) = (1-\phi_z)\mu_z - \sigma_z \mathbb{E}\,|u_t|$ and initialization

$$\frac{\partial \widehat{z}_1}{\partial \psi_i} = \begin{cases} 1 & \text{if } \psi_i = \mu_z, \\ 0 & \text{otherwise.} \end{cases}$$

The components are given by

$$\frac{\partial \phi_z \widehat{z}_{t-1}}{\partial \psi_i} = \begin{cases} \phi_z \frac{\partial \widehat{z}_{t-1}}{\partial \phi_z} + \widehat{z}_{t-1} & \text{if } \psi_i = \phi_z, \\ \phi_z \frac{\partial \widehat{z}_{t-1}}{\partial \psi_i} & \text{otherwise,} \end{cases}$$

$$\frac{\partial \sigma_z \exp\left(-\widehat{z}_{t-1}/2\right)}{\partial \psi_i} = \begin{cases} \sigma_z \frac{\partial \exp\left(-\widehat{z}_{t-1}/2\right)}{\partial \psi_i} + \exp\left(-\widehat{z}_{t-1}/2\right) & \text{if } \psi_i = \sigma_z, \\ \sigma_z \frac{\partial \exp\left(-\widehat{z}_{t-1}/2\right)}{\partial \psi_i} & \text{otherwise,} \end{cases}$$

$$\frac{\partial \exp\left(-\widehat{z}_{t-1}/2\right)}{\partial \psi_i} = \frac{\partial \exp\left(-\widehat{z}_{t-1}/2\right)}{\partial \widehat{z}_{t-1}} \frac{\partial \widehat{z}_{t-1}}{\partial \psi_i} = -\frac{1}{2} \exp\left(-\widehat{z}_{t-1}/2\right) \frac{\partial \widehat{z}_{t-1}}{\partial \psi_i},$$

$$\frac{\partial \gamma}{\partial \psi_i} = \begin{cases} 1-\phi_z & \text{if } \psi_i = \mu_z, \\ -\mu_z & \text{if } \psi_i = \phi_z, \\ -\mathbb{E}\,|u_t| & \text{if } \psi_i = \sigma_z, \\ -\sigma_z \frac{\partial \mathbb{E}\,|u_t|}{\partial s} & \text{if } \psi_i = s, \\ -\sigma_z \frac{\partial \mathbb{E}\,|u_t|}{\partial \lambda} & \text{if } \psi_i = \lambda. \end{cases}$$

The terms $\mathbb{E}\,|u_t|$, $\partial\,\mathbb{E}\,|u_t|/\partial s$, and $\partial\,\mathbb{E}\,|u_t|/\partial\lambda$ must be computed numerically. For example, Lemma 1 can be used to efficiently compute both $\mathbb{E}\,|u_t|$ and $\partial\,\mathbb{E}\,|u_t|/\partial s$ via simulation by using the same variates. The component $\partial\,\mathbb{E}\,|u_t|/\partial\lambda$ should not be evaluated this way because it would be too expensive due to the integral in (17). It is therefore more efficient to compute this element of the score vector by finite difference.