

TI 2026-082/II
Tinbergen Institute Discussion Paper

Stochastic Volatility Models with Stable Disturbances: Estimation via linearization

*Fernando Roders*¹
*Siem Jan Koopman*²
*Karim Moussa*³

¹ a.s.r. Asset Management

² Vrije Universiteit Amsterdam

³ Vrije Universiteit Amsterdam

Stochastic Volatility Models with Stable Disturbances: Estimation via linearization

Fernando Roders, Siem Jan Koopman^{§*} , Karim Moussa 

§ Vrije Universiteit Amsterdam and Tinbergen Institute, the Netherlands

September 30, 2026

Abstract

In this study, we propose to estimate the parameters of stochastic volatility models with stable disturbances using methods based on linearization. Specifically, we obtain a linear state space form of the model by taking the logarithm of the squared financial returns, which, together with analytical expressions for logarithmic and fractional lower order moments of the stable distribution, forms the basis of the proposed quasi-maximum likelihood and generalized method of moments estimators. The methods are straightforward to implement and do not require simulation. We provide details of the implementation and study the finite-sample performance of the two estimation methods in a Monte Carlo study. The empirical relevance is illustrated by studying model-implied volatility estimates from time series of daily returns from stock indices, exchange rates, and cryptocurrencies.

Keywords: α -stable distribution; fractional lower order moments; generalized method of moments; logarithmic moments; quasi-maximum likelihood.

MSC Codes: 60E07; 62F10; 62M10; 62M20.

*Corresponding author. E-mail: s.j.koopman@vu.nl. Fernando Roders is currently with a.s.r. Asset Management. The views expressed in this paper do not necessarily reflect those of his employer.

1 Introduction

1.1 The Stable Distribution

The stable distribution (e.g., Samoradnitsky & Taqqu 1994; Nolan 2020) has received considerable interest in statistics and related fields. The generalized central limit theorem (Gnedenko & Kolmogorov, 1954) characterizes it as the limit distribution for convergent sums of independent and identically distributed (IID) random variables. Since measurements can often be thought of as the aggregation of many independent effects, the stable distribution has been used in the analysis of a wide range of phenomena, including financial asset returns (Mandelbrot, 1963; McCulloch, 1996), network traffic in communications (Karasaridis & Hatzinakos, 2001), noise in sound recordings Lombardi & Godsill (2006), and the number of infections in a pandemic (Machado & Lopes, 2020). The distribution is frequently used in signal processing (Nikias & Shao, 1995; Nolan, 2020), where the latter reference also discusses various applications in econometrics, physics, biology, and other fields.

The stable distribution has several appealing theoretical properties, such as being closed under summation, and its heavy tails make it useful for applications where extreme values are common. However, its practical use has been limited by the absence of a closed-form probability density function (PDF) in general. Additionally, except for the special case of the normal distribution, the distribution does not have a bounded second moment, which presents difficulties for moment-based estimation. However, despite these challenges, many estimation procedures have been developed, with various approaches focusing on the time series setting (e.g., Qiou & Ravishanker, 1998; Lombardi & Godsill, 2006; Nolan & Ravishanker, 2009; Peters et al., 2011; Chou-Chen & Morettin, 2020). In particular, for the modeling of time-varying scales, Liu & Brorsen (1995) and Paoella (2016) consider GARCH models by specifying financial returns as conditionally stable. Furthermore, in recent years, the stable distribution is increasingly being used within stochastic volatility (SV) models, which we now discuss in more detail below.

1.2 Stochastic Volatility Models with Stable Disturbances

The use of stable disturbances in SV models was initiated by Casarin (2004), who considered the Bayesian analysis of a Markov-switching specification for the volatility. Other specifications of the SV model with stable disturbances were subsequently proposed by various authors for applications to data characterized by extreme movements. Examples include returns on exchange rates (Lombardi & Calzolari, 2009; Meintanis & Taufer, 2012; Calvet & Czellar, 2015), propane prices (Vankov et al., 2019), index returns (Sampaio &

Morettin, 2020), electricity prices (Müller & Uhl, 2021), Bitcoin (Blasques et al., 2025), and Tesla stock (Moussa & Koopman, 2026).

In general, the estimation of SV models is considered challenging, as the likelihood corresponds to a high-dimensional integral with respect to the unobserved volatilities; see the survey by Broto & Ruiz (2004) for a discussion of estimation methods. For specifications involving stable disturbances, the estimation task is further complicated by the intractability of the stable density and the limited number of moments being available. For this reason, the works cited above all rely on numerical approximations for estimation, with Monte Carlo simulation being the main technique used.

1.3 Contribution

We propose to estimate the parameters of SV models with stable disturbances using methods that rely on linearization and are straightforward to implement. Specifically, the linearization is implied by taking the logarithm of the squared financial returns, from which we obtain a linear state space form of the SV model. Importantly, linearization allows parameter estimation without resorting to the use of simulation methods (Broto & Ruiz, 2004, Sec.3.2). We consider two widely used estimation approaches and extend these to the current setting by using analytical expressions for the logarithmic and fractional lower order moments (FLOMs) of the stable distribution.

We first consider the quasi-maximum likelihood (QML) approach proposed by Harvey et al. (1994) for SV models with Gaussian and Student’s t disturbances. It adopts the Kalman filter to evaluate a quasi-Gaussian likelihood function which is used for parameter estimation. We extend this method for the SV model with symmetric stable disturbances and discuss how the approach can be extended to other model specifications. Notably, the approach also provides a straightforward procedure to extract the unobserved volatility.

The second approach is based on the generalized method of moments (GMM; Hansen, 1982), which was first used for SV models by Melino & Turnbull (1990) and extensively studied by Andersen & Sørensen (1996). GMM can provide a non-simulated benchmark for QML, and it also allows for straightforward computation of standard errors. Moreover, by incorporating signed moments and FLOMs, it allows for the estimation of parameters in an SV model with asymmetric stable disturbances.

The use of linearization for estimating SV models with stable disturbances was also considered by Müller & Uhl (2021), who rely on mixtures of normals to approximate the stable and exp-stable distributions for Bayesian inference. In contrast, this study adopts a classical approach, using analytical moments to circumvent numerical approximations, which results in simple estimators that do not require the use of simulation. These estimators can be used either directly, or as input for other comprehensive es-

estimation procedures such as Monte Carlo maximum likelihood or Markov chain Monte Carlo methods. In so doing, this study connects to the previous work on simulation-based estimation mentioned in the previous section. In addition, it contributes to the statistical and signal processing literature by building on established estimation methods that use analytical logarithmic moments of the stable distribution (Ma & Nikias, 1995a; Kuruoglu, 2001; Wang et al., 2015; Lévy Véhel et al., 2021) and extending this approach to the case of stochastic and time-varying scales. This extension may be useful in network communication and, more generally, in systems where human interactions have an important impact on the dynamics, given that clustering of social activity is well known to occur during special events (e.g., the football world cup). The SV model could provide a natural framework to model time-varying uncertainty in such settings where the stable distribution is often used.

The rest of this paper is organized as follows. Section 2 presents the SV model with stable disturbances, along with its linearized form. Section 3 discusses QML estimation of this model. Section 4 introduces the GMM estimators. Section 5 develops a simulation study to investigate the finite-sample performance of the estimators. Section 6 presents empirical applications of the proposed methods. Section 7 provides discussion and concludes the paper. The appendix contains proofs and supplementary material.

2 Stochastic Volatility with Stable Disturbances

2.1 Stochastic Volatility Model

We consider the SV model (Lombardi & Calzolari, 2009; Vankov et al., 2019) as

$$\begin{aligned} y_t &= \exp(h_t/2)\varepsilon_t, & \varepsilon_t &\sim \mathcal{S}(\alpha, \beta), \\ h_{t+1} &= \mu + \phi(h_t - \mu) + \eta_t, & \eta_t &\sim \mathcal{N}(0, \sigma^2), \end{aligned} \tag{1}$$

for discrete times $t = 1, \dots, T$. Here, y_t denotes the return on a financial asset at time t , the latent variable h_t determines the corresponding scale or volatility, $\exp(h_t/2)$, and the disturbances η_t and ε_t are assumed to be mutually and serially independent, as well as independent of the initial state h_1 . We assume $0 < \phi < 1$ for stationarity and initialize the model via $h_1 \sim \mathcal{N}(\mu, \sigma^2/(1 - \phi^2))$, where $\mu \in \mathbb{R}$ and $\sigma \geq 0$.

The observation disturbances, ε_t , follow the stable distribution, which is characterized by a tail index $\alpha \in (0, 2]$ and an asymmetry parameter $\beta \in [-1, 1]$. We use the first parameterization in Nolan (2020, Ch. 1.5), which is the most commonly used and has the

following characteristic function:

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

The stable distribution generalizes the normal distribution ($\alpha = 2$), the Cauchy distribution ($\alpha = 1, \beta = 0$), and the Lévy distribution ($\alpha = 0.5, \beta = 1$). However, apart from these special cases, the density is not available in closed form. Bounded moments $\mathbb{E}|\varepsilon_t|^p < \infty$ exist for $p < \alpha$, resulting in heavy-tailed shocks in the returns when $\alpha < 2$. Throughout, we will make the mild assumption that $\alpha > 1$ to ensure a finite mean. In this case, a convenient property of the adopted parameterization is that $\mathbb{E}[\varepsilon_t] = 0$, so that the mean is not affected by the tail and asymmetry parameters, which simplifies their interpretation. Furthermore, it ensures that the returns $\{y_t\}$ form a martingale difference sequence, which aligns with the efficient market hypothesis.

2.2 Linearized Form

For SV models, a challenging aspect is that the likelihood corresponds to a high-dimensional integral with respect to the path of the unobserved states, $h_{1:T}$. The limited tractability of the stable distribution further complicates the already challenging problem of parameter estimation. To simplify the estimation task, we apply the following data transformation (Harvey et al., 1994)

$$\tilde{y}_t = \log(y_t^2)$$

to obtain the linear state space model (SSM) given by

$$\begin{aligned} \tilde{y}_t &= \omega + h_t + \tilde{\varepsilon}_t, & \tilde{\varepsilon}_t &= \log \varepsilon_t^2 - \omega, \\ h_{t+1} &= \mu + \phi(h_t - \mu) + \sigma \eta_t, & \omega &= \mathbb{E} \log \varepsilon_t^2, \end{aligned} \quad (3)$$

where $\eta_t \sim \mathcal{N}(0, 1)$ as before. The transformed observation disturbance $\tilde{\varepsilon}_t$ is defined having a mean of zero, while its variance, along with the intercept ω , can be determined via the first two moments of the exp-stable distribution using Lemma 3.19 of Nolan (2020). For $\alpha \neq 1$, this yields

$$\omega = 2 \left(\gamma_{\text{Euler}} (1/\alpha - 1) - \frac{1}{\alpha} \log[\cos(\alpha\theta)] \right), \quad \text{Var}[\tilde{\varepsilon}_t] = 4 \left(\frac{\pi^2(1 + 2/\alpha^2)}{12} - \theta^2 \right), \quad (4)$$

where $\gamma_{\text{Euler}} \approx 0.577$ is Euler's constant and (Nolan, 2020, Eq.3.22) with

$$\theta = \theta(\alpha, \beta) = \frac{1}{\alpha} \arctan \left[\beta \tan \left(\frac{\pi\alpha}{2} \right) \right]. \quad (5)$$

The linearized form makes the analysis of the SV model with stable disturbances considerably more tractable. As such, it will be central in the estimation methods proposed in the following sections.

3 Quasi-Maximum Likelihood Estimation

The QML estimation method (Harvey et al., 1994) uses the Kalman filter to estimate the unobserved volatility and the parameters of the SV model based on its linearized form in (3). To ensure identifiability, we assume $\beta = 0$ for parameter estimation, as discussed further below. This corresponds to the SV model with symmetric stable disturbances considered by Lombardi & Calzolari (2009).

The QML method makes use of the fact that, for linear Gaussian SSMs, the Kalman filter can be used to efficiently compute the log-likelihood. The process in (3) is linear but non-Gaussian, due to the exp-stable disturbance $\tilde{\varepsilon}_t$. However, by treating the disturbance as if it were normally distributed, we obtain an approximate linear Gaussian model to which the Kalman filter can be applied. Using $\xi_t \approx \tilde{\varepsilon}_t$ with $\xi_t \sim \mathcal{N}(0, \text{Var}[\tilde{\varepsilon}_t])$, the resulting model in state space form (Harvey, 1990) is given by

$$\begin{aligned} \tilde{y}_t &= Z_t h_t + d_t + \xi_t, & \xi_t &\sim \mathcal{N}(0, H_t), \\ h_{t+1} &= T_t h_t + c_t + R_t \eta_t, & \eta_t &\sim \mathcal{N}(0, Q_t), \end{aligned} \quad (6)$$

with unobserved state h_t and system matrices $Z_t = 1$, $d_t = \omega$, $H_t = \text{Var}[\tilde{\varepsilon}_t]$, $T_t = \phi$, $c_t = \mu(1 - \phi)$, $R_t = \sigma$, and $Q_t = 1$.

Writing $\tilde{Y}_t = (\tilde{y}_1, \dots, \tilde{y}_t)$, the Gaussian log-likelihood based on the SSM in (6) is

$$\log L_G(\tilde{Y}_T) = -\frac{T}{2} \log 2\pi - \frac{1}{2} \sum_{t=1}^T \log |F_t| - \frac{1}{2} \sum_{t=1}^T v_t' F_t^{-1} v_t, \quad (7)$$

with prediction errors $v_t = \tilde{y}_t - \mathbb{E}[\tilde{y}_t | \tilde{Y}_{t-1}]$ and variances $F_t = \text{Var}[\tilde{y}_t | \tilde{Y}_{t-1}]$, where we define $\tilde{Y}_0 = \emptyset$ to yield the unconditional mean and variance at $t = 1$. The terms v_t and F_t can be computed recursively for $t = 1, \dots, T$ using the Kalman filter; we refer to Harvey (1990) for the recursions. The quasi-log likelihood in (7) can be maximized with respect to the model parameters. Under suitable regularity conditions, the resulting QML estimator is consistent and asymptotically normal (Harvey, 1990, Ch.4.5.5). In particular, the transformed disturbance $\tilde{\varepsilon}_t$ satisfies the requirement of having four bounded moments, and it can be shown to have bounded moments of all orders $p > 0$ (Nolan, 2020, p.131).

Regarding identifiability, note that the log-likelihood in (7) can equivalently be written in terms of the mean and autocovariance function of the transformed observations (Durbin

& Koopman, 2012, Ch.4.13), which are given by

$$\begin{aligned}\mathbb{E}[\tilde{y}_t] &= \omega + \mu, \\ \text{Var}(\tilde{y}_t) &= \frac{\sigma^2}{1 - \phi^2} + \text{Var}[\tilde{\varepsilon}_t], \\ \text{Cov}(\tilde{y}_t, \tilde{y}_{t+k}) &= \phi^k \frac{\sigma^2}{1 - \phi^2} \quad k \in \mathbb{N}.\end{aligned}\tag{8}$$

From Equations (4) and (8), it is seen that the parameters α and β enter the likelihood only through the mean and variance. Since the mean also contains μ , this leaves two equations to determine three unknowns, so the parameters cannot be identified under the general stable distribution. We therefore impose $\beta = 0$, yielding the symmetric stable distribution, under which the parameters are identifiable via the quasi-log likelihood.

To estimate the unobserved states, the Kalman filter for the SSM in (6) can be used to compute the expectations of h_t conditional on $\tilde{Y}_{t-1}$ (forecasting), $\tilde{Y}_t$ (filtering), or $\tilde{Y}_T$ (smoothing). This yields the minimum variance linear unbiased estimator of the states based on the transformed observations. Estimates of the volatility,

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

for $t = 1, \dots, T$, can then be obtained using corresponding conditional means and variances that are output from the Kalman filter or smoother:

$$\hat{\sigma}_{t|k} \equiv \mathbb{E}[\sigma_t | \tilde{Y}_k] = \exp\left(\mathbb{E}[h_t | \tilde{Y}_k]/2 + \text{Var}[h_t | \tilde{Y}_k]/8\right).\tag{9}$$

The variance term provides a bias-correction based on the log-normality of σ_t . This filtering and smoothing approach remains applicable with the general stable distribution ($\beta \neq 0$), since it is based on given model parameters. The next section presents a moment-based method to estimate the parameters in the corresponding SV model.

4 Generalized Method of Moments Estimation

This section presents GMM estimators for the SV model in Equation (1). We start by considering an estimator for the model in which the observation disturbances follow the symmetric stable distribution. The resulting ‘‘symmetric GMM’’ (S-GMM) estimator is then extended to accommodate the asymmetric stable distribution in Section 4.2.

4.1 Symmetric Stable Distribution ($\beta = 0$)

Let $\psi \in \mathbb{R}^{N_\psi}$ denote the parameter vector to be estimated, with $N_\psi \in \mathbb{N}$ its number of elements. In this section, we assume $\beta = 0$, so that $\psi = (\mu, \phi, \sigma, \alpha)'$ and $N_\psi = 4$. GMM estimation of ψ is based on a vector function $g(y_t, \psi): \mathbb{R} \times \mathbb{R}^{N_\psi} \rightarrow \mathbb{R}^M$ satisfying the $M \geq N_\psi$ moment (or orthogonality) conditions

$$\mathbb{E}[g(y_t, \psi_0)] = 0$$

when y_t is generated by the SV model with true parameter vector ψ_0 . The components of $g(y_t, \psi)$ are used to express distances between selected population moments and their sample analogues, and the GMM approach is designed to minimize these distances. Specifically, for a given choice of positive definite weighting matrix $\widehat{W}$ (discussed further below), the GMM estimator is defined by

$$\widehat{\psi}_{\text{GMM}} \in \arg \min_{\psi} \left(\frac{1}{T} \sum_{t=1}^T g(y_t, \psi) \right)' \widehat{W} \left(\frac{1}{T} \sum_{t=1}^T g(y_t, \psi) \right).$$

Under suitable regularity conditions (Hansen 1982; Hayashi 2000, Ch. 7), the GMM estimator is consistent and asymptotically normal, with asymptotic variance matrix

$$\text{Avar}(\widehat{\psi}_{\text{GMM}}) = (G'WG)^{-1}G'WSWG(G'WG)^{-1}, \quad (10)$$

where $W = \text{plim}_{T \rightarrow \infty} \widehat{W}$ is positive definite, $G = \mathbb{E}[\frac{\partial g(y_t, \psi_0)}{\partial \psi'}]$, and S is the long-run variance matrix of $\{g(y_t, \psi_0)\}$.

The S-GMM estimator for the SV model is based on the moments of $\tilde{y}_t = \log(y_t^2)$ in Equation (8), which provide a convenient way to circumvent the limited number of bounded moments and intractability of the stable distribution. The first component of g is given by

$$g_1(y_t, \psi) = \tilde{y}_t - \omega - \mu. \quad (11)$$

Together with the variance and $K \geq 2$ autocovariances in (8) for $k = 1, \dots, K$, this provides $M = K + 2 \geq 4$ moment conditions to estimate ψ .

What remains is to choose M (via K), as well as the weighting matrix $\widehat{W}$, both of which impact the performance of the estimator, and the effects are related. When the number of moment conditions equals the number of parameters to be estimated, the weighting matrix is irrelevant. However, Ruiz (1994) and Andersen & Sørensen (1996) find that the just-identified estimator performs poorly for the standard SV model with Gaussian disturbances, suggesting that $M > N_\psi$ should be used. In the over-identified

case, the choice of the weighting matrix impacts the efficiency of the estimator, as can be seen from the asymptotic variance matrix in (10). In particular, asymptotic optimality is achieved when $\text{plim } \widehat{W} = S^{-1}$, which yields the simplified expression $(G'S^{-1}G)^{-1}$ for the asymptotic variance matrix. However, in finite samples, other choices may perform better, noting that the asymptotically optimal choice requires the estimation of $M(M+1)/2$ elements of W using a sample of size T . This matter is discussed in detail by Andersen & Sørensen (1996), who note that M represents a trade-off where additional moments provide more information but also complicate the estimation of the weighting matrix.

In our initial simulation experiments, we observe that using the identity matrix as weighting matrix provided better results than the two-step and iterated approaches based on the asymptotically optimal choice. This is particularly evident for larger values of M , which typically yielded better performance. Using this choice for the weighting matrix, we select $K = 25$ ($M = 27$) by optimizing the accuracy of the S-GMM estimator on simulated data, based on several data generating processes (DGPs) and sample sizes, over a set of candidate tuning parameters. Further details of this procedure are provided in the Appendix B.

Lastly, we note that in our case, as for diagonal weighting matrices more generally, it is optimal to estimate μ via

$$\widehat{\mu}_{\text{GMM}} = \frac{1}{T} \sum_{t=1}^T \widetilde{y}_t - \omega,$$

since this parameter enters the GMM objective function only via g_1 in (11). This reduces the dimension of the optimization problem by one.

4.2 Asymmetric Stable Distribution ($\beta \neq 0$)

Extending the GMM estimator from the previous section to estimate β requires an additional moment condition that depends on α , β , or both. We consider two possible choices, which make use of FLOMs or signed moments of the stable distribution (Nolan, 2020, Ch.3.7), which have previously been used for estimation in the static setting (e.g., Ma & Nikias, 1995b; Kuruoglu, 2001). These moments of the stable distribution are given below.

Lemma 1. *Let $\varepsilon_t \sim \mathcal{S}(\alpha, \beta)$ with $\alpha \in (1, 2)$, and suppose $p \in (-1, \alpha)$.*

(i) *The p -th order absolute moment (or FLOM) is finite and given by*

$$\mathbb{E}|\varepsilon_t|^p = \begin{cases} 2 \cos(\alpha\theta)^{-1/\alpha} \Gamma(1 - 1/\alpha) \cos(\theta)/\pi & p = 1, \\ c(\alpha, \beta, p) \cos(p\theta)/\cos(p\pi/2) & p \neq 1, \end{cases} \quad (12)$$

where $\Gamma(\cdot)$ denotes the Gamma function, $\theta = \theta(\alpha, \beta)$ is given in (5), and

$$c(\alpha, \beta, p) = \cos(\alpha\theta)^{-p/\alpha} \frac{\Gamma(1 - p/\alpha)}{\Gamma(1 - p)}.$$

(ii) Define the signed power by $x^{(p)} = \text{sgn}(x)|x|^p$, with $\text{sgn}(x) = x/|x|$ for $x \neq 0$ and $\text{sgn}(0) = 0$. The p -th order signed moment is finite and given by

$$\mathbb{E}[\varepsilon_t^{(p)}] = \begin{cases} 2\theta/\pi & p = 0, \\ 0 & p = 1, \\ c(\alpha, \beta, p) \sin(p\theta)/\sin(p\pi/2) & \text{otherwise.} \end{cases} \quad (13)$$

The first moment condition used to extend the GMM estimator utilizes the sign of the observations, based on the signed moment in (13) for $p = 0$.

Proposition 1. Suppose that $\{y_t\}$ is generated by the SV process with stable disturbances in (1) such that $\alpha \in (1, 2)$. Then,

$$\mathbb{E}[\text{sgn}(y_t)] = d_\alpha^{-1} \arctan[\beta \tan(d_\alpha)], \quad d_\alpha = \pi\alpha/2. \quad (14)$$

Our first asymmetric GMM (“A-GMM”) estimator is obtained by augmenting the S-GMM estimator from the previous section with the following moment condition:

$$g_{\text{sgn}}(y_t, \psi) = \text{sgn}(y_t) - \mathbb{E}[\text{sgn}(y_t)].$$

This condition depends only on the sign of the data and is attractive for a number of reasons. First, the sign exhibits limited sensitivity to extreme values, and its boundedness ensures finite moments of any order. Furthermore, the moment condition contains only on the parameters α and β , and, since $\text{sgn}(y_t) = \text{sgn}(\varepsilon_t)$, we have that $\{g_{\text{sgn}}\}$ is serially independent. Therefore, the standard law of large numbers and central limit theorems for IID data apply, despite the presence of volatility clustering and heavy-tailed disturbances.

To obtain a second A-GMM estimator, we follow Kuruoglu (2001), who considered the ratio of the moments in Lemma 1 to obtain a simplified expression. The same approach can be used for the stochastic volatility model in (1).

Proposition 2. Suppose that $\{y_t\}$ is generated by the SV process with stable disturbances in (1) such that $\alpha \in (1, 2)$ and $\phi \in (-1, 1)$. For $p \in (-1, \alpha)$, the p -th order absolute and signed moments of the observations are finite and given by

$$\mathbb{E}|y_t|^p = \mathbb{E}\sigma_t^p \cdot \mathbb{E}|\varepsilon_t|^p, \quad \mathbb{E}[y_t^{(p)}] = \mathbb{E}\sigma_t^p \cdot \mathbb{E}[\varepsilon_t^{(p)}],$$

where $\mathbb{E}|\varepsilon_t|^p$ and $\mathbb{E}[\varepsilon_t^{(p)}]$ are given in Lemma 1, and with

$$\mathbb{E}[\sigma_t^p] = \mathbb{E}[\exp(h_t/2)^p] = \exp\left(\frac{p}{2}\left[\mu + \frac{p\sigma^2}{4(1-\phi^2)}\right]\right).$$

Furthermore, if $0 \neq p \neq 1$,

$$\frac{\mathbb{E}[y_t^{(p)}]}{\mathbb{E}|y_t|^p} = r(\alpha, \beta, p) \equiv \frac{\tan\left(\frac{p}{\alpha} \arctan\left[\beta \tan\left(\frac{\pi\alpha}{2}\right)\right]\right)}{\tan(p\pi/2)}. \quad (15)$$

The second A-GMM estimator is defined by augmenting the S-GMM estimator with the following moment condition:

$$g_r(y_t, \psi) = y_t^{<p>} - |y_t|^p \cdot r(\alpha, \beta, p) \quad p \in (-1, \alpha), \quad 0 \neq p \neq 1.$$

This condition reduces to g_{sgn} in the limit $p \rightarrow 0$, and similarly, it depends only on the parameters α and β . However, its use requires selecting a suitable value for p , which is treated as an additional tuning parameter. For both A-GMM variants, the tuning parameters were selected using the procedure from Appendix B as before. This resulted in $K = 20$ for the sign variant and $(p, K) = (0.1, 40)$ for the ratio variant.

5 Monte Carlo study

We have presented two estimators for the symmetric stable version of the SV model (QML and S-GMM), along with two variants of the A-GMM estimator for the asymmetric stable variant. Before studying their performance through various simulation studies, we briefly reflect on the qualitative differences between the estimators. Regarding the relative performance of QML and GMM, the literature reports mixed results for the SV model with normal disturbances (Jacquier et al., 1994; Ruiz, 1994). Two advantages of the QML estimator are that it also yields a corresponding estimator of the unobserved volatility and does not require the selection of tuning parameters. Advantages of the GMM approach are that standard errors are easily obtained (see Appendix C) and that the estimator can be readily extended to the asymmetric case. Comparing the two A-GMM estimators, the desirable properties of g_{sgn} suggest that this variant is preferred to the ratio variant in practice. We now further investigate the relative performance of the proposed estimators using simulation studies.

5.1 Results for QML and S-GMM

In this section, we compare the QML and S-GMM estimators by applying them to data generated by the SV model symmetric stable disturbances ($\beta = 0$). Two settings are considered for the remaining parameters, taken from Lombardi & Calzolari (2009):

- I. $\sigma = 0.35$, $\phi = 0.9$, $\mu = -0.7/(1 - \phi) = -7$;
- II. $\sigma = 0.06$, $\phi = 0.98$, $\mu = -0.15/(1 - \phi) = -7.5$.

In both cases, several values of α are considered: 1.8, 1.9, 1.95, 1.99. For each combination of parameters, the SV model in (1) is used to simulate 1000 samples of sizes $T \in \{1000, 3000, 5000\}$. The proposed estimators are applied to each of the simulated samples to estimate the parameters, and we compare the estimates with the true values based on the corresponding root mean squared error (RMSE) values.

Table 1 presents the results for the first parameter set. As the sample size increases, the RMSE values decrease, illustrating the consistency of the estimators. A strictly lower RMSE value up to three decimals is indicated by italicized entries. This is achieved by QML in 26 cases and S-GMM in five cases; the performance is equal in the remaining 17 cases. The QML estimates of σ are generally more accurate, while for $T = 1000$, the S-GMM estimates of ϕ achieve a lower RMSE for the values of α other than 1.95. Overall, the two estimators exhibit a very similar performance.

Table 2 presents the results for the second parameter set, which is characterized by a higher persistence of the volatility, with the autoregressive coefficient $\phi = 0.98$ being near one. Both methods obtain higher RMSE values in estimating this parameter, indicating that this task is now more challenging. This is also the parameter for which the largest differences in accuracy occur: where S-GMM tends to be better for the smallest sample size ($T = 1000$), QML generally performs better for the larger sample sizes. In particular, with $T = 5000$, QML performs at least as good as S-GMM for each parameter, with reductions in the RMSE up to about 0.05 in estimating ϕ for $\alpha \leq 1.9$. In total, QML achieves the lowest RMSE 21 times and S-GMM in 13 cases.

Overall, the results indicate that QML tends to be more accurate than S-GMM, with their performance typically being similar. When the volatility process has a higher degree of persistence, the difference in performance can be more pronounced. In this case, S-GMM seems to be more accurate in estimating the autoregressive coefficient in small samples, while QML performs better with larger samples.

Table 1: Root mean squared error values for simulation study based on first parameter set from Lombardi & Calzolari (2009).

α	T	QML			S-GMM		
		1000	3000	5000	1000	3000	5000
1.8	μ	0.155	<i>0.093</i>	0.071	0.155	0.094	0.071
	ϕ	0.092	<i>0.030</i>	<i>0.023</i>	<i>0.072</i>	0.032	0.025
	σ	0.134	<i>0.067</i>	<i>0.052</i>	0.134	0.072	0.056
	α	0.152	0.107	<i>0.081</i>	0.152	0.107	0.083
1.9	μ	0.150	0.089	<i>0.072</i>	0.150	0.089	0.073
	ϕ	0.078	0.030	<i>0.022</i>	<i>0.068</i>	0.030	0.023
	σ	0.117	<i>0.064</i>	<i>0.051</i>	0.117	0.066	0.053
	α	0.135	<i>0.093</i>	<i>0.081</i>	0.135	0.094	0.082
1.95	μ	0.148	0.089	0.068	0.148	0.089	0.068
	ϕ	<i>0.055</i>	<i>0.028</i>	<i>0.021</i>	0.056	0.029	0.022
	σ	<i>0.110</i>	<i>0.063</i>	<i>0.048</i>	0.116	0.067	0.050
	α	<i>0.126</i>	0.083	0.069	0.127	0.083	0.069
1.99	μ	0.150	0.081	0.066	0.150	0.081	<i>0.065</i>
	ϕ	0.079	<i>0.027</i>	<i>0.021</i>	<i>0.068</i>	0.028	0.022
	σ	<i>0.108</i>	<i>0.061</i>	<i>0.046</i>	0.118	0.065	0.050
	α	<i>0.124</i>	<i>0.081</i>	0.064	0.126	0.082	0.064

Note: Results based on 1000 simulated paths from the SV model in (1), with parameters $\beta = 0$ (symmetric), $\sigma = 0.35$, $\phi = 0.9$, $\mu = -0.7/(1 - \phi) = -7$, and various values of α . Italicized entries indicate a strictly lower root mean squared error value up to three reported decimals.

5.2 Results for A-GMM Estimators

For the A-GMM estimators, we use a similar setup as in the previous section, but different data generating processes (DGPs). Starting from the two parameter sets described in the previous section, we consider two values of β : -0.3 (mild asymmetry) and -0.6 (pronounced asymmetry). The tail index $\alpha = 1.9$ is used throughout. For higher values of α , the estimation of the asymmetry parameter β may become less useful. We notice that β is not even identified in the limiting case of $\alpha \rightarrow 2$ which reduces the stable distribution to the normal distribution.

Table 3 presents the results of this case. In 22 cases, the sign variant achieves the lowest RMSE value, the ratio variant in 18 cases, with equal performance in the remaining 20 cases. The performance is generally very similar, with the absolute difference in RMSE exceeding 0.01 in only seven out of 60 cases. In five of these cases, the sign variant outperforms the ratio variant, including the largest difference in RMSE of 0.034 points in estimating $\beta = -0.3$ with $T = 1000$. Overall, the sign variant of A-GMM tends to be either more accurate than or comparable to the ratio variant in estimating β .

Table 2: Root mean squared error values for simulation study based on second parameter set from Lombardi & Calzolari (2009).

α	T	QML			S-GMM		
		1000	3000	5000	1000	3000	5000
1.8	μ	0.142	0.089	<i>0.068</i>	0.142	0.089	0.069
	ϕ	0.420	<i>0.235</i>	<i>0.139</i>	<i>0.321</i>	0.289	0.195
	σ	0.152	0.121	<i>0.062</i>	0.152	<i>0.110</i>	0.072
	α	<i>0.152</i>	0.108	0.080	0.153	<i>0.106</i>	0.080
1.9	μ	0.139	<i>0.084</i>	0.065	<i>0.138</i>	0.085	0.065
	ϕ	0.504	<i>0.242</i>	<i>0.148</i>	<i>0.360</i>	0.257	0.196
	σ	0.145	0.087	<i>0.044</i>	<i>0.140</i>	<i>0.082</i>	0.058
	α	<i>0.126</i>	<i>0.090</i>	0.076	0.128	0.091	0.076
1.95	μ	0.142	0.083	0.062	<i>0.141</i>	0.083	0.062
	ϕ	0.520	<i>0.237</i>	<i>0.167</i>	<i>0.350</i>	0.242	0.178
	σ	0.117	0.077	<i>0.046</i>	0.117	<i>0.074</i>	0.058
	α	0.130	<i>0.085</i>	0.065	0.130	0.086	0.065
1.99	μ	0.136	0.079	0.061	0.136	0.079	0.061
	ϕ	0.530	0.238	<i>0.178</i>	<i>0.344</i>	<i>0.228</i>	0.218
	σ	0.110	<i>0.066</i>	<i>0.037</i>	<i>0.107</i>	0.071	0.051
	α	<i>0.122</i>	<i>0.078</i>	<i>0.066</i>	0.123	0.079	0.067

Note: Results based on 1000 simulated paths from the SV model in (1), with parameters $\beta = 0$ (symmetric), $\sigma = 0.06$, $\phi = 0.98$, $\mu = -0.15/(1 - \phi) = -7.5$, and various values of α . Italicized entries indicate a strictly lower root mean squared error value up to three reported decimals.

We also consider the simulation setup used in Blasques et al. (2025), using the SV model to generate 1000 samples of sizes 1000 and 3000 with the following parameters: $\alpha = 1.75$, $\mu = -0.2$, $\phi = 0.95$, and $\sigma = 0.2$, as in Vankov et al. (2019), and $\beta = 0$ (symmetric) or $\beta = 0.5$ (asymmetric). Table 4 shows the results of the simulation study. The sign and ratio variants typically yield similar performance. However, the sign variant attains a higher accuracy in estimating β in all cases, with differences up to 0.018 RMSE points for $T = 1000$ in the symmetric case.

For comparison, we also report results for the indirect inference (II) estimator of Blasques et al. (2025), converted from their Table 1. The A-GMM estimators perform better in estimating ϕ and σ , while the II estimator attains lower RMSE values for the remaining parameters. The II estimator performs relatively well in estimating β ; for α , the differences are of smaller magnitude, while for μ they are negligible. In sum, the A-GMM variants exhibit similar performance, with the sign variant performing better in estimating β . Compared to the II estimator, the best approach depends on the parameters, where the A-GMM estimators can provide a better performance for some ϕ and σ and a similar

Table 3: Root mean squared error values for simulation study based on parameter sets from Lombardi & Calzolari (2009) and various values of $\beta \neq 0$.

β	T	Sign			Ratio		
		1000	3000	5000	1000	3000	5000
-0.3	μ	0.152	0.085	0.070	<i>0.151</i>	0.085	0.070
	ϕ	<i>0.058</i>	<i>0.031</i>	0.023	0.060	0.033	0.023
	σ	0.123	<i>0.070</i>	<i>0.052</i>	<i>0.122</i>	0.074	0.054
	α	0.133	0.094	0.080	<i>0.131</i>	0.094	0.080
	β	<i>0.401</i>	<i>0.389</i>	<i>0.359</i>	0.435	0.401	0.372
-0.6	μ	0.153	0.092	0.073	<i>0.152</i>	<i>0.091</i>	0.073
	ϕ	<i>0.065</i>	<i>0.028</i>	<i>0.022</i>	0.076	0.029	0.023
	σ	0.120	<i>0.064</i>	<i>0.049</i>	<i>0.119</i>	0.066	0.053
	α	0.131	0.094	0.081	<i>0.130</i>	0.094	<i>0.080</i>
	β	<i>0.540</i>	0.432	0.389	0.548	0.432	<i>0.386</i>
-0.3	μ	0.136	0.084	0.066	0.136	0.084	0.066
	ϕ	<i>0.347</i>	0.229	0.206	0.363	0.229	<i>0.204</i>
	σ	<i>0.122</i>	0.085	0.067	0.125	<i>0.077</i>	<i>0.055</i>
	α	0.129	0.090	<i>0.078</i>	<i>0.128</i>	0.090	0.079
	β	<i>0.413</i>	<i>0.397</i>	<i>0.360</i>	0.421	0.406	0.366
-0.6	μ	0.140	0.082	0.069	0.140	0.082	0.069
	ϕ	<i>0.359</i>	<i>0.256</i>	0.219	0.368	0.257	0.219
	σ	0.121	0.085	0.062	<i>0.116</i>	<i>0.077</i>	<i>0.050</i>
	α	0.126	0.085	0.074	0.126	0.085	0.074
	β	<i>0.523</i>	0.431	0.388	0.528	<i>0.428</i>	<i>0.386</i>

Note: Results based on 1000 simulated paths from the SV model in (1), with the parameters $\alpha = 1.9$, various values of β , and either $\sigma = 0.35$, $\phi = 0.9$, $\mu = -0.7/(1 - \phi) = -7$ (top) or $\sigma = 0.06$, $\phi = 0.98$, $\mu = -0.15/(1 - \phi) = -7.5$ (bottom). Italicized entries indicate a strictly lower root mean squared error value up to three reported decimals.

RMSE for μ . These results seem encouraging, noting that the A-GMM methods do not require the use of simulation and are computationally light, where the II method is computationally more intensive.

6 Empirical Applications

In this section, we apply the QML and GMM estimators to six data sets comprising the daily returns of two stock indices, two exchange rates, and two cryptocurrencies. The stock indices are the S&P 500 ($T = 24664$) and the NASDAQ ($T = 13891$). The exchange rates are the US dollar against the Brazilian Real (USDBRL; $T = 5363$) and the British versus the Egyptian Pound (GBPEGP; $T = 3463$). The cryptocurrencies are Bitcoin

Table 4: Root mean squared error values for simulation study based on parameters from Blasques et al. (2025).

Sample size	Estimator	μ	ϕ	σ	α	β
$\beta = 0$ (symmetric)						
1000	Sign	0.174	<i>0.091</i>	0.110	0.151	0.325
	Ratio	0.174	0.096	<i>0.101</i>	0.150	0.343
	II	<i>0.159</i>	0.133	0.131	<i>0.072</i>	<i>0.18</i>
3000	Sign	0.104	0.024	0.058	0.096	0.208
	Ratio	0.104	<i>0.022</i>	<i>0.051</i>	0.095	0.220
	II	<i>0.103</i>	0.077	0.069	<i>0.047</i>	<i>0.09</i>
$\beta = 0.5$ (asymmetric)						
1000	Sign	0.169	<i>0.082</i>	0.111	0.146	0.341
	Ratio	0.169	0.087	0.104	0.146	0.342
	II	<i>0.159</i>	0.130	0.104	<i>0.073</i>	<i>0.267</i>
3000	Sign	0.100	0.025	0.059	0.093	0.269
	Ratio	0.101	<i>0.023</i>	<i>0.052</i>	0.093	0.272
	II	<i>0.099</i>	0.059	0.061	<i>0.045</i>	<i>0.154</i>

Note: Results are reported for the two variants of the A-GMM estimator. The indirect inference (II) results, converted from Table 1 of Blasques et al. (2025), are included as a benchmark. The results are based on 1000 simulated paths from the SV model in (1), with $\alpha = 1.75$, $\mu = -0.2$, $\phi = 0.95$, and $\sigma = 0.2$, and $\beta = 0$ (symmetric) or $\beta = 0.5$ (asymmetric). Italicized entries indicate a strictly lower root mean squared error value up to three reported decimals.

(BTCUSD; $T = 4197$) and Ethereum ($T = 3048$).

Table 5 presents the parameter estimates. We first focus on the estimators for the model with symmetric stable disturbances. The estimates of the QML and S-GMM estimators are very similar in almost all cases. The biggest differences are in the estimates of σ , but these can be explained by the corresponding estimate of ϕ , which together determine the unconditional variance of $\text{Var}[h_t] = \sigma^2/(1 - \phi^2)$.

For the S-GMM estimator, we also report the significance levels, based on the asymptotic standard errors computed as detailed in Appendix C. In almost all cases, the estimates are significant at the 1% level (***). For the parameter α , we test the conditionally normal null hypothesis $H_0: \alpha = 2$ against the conditionally stable alternative $H_a: \alpha < 2$. Because the parameter α is located on the boundary of the parameter space under the null hypothesis, the t -statistic follows a half-normal distribution under the null hypothesis (e.g., Andrews, 2002, p. 531). The squared t -statistic then follows an equally weighted mixture of a chi-squared distribution with one degree of freedom and the degenerate “distribution” at zero. In practice, this means we simply compute the γ -level critical value (e.g., $\gamma = 1\%$) by using the 2γ -level critical value of the chi-squared distribution with one

Table 5: Parameter estimates for SV model in (1).

Asset	Estimator	$\hat{\mu}$	$\hat{\phi}$	$\hat{\sigma}$	$\hat{\alpha}$	$\hat{\beta}$
S&P 500	QML	-1.205	0.994	0.106	1.837	-
	S-GMM	-1.205***	0.991***	0.134***	1.855***	-
	Sign	-1.211***	0.992***	0.127***	1.849***	-0.325***
	Ratio	-1.212***	0.992***	0.124***	1.848***	-0.334***
NASDAQ	QML	-1.022	0.994	0.112	1.807	-
	S-GMM	-1.017***	0.990***	0.154***	1.832***	-
	Sign	-1.051***	0.990***	0.151***	1.806***	-0.677***
	Ratio	-1.049***	0.992***	0.130***	1.801***	-0.656***
USDBRL	QML	-1.355	0.985	0.138	1.445	-
	S-GMM	-1.355***	0.982***	0.155***	1.449***	-
	Sign	-1.354***	0.972***	0.199***	1.458***	0.069**
	Ratio	-1.363***	0.987***	0.126***	1.440***	0.055**
GBPEGP	QML	-2.428	0.986	0.109	1.358	-
	S-GMM	-2.433***	1.000***	0.009	1.350***	-
	Sign	-2.655***	0.999***	0.022	1.266***	0.211***
	Ratio	-2.619***	0.999***	0.022	1.279***	0.203***
Bitcoin	QML	0.759	0.981	0.205	1.533	-
	S-GMM	0.761***	0.981***	0.202***	1.532***	-
	Sign	0.759***	0.983***	0.191***	1.528***	0.007
	Ratio	0.764***	0.979***	0.215***	1.538***	0.002
Ethereum	QML	1.448	0.985	0.137	1.512	-
	S-GMM	1.445***	0.969***	0.212***	1.537***	-
	Sign	1.448***	0.964***	0.231***	1.544***	-0.002
	Ratio	1.435***	0.980***	0.160***	1.518***	-0.005

Note: Significance levels are denoted by *** (1%), ** (5%), and * (10%), and are reported only for the GMM estimators.

degree of freedom. For all six data sets, the conditionally normal hypothesis is rejected in favor of the conditionally stable one. For the exchange rates and cryptocurrencies, the estimates of α are lowest, between 1.3 – 1.6, corresponding to considerably heavy tailed distributions for the observation disturbances.

Table 5 also provides the parameter estimates for the asymmetric stable version of the SV model. The two variants of the A-GMM estimator provide similar results, both in terms of estimates and significance levels, with most estimates being significant at the 1% level. The insignificance of the asymmetry parameter for Bitcoin aligns with the findings from Blasques et al. (2025), while the corresponding result for Ethereum suggests that the symmetric conditional distribution could be most appropriate for cryptocurrencies more generally. For the other time series, the asymmetry parameter is found to be significantly

different from zero, with the estimated asymmetry being strongest for the S&P 500, the NASDAQ, and the GBPEGP series. Inclusion of the asymmetry parameter appears to have little impact on the remaining parameters, with the estimates being well aligned with those obtained using QML and S-GMM.

As previously noted, the QML approach also provides a simple way to estimate the unobserved volatility (Harvey et al., 1994), for which we use the expression in Equation (9). We use the Kalman smoother based on the state space form in (6) to compute the smoothed volatilities $\mathbb{E}[\sigma_t|\tilde{Y}_T]$ for $t = 1, \dots, T$. We compute the smoothed estimates for the six time series, in each case setting the model parameters to the estimates from the sign variant of A-GMM reported in Table 5.

Figures 1 and 2 show the smoothed volatilities $\hat{\sigma}_{t|T}$ together with the absolute returns $|y_t|$. As expected, the estimated volatility tends to increase during periods of high absolute returns. However, the volatility is considerably smoother than the returns, noting that part of the movements in y_t is attributed to the heavy-tailed observation disturbances, which are assumed to follow a stable distribution. The returns exhibit the largest movements for the cryptocurrencies (shown in the bottom plots), as well as for the GBPEGP exchange rate (Figure 2, middle plot). In the latter case, the volatility movements are less apparent, as the estimated SV process corresponds to small values of μ and σ and the plot is dominated by two extreme returns. For the equity indices (top plots) and the USDBRL exchange rate (Figure 1, middle plot), the most extreme returns are smaller in magnitude, and the smoothed volatility tends to peak around these observations, while remaining relatively insensitive to the outliers.

7 Discussion

In this study, we have proposed methods for estimating SV models with stable disturbances via linearization, which are straightforward to implement and do not require the use of simulation. The use of linearization aligns our work with related studies in the engineering and statistics literature (Ma & Nikias, 1995a; Kuruoglu, 2001; Wang et al., 2015; Lévy Véhel et al., 2021) and extends this strategy to allow for stochastic and time-varying scales. For the model with symmetric stable distribution, QML and GMM estimators are considered. To estimate the asymmetric version of the model, two extensions of the GMM estimator are considered.

The performance of the estimators is investigated by means of various simulation studies. For the SV model with symmetric stable disturbances, the QML estimator tends to be more accurate than S-GMM, though their performance is generally similar. For the model with an asymmetric stable distribution, the two versions of A-GMM generally

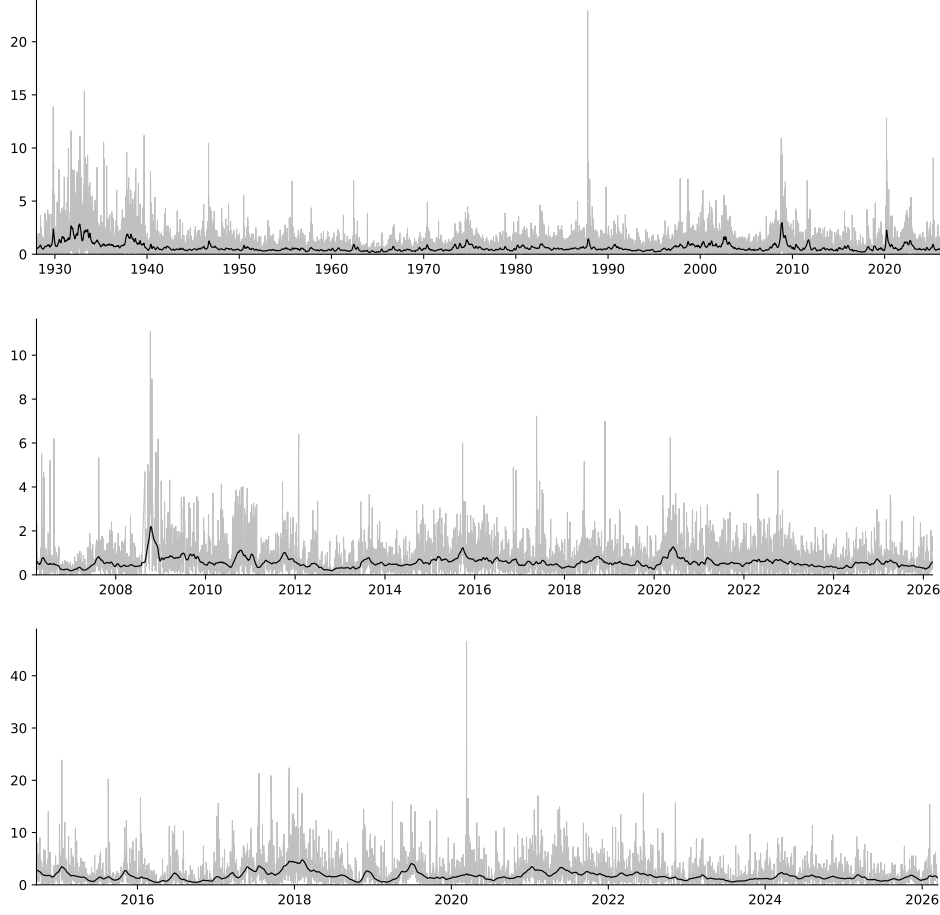


Figure 1: Smoothed volatility $\hat{\sigma}_{t|T} = \mathbb{E}[\sigma_t|\tilde{Y}_T]$ (black) and absolute returns $|y_t|$ (grey), computed with the Kalman smoother based on the state space form in (6) and the A-GMM sign estimates of Table 5. From top to bottom: S&P 500, USDBRL, Bitcoin.

exhibit similar performance, though the sign variant tends to be more accurate than the ratio variant in estimating β . From a qualitative point of view, the QML and sign A-GMM estimators also have the advantage of requiring fewer tuning parameters (zero and one parameter, respectively) than the S-GMM and ratio A-GMM estimators. Furthermore, the limited sensitivity of the sign sample moment to the returns used in the corresponding variant can be advantageous for data with frequent outliers.

The methods have also been applied empirically, to six data sets comprising two stock indices, two exchange rates, and two cryptocurrencies. The different methods yield similar estimates, which supports their validity. Considerable asymmetry is found in the conditional distribution of the returns for the S&P 500, the NASDAQ, and the GBPEGP exchange rate series. The QML approach also allows for straightforward estimation of the unobserved volatility via the Kalman filter (Harvey et al., 1994). The smoothed estimates of the volatility are intuitive, increasing during periods of high absolute returns, while

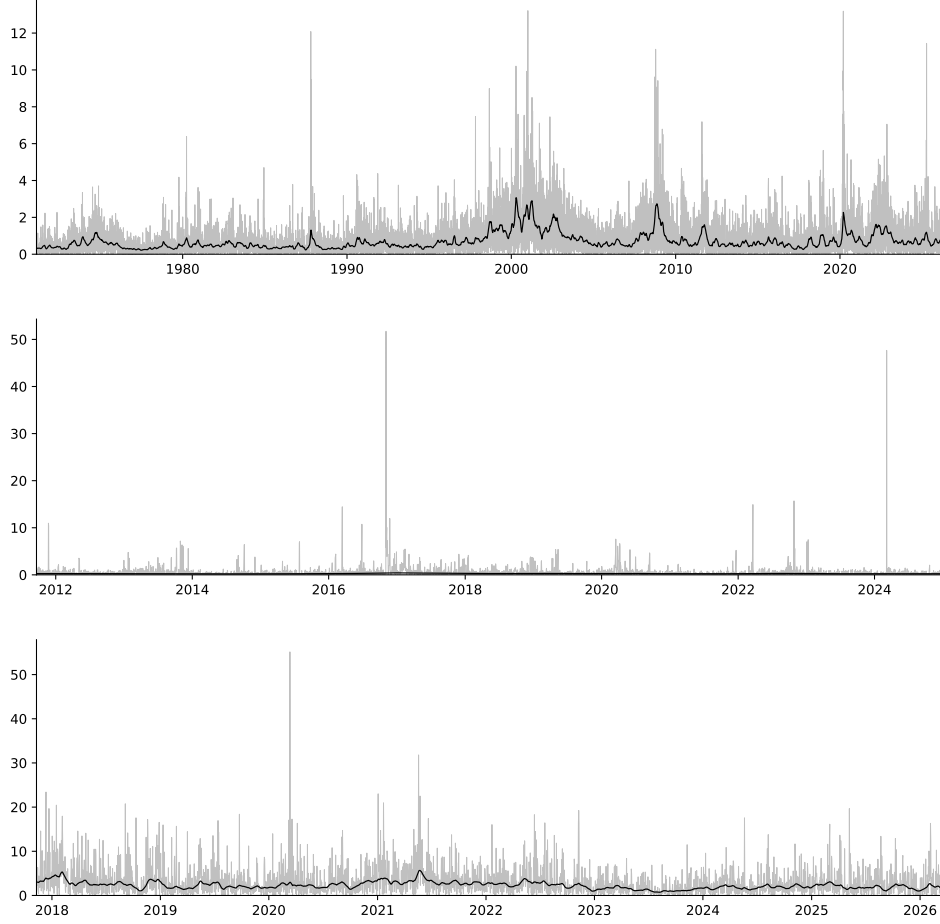


Figure 2: Smoothed volatility $\hat{\sigma}_{t|T}$ (black) and absolute returns $|y_t|$ (grey), continued. From top to bottom: NASDAQ, GBPEGP, Ethereum.

remaining relatively insensitive to the outliers.

The proposed methods can be extended in several ways. A possible approach to extend the QML estimator to the asymmetric case is to solve the expression for $\mathbb{E}[\text{sgn}(y_t)]$ given in (14) for β . Substituting the sample average of the signs for their expectation, this yields the following estimator (for given α):

$$\beta^* = \frac{\tan \left[d_\alpha T^{-1} \sum_{t=1}^T \text{sgn}(y_t) \right]}{\tan(d_\alpha)}, \quad d_\alpha = \pi\alpha/2.$$

In this way, β is “concentrated out” of the quasi-log likelihood optimization, with no further modifications needed to the QML procedure. Notably, this estimate need not be equivalent to the corresponding A-GMM estimate, since β also impacts the range of $\text{Var}[\tilde{\varepsilon}_t]$ and, therefore, the fit to the sample moment corresponding to $\text{Var}[\tilde{y}_t]$. Lower values of the quadratic form objective may therefore be attained by including the asymmetry parameter in the optimization, as was done in Section 4.2.

An extension of the model is the regime-switching variant proposed by Casarin (2004). With $r \in \mathbb{N}$ regimes, the model is given by

$$\begin{aligned} y_t &= \exp(h_t/2)\varepsilon_t, & \varepsilon_t &\sim \mathcal{S}(\alpha, 0), \\ h_{t+1} &= \mu_{s_t} + \phi(h_t - \mu_{s_t}) + \eta_t, & \eta_t &\sim \mathcal{N}(0, \sigma^2), \\ \mathbb{P}(s_{t+1} = j \mid s_t = i) &= p_{ij}, & \sum_{j=1}^r p_{ij} &= 1, \quad p_{ij} \geq 0, \quad i, j = 1, \dots, r. \end{aligned}$$

The model has r different possible means $\mu_1, \dots, \mu_r$ for h_t , and, to ensure valid transition probabilities, $p_{ir} = 1 - \sum_{j=1}^{r-1} p_{ij}$ is defined for $i = 1, \dots, r$. The means and transition probabilities together yield a total of r^2 free parameters, with small values of r retaining parsimony. By adopting the corresponding state space form by analogy to (6), this version of the model may be estimated straightforwardly via the QML approach, in which the approximate filter of Kim (1994) replaces the usual Kalman filter.

The above extensions illustrate the flexibility of the QML approach in accommodating modifications to the volatility process. Using the state space form makes it straightforward to allow for alternative dynamics, such as multifactor volatility processes, which can emulate long-memory behavior. Further consideration of these and other extensions is left to future research.

Conflicts of Interest

The authors declare no conflicts of interest.

ORCID

Siem Jan Koopman  <https://orcid.org/0000-0002-4440-9524>

Karim Moussa  <https://orcid.org/0000-0001-7760-3529>

References

- Andersen, T. G., & Sørensen, B. E. (1996). Gmm estimation of a stochastic volatility model: A monte carlo study. *Journal of Business & Economic Statistics*, 14(3), 328–352.
- Andrews, D. W. K. (2002). Generalized method of moments estimation when a parameter is on a boundary. *Journal of Business & Economic Statistics*, 20(4), 530–544.

- Blasques, F., Koopman, S. J., & Moussa, K. (2025). Asymmetric stable stochastic volatility models: estimation, filtering, and forecasting. *Journal of Time Series Analysis*, 46(6), 1098–1124.
- Broto, C., & Ruiz, E. (2004). Estimation methods for stochastic volatility models: a survey. *Journal of Economic surveys*, 18(5), 613–649.
- Calvet, L. E., & Czellar, V. (2015). Accurate methods for approximate bayesian computation filtering. *Journal of Financial Econometrics*, 13(4), 798–838.
- Casarin, R. (2004). Bayesian inference for generalised markov switching stochastic volatility models. *CEREMADE Journal Working Paper*(0414).
- Chou-Chen, S. W., & Morettin, P. A. (2020). Indirect inference for locally stationary arma processes with stable innovations. *Journal of Statistical Computation and Simulation*, 90(17), 3106–3134.
- Durbin, J., & Koopman, S. J. (2012). *Time series analysis by state space methods* (Vol. 38). OUP Oxford.
- Gnedenko, B. V., & Kolmogorov, A. N. (1954). *Limit distributions for sums of independent random variables*. Addison-Wesley Reading, MA.
- Hansen, L. P. (1982). Large sample properties of generalized method of moments estimators. *Econometrica: Journal of the econometric society*, 1029–1054.
- Harvey, A. (1990). *Forecasting, structural time series models and the kalman filter*. Cambridge university press.
- Harvey, A., Ruiz, E., & Shephard, N. (1994). Multivariate stochastic variance models. *The Review of Economic Studies*, 61(2), 247–264.
- Hayashi, F. (2000). *Econometrics*. Princeton, NJ: Princeton University Press.
- Jacquier, E., Polson, N. G., & Rossi, P. E. (1994). Bayesian analysis of stochastic volatility models. *Journal of Business & Economic Statistics*, 20(1), 69–87.
- Karasaridis, A., & Hatzinakos, D. (2001). Network heavy traffic modeling using/spl alpha/-stable self-similar processes. *IEEE Transactions on Communications*, 49(7), 1203–1214.
- Kim, C.-J. (1994). Dynamic linear models with markov-switching. *Journal of econometrics*, 60(1-2), 1–22.

- Kuruoglu, E. E. (2001). *Density parameter estimation of skewed-stable distributions* (Vol. 49).
- Lévy Véhel, J., Philippe, A., & Robet, C. (2021). Explicit and combined estimators for parameters of stable distributions. *Statistics*, 55(4), 736–764.
- Liu, S.-M., & Brorsen, B. W. (1995). Maximum likelihood estimation of a garch-stable model. *Journal of Applied Econometrics*, 10(3), 273–285.
- 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 α -stable noise. *IEEE Transactions on Signal Processing*, 54(2), 775–779.
- Ma, X., & Nikias, C. L. (1995a). *Parameter estimation and blind channel identification in impulsive signal environments* (Vol. 43).
- Ma, X., & Nikias, C. L. (1995b). Parameter estimation and blind channel identification in impulsive signal environments. *IEEE Transactions on signal processing*, 43(12), 2884–2897.
- Machado, J. T., & Lopes, A. M. (2020). Rare and extreme events: the case of covid-19 pandemic: Ja tenreiro machado, am lopes. *Nonlinear dynamics*, 100(3), 2953–2972.
- Mandelbrot, B. (1963). The variation of certain speculative prices. *Journal of business*, 36(4), 394.
- McCulloch, J. H. (1996). 13 financial applications of stable distributions. *Handbook of statistics*, 14, 393–425.
- 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(1-2), 239–265.
- Moussa, K., & Koopman, S. J. (2026). Simulation smoothing for nonlinear non-gaussian state space models using machine learning methods.
- Müller, G., & Uhl, S. (2021). Estimation of time-varying autoregressive stochastic volatility models with stable innovations. *Statistics and Computing*, 31(3), 36.

- Newey, W. K., & West, K. D. (1987, May). A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica*, 55(3), 703–708. doi: 10.2307/1913610
- Nikias, C. L., & Shao, M. (1995). *Signal processing with alpha-stable distributions and applications*. Wiley-Interscience.
- Nolan, J. P. (2020). *Univariate stable distributions*. Springer.
- Nolan, J. P., & Ravishanker, N. (2009). Simultaneous prediction intervals for arma processes with stable innovations. *Journal of Forecasting*, 28(3), 235–246.
- Paolella, M. S. (2016). Stable-garch models for financial returns: Fast estimation and tests for stability. *Econometrics*, 4(2), 25.
- Peters, G. W., Kannan, B., Lasscock, B., Mellen, C., & Godsill, S. (2011). Bayesian cointegrated vector autoregression models incorporating alpha-stable noise for inter-day price movements via approximate bayesian computation. *Bayesian Analysis*, 6(4), 755–792. doi: 10.1214/11-BA628
- Qiou, Z., & Ravishanker, N. (1998). Bayesian inference for time series with stable innovations. *Journal of Time Series Analysis*, 19(2), 235–249.
- Ruiz, E. (1994). Quasi-maximum likelihood estimation of stochastic volatility models. *Journal of econometrics*, 63(1), 289–306.
- Samoradnitsky, G., & Taqqu, M. S. (1994). *Stable non-gaussian random processes: stochastic models with infinite variance*. Chapman & Hall/CRC.
- Sampaio, J. M., & Morettin, P. A. (2020). Stable randomized generalized autoregressive conditional heteroskedastic models. *Econometrics and Statistics*, 15, 67–83.
- 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, 29–52.
- Wang, X., Li, K., Gao, P., & Meng, S. (2015). Research on parameter estimation methods for alpha stable noise in a laser gyroscope’s random error. *Sensors*, 15(8), 18550–18564.

Appendix

A Proofs

A.1 Lemma 1

The lemma follows by collecting the results from Nolan (2020, pp.115-116), with particular attention paid to the special cases. The expression for $p \neq 1$ in Part (i) is given in Corollary 3.5 (c) of Nolan (2020). For $p = 1$, we take the corresponding limit $p \rightarrow 1$, noting that

$$\lim_{p \rightarrow 1} \Gamma(1-p) \cos(p\pi/2) = \pi/2.$$

For Part (ii), we first consider $p = 1$. By definition, $x^{(1)} = x$, and hence

$$\mathbb{E}[\varepsilon_t^{(1)}] = \mathbb{E}[\varepsilon_t] = 0.$$

For the remaining nonzero values of p , the moment is given in Corollary 3.5 (c) of Nolan (2020). For $p = 0$, we take the corresponding limit, noting that

$$\lim_{p \rightarrow 0} c(\alpha, \beta, p) \sin(p\theta) / \sin(p\pi/2) = 2\theta/\pi.$$

A.2 Proposition 1

Since $\sigma_t = \exp(h_t/2) > 0$, we have that $\text{sgn}(y_t) = \text{sgn}(\varepsilon_t) = \varepsilon_t^{(0)}$, where the latter follows by definition of the signed power (Nolan, 2020, Lemma 3.16 (c)). Equation (14) then follows from Equation (13) with $p = 0$ and substituting Equation (5) for θ .

A.3 Proposition 2

Since $\phi \in (-1, 1)$, it follows that the process for h_t is strictly stationary, and the same holds for $\sigma_t = \exp(h_t/2)$ and $y_t = \sigma_t \varepsilon_t$, so their moments are time-invariant. Since

$$\frac{p}{2} \cdot h_t \sim \mathcal{N}\left(\frac{p\mu}{2}, \frac{(p\sigma)^2}{4[1-\phi^2]}\right),$$

using the log normality of $\exp(ph_t/2)$, together with $\sigma_t > 0$, we obtain

$$\mathbb{E}|\sigma_t|^p = \mathbb{E}[\sigma_t^p] = \mathbb{E}\left[\exp\left(\frac{p}{2} \cdot h_t\right)\right] = \exp\left(\frac{p}{2} \left[\mu + \frac{p\sigma^2}{4(1-\phi^2)}\right]\right).$$

The assumptions in Proposition 2 further ensure that Lemma 1 applies, so that the p -th order absolute and signed moments of ε_t are finite. Combining finiteness of the above moments with independence of σ_t and ε_t , it follows that the absolute moment of y_t can be decomposed as

$$\mathbb{E}|y_t|^p = \mathbb{E}|\sigma_t|^p \cdot \mathbb{E}|\varepsilon_t|^p = \mathbb{E}\sigma_t^p \cdot \mathbb{E}|\varepsilon_t|^p.$$

Similarly, noting that $\text{sgn}(y_t) = \text{sgn}(\varepsilon_t)$ as $\sigma_t > 0$, we obtain the signed moment

$$\mathbb{E}[y_t^{(p)}] = \mathbb{E}[\text{sgn}(y_t) \cdot |y_t|^p] = \mathbb{E}[\sigma_t^p \cdot \text{sgn}(\varepsilon_t) \cdot |\varepsilon_t|^p] = \mathbb{E}\sigma_t^p \cdot \mathbb{E}[\varepsilon_t^{(p)}].$$

Lastly, we note that the p -th order moment of σ_t cancels out in the ratio $\mathbb{E}[y_t^{(p)}]/\mathbb{E}|y_t|^p = \mathbb{E}[\varepsilon_t^{(p)}]/\mathbb{E}|\varepsilon_t|^p$. We therefore obtain Equation (15) by substituting the corresponding expressions from Lemma 1 for $p \neq 1$.

B Selection of GMM Tuning parameters

To apply the GMM estimators from Section 4 in practice, it is needed to select suitable tuning parameters. For the S-GMM estimator and the sign variant of the A-GMM estimator, the sole tuning parameter is the number of moment conditions M , or equivalently, the number $K \geq 2$ of autocovariances. For the ratio variant of the A-GMM estimator, both p and K have to be chosen. We select the tuning parameters by optimizing the accuracy of the estimators, aggregated over several data generating processes (DGPs) and sample sizes, over a set of candidate tuning parameters.

We consider the following candidate values:

$$K \in \{10, 15, 20, 25, 30, 35, 40, 50\}, \quad p \in \{0.1, 0.15, 0.2, 0.25, 0.3, 0.4, 0.5, 0.6, 0.75\},$$

where p is used only in the ratio variant of the A-GMM estimator. As DGPs, we consider the SV model in (1) with parameters from the simulation study in Section 5.1 of Blasques et al. (2025). This corresponds to $\mu = -0.2$, $\phi = 0.95$, $\sigma = 0.2$, and $\alpha = 1.75$, as in Vankov et al. (2019), for both symmetric ($\beta = 0$) and asymmetric ($\beta = 0.5$) stable distributions. We use the sample sizes $T = 1000$ and $T = 3000$ as in Blasques et al. (2025). This results in four distinct scenarios, defined as a combination of the DGP (symmetric/asymmetric) and the sample size T .

For each scenario, we use the SV model to simulate 1000 samples of size T . We then apply the proposed estimators to each sample, where every candidate tuning parameter corresponds to a different instance of the estimator. For each such estimator, this yields 1000 parameter estimates, which are used to compute corresponding mean squared er-

rors (MSEs) relative to the true values. The sum of the MSE values for the five model parameters can then be computed. However, the resulting summed MSE (SMSE) values may not be comparable between scenarios, which complicates the tuning parameter selection. To this end, denote the minimum SMSE value for a particular estimator in a given scenario by SMSE^* , and define the excess SMSE for tuning parameter choice i (i.e., i th value of K , or i th combination of p and K) relative to the minimum:

$$\text{ESMSE}_i = 100 \cdot \frac{\text{SMSE}_i - \text{SMSE}^*}{\text{SMSE}^*}.$$

The ESMSE values are expressed as percentages, which facilitates their comparison over different scenarios.

We then select the tuning parameters by applying a mini-max rule to the ESMSE values over the four scenarios. Specifically, let $\overline{\text{ESMSE}}_i$ denote the largest ESMSE value over the four scenarios attained by the i th tuning parameter candidate, we select the candidate having the *smallest* value for $\overline{\text{ESMSE}}_i$.

As a robustness check, we also considered minimization of the mean ESMSE values over the different scenarios, instead of the mini-max rule. In all cases, this led to the same tuning parameters being selected.

C Computing Standard Errors in GMM

For the GMM estimator with identity weighting matrix, the asymptotic distribution is

$$\sqrt{T}(\hat{\psi} - \psi_0) \xrightarrow{d} \mathcal{N}(0, V), \quad V = (G'G)^{-1} G' S G (G'G)^{-1},$$

where $G \equiv \partial g(\psi_0)/\partial \psi'$ and $S \equiv \lim_{T \rightarrow \infty} \text{Var}[\sqrt{T} g(\psi_0)]$ is the long-run covariance of the moment vector. With G given in closed form by (D.14), the remaining task for finite-sample inference is to estimate S .

Since the moments are constructed from time series data, serial correlation and heteroskedasticity in the moment errors cannot be ruled out. Therefore, we compute the GMM standard errors using a Newey-West heteroskedasticity and autocorrelation consistent (HAC) estimator for the long-run covariance matrix of the moment conditions. This provides a robust estimate of S in the sandwich covariance matrix.

The estimator accounts for potential heteroskedasticity and autocorrelation by adding weighted lag covariances of the moment vector. We use the Newey-West estimator with

a Bartlett kernel:

$$\hat{S} = \hat{\Gamma}_0 + \sum_{\ell=1}^{L_T} \left(1 - \frac{\ell}{L_T + 1}\right) (\hat{\Gamma}_\ell + \hat{\Gamma}_\ell'), \quad \hat{\Gamma}_\ell = \frac{1}{T} \sum_{t=\ell+1}^T \tilde{g}_t \tilde{g}_{t-\ell}',$$

with $\tilde{g}_t \equiv g_t(\hat{\psi}) - \bar{g}(\hat{\psi})$ the demeaned moment vector and default bandwidth $L_T = \lfloor 4(T/100)^{2/9} \rfloor$ (Newey & West, 1987). The Bartlett-kernel construction guarantees that $\hat{S}$ is positive semi-definite for any L_T . The significance levels reported in the empirical applications are based on standard errors computed via

$$\widehat{\text{Var}}(\hat{\psi}) = \frac{1}{T} (\hat{G}'\hat{G})^{-1} \hat{G}' \hat{S} \hat{G} (\hat{G}'\hat{G})^{-1}, \quad \text{SE}_j = \sqrt{\widehat{\text{Var}}(\hat{\psi})_{jj}}.$$

D Jacobian G for GMM

Below we derive the Jacobian $G(\psi) \equiv \partial g(\psi)/\partial \psi'$ of the GMM moment vector for the SV model with asymmetric stable disturbances in Equation (1). In this section only, we define, with slight abuse of notation, $\tilde{y}_t \equiv \log |y_t|$ and $\tilde{\varepsilon}_t \equiv \log |\varepsilon_t|$, so that these quantities are one-half of their counterparts in the main text. This does not affect the results. The quantity θ is as given in Equation (5). We derive the per-parameter partial derivatives directly in Section D.2, assemble the Jacobian row by row in Section D.3, and display the full matrix in Section D.4. The special case $\beta = 0$ is treated in Appendix D.5.

D.1 Model and Moment Vector

The SV model with stable disturbances is given by

$$y_t = \exp(h_t/2) \varepsilon_t, \quad h_{t+1} = \mu + \phi(h_t - \mu) + \sigma \eta_t,$$

with $\eta_t \sim \mathcal{N}(0, 1)$ and $\varepsilon_t \sim \mathcal{S}(\alpha, \beta)$. We write $\psi \equiv (\mu, \phi, \sigma, \alpha, \beta)'$ for the parameter vector. We focus on the A-GMM estimators, which contain all the moment conditions used in the S-GMM estimator.

The moment conditions are the mean, variance, K autocovariances of $\tilde{y}_t = \log |y_t|$, and an additional condition including β , which is the sign condition for the sign variant and the FLOM ratio for the ratio variant.

This gives the following moment conditions:

$$g_{1,t}(\psi) = \tilde{y}_t - \left(\frac{1}{2}\mu + \mathbb{E}[\tilde{\varepsilon}_t]\right), \quad (\text{D.1})$$

$$g_{2,t}(\psi) = (\tilde{y}_t - \bar{\tilde{y}})^2 - \left(\frac{\sigma^2}{4(1-\phi^2)} + \text{Var}[\tilde{\varepsilon}_t]\right), \quad (\text{D.2})$$

$$g_{2+k,t}(\psi) = (\tilde{y}_t - \bar{\tilde{y}})(\tilde{y}_{t-k} - \bar{\tilde{y}}) - \frac{\sigma^2}{4(1-\phi^2)}\phi^k, \quad k = 1, \dots, K, \quad (\text{D.3})$$

$$g_{M,t}(\psi) = \text{sgn}(y_t) - \frac{2\theta}{\pi} \quad (\text{sign variant}), \quad (\text{D.4})$$

$$g_{M,t}(\psi) = |y_t|^p \left[\text{sgn}(y_t) - \frac{\tan(p\theta)}{\tan(p\pi/2)} \right] \quad (\text{ratio variant}), \quad (\text{D.5})$$

with $M \equiv K + 3$. The corresponding theoretical log-absolute moments are (Lemma 3.19 of Nolan (2020)):

$$\mathbb{E}[\tilde{\varepsilon}_t] = \gamma_{\text{Euler}}\left(\frac{1}{\alpha} - 1\right) - \frac{1}{\alpha} \log \cos(\alpha \theta), \quad \text{Var}[\tilde{\varepsilon}_t] = \frac{\pi^2}{12} \left(1 + \frac{2}{\alpha^2}\right) - \theta^2, \quad (\text{D.6})$$

with $\theta = \alpha^{-1} \arctan(\beta \tan(\pi\alpha/2))$.

The sample moment vector is $g(\psi) = T^{-1} \sum_t g_t(\psi)$ (the autocovariance entries sum over $t = k+1, \dots, T$). Because the sample part of every moment is free of ψ , the Jacobian reduces to the negative gradient of the theoretical part, which is what we now compute. The only exception is the ratio condition in (D.5).

D.2 Partial Derivatives

To assemble the Jacobian we need, for each parameter in $\psi = (\mu, \phi, \sigma, \alpha, \beta)'$, the partial derivative of every model quantity that appears in the moment conditions. Six such quantities appear: θ , the log-stable mean $\mathbb{E}[\tilde{\varepsilon}_t]$, the log-stable variance $\text{Var}[\tilde{\varepsilon}_t]$, the variance term $\sigma^2/(4(1-\phi^2))$ in the SV component, the FLOM ratio $\tan(p\theta)/\tan(p\pi/2)$ and the mean of the sign $2\theta/\pi$. Each is treated in its own subsection below.

D.2.1 Derivatives of θ with Respect to α and β

Write $\theta = A/\alpha$ with $A \equiv \arctan(\beta \tan(\pi\alpha/2))$ as a local shorthand for this subsection only.

The derivative of $\tan(\pi\alpha/2)$ uses $\tan'(x) = \sec^2(x) = 1 + \tan^2(x)$ and the chain rule with inner derivative $\pi/2$:

$$\frac{d}{d\alpha} \tan\left(\frac{\pi\alpha}{2}\right) = \frac{\pi}{2} \left(1 + \tan^2\left(\frac{\pi\alpha}{2}\right)\right).$$

For $A = \arctan(u)$ with $u = \beta \tan(\pi\alpha/2)$, we have $\arctan'(u) = 1/(1+u^2)$. The chain rule gives

$$\begin{aligned}\frac{\partial A}{\partial \alpha} &= \frac{1}{1 + \beta^2 \tan^2(\pi\alpha/2)} \cdot \beta \cdot \frac{\pi}{2} (1 + \tan^2(\pi\alpha/2)) = \frac{\beta (\pi/2) (1 + \tan^2(\pi\alpha/2))}{1 + \beta^2 \tan^2(\pi\alpha/2)}, \\ \frac{\partial A}{\partial \beta} &= \frac{1}{1 + \beta^2 \tan^2(\pi\alpha/2)} \cdot \tan(\pi\alpha/2) = \frac{\tan(\pi\alpha/2)}{1 + \beta^2 \tan^2(\pi\alpha/2)}.\end{aligned}$$

Applying the quotient rule to $\theta = A/\alpha$ then gives

$$\frac{\partial \theta}{\partial \alpha} = \frac{1}{\alpha} \frac{\partial A}{\partial \alpha} - \frac{\theta}{\alpha}, \quad \frac{\partial \theta}{\partial \beta} = \frac{1}{\alpha} \frac{\partial A}{\partial \beta}. \quad (\text{D.7})$$

The β -derivative follows directly since α does not depend on β . The α -derivative follows from $\partial(A/\alpha)/\partial\alpha = (\alpha \partial A/\partial\alpha - A)/\alpha^2$ and substituting $A = \alpha\theta$ in the second term.

D.2.2 Derivatives of $\mathbb{E}[\tilde{\varepsilon}_t]$

From (D.6), $\mathbb{E}[\tilde{\varepsilon}_t] = \gamma_{\text{Euler}}(\alpha^{-1} - 1) - \alpha^{-1} \log \cos(\alpha\theta)$. Both factors depend on α , so we apply the product rule to the second term.

The derivative of $\log \cos(\alpha\theta)$ with respect to either α or β uses $(d/du) \log \cos(u) = -\tan(u)$ and the chain rule. The product rule for $\alpha\theta$ gives $\partial(\alpha\theta)/\partial\alpha = \theta + \alpha \partial\theta/\partial\alpha$ and $\partial(\alpha\theta)/\partial\beta = \alpha \partial\theta/\partial\beta$ (since α is constant in β). Therefore

$$\frac{\partial}{\partial \alpha} \log \cos(\alpha\theta) = -\tan(\alpha\theta) \left(\theta + \alpha \frac{\partial \theta}{\partial \alpha} \right), \quad \frac{\partial}{\partial \beta} \log \cos(\alpha\theta) = -\alpha \tan(\alpha\theta) \frac{\partial \theta}{\partial \beta}.$$

For the α -derivative of $\mathbb{E}[\tilde{\varepsilon}_t]$, the first term contributes $-\gamma_{\text{Euler}}/\alpha^2$, and the second term, by the quotient rule applied to $\log \cos(\alpha\theta)/\alpha$, contributes

$$-\frac{\partial}{\partial \alpha} \left(\frac{\log \cos(\alpha\theta)}{\alpha} \right) = -\frac{\alpha \partial \log \cos(\alpha\theta)/\partial \alpha - \log \cos(\alpha\theta)}{\alpha^2} = \frac{\log \cos(\alpha\theta)}{\alpha^2} - \frac{1}{\alpha} \frac{\partial \log \cos(\alpha\theta)}{\partial \alpha}.$$

Combining,

$$\frac{\partial \mathbb{E}[\tilde{\varepsilon}_t]}{\partial \alpha} = -\frac{\gamma_{\text{Euler}}}{\alpha^2} + \frac{\log \cos(\alpha\theta)}{\alpha^2} - \frac{1}{\alpha} \frac{\partial \log \cos(\alpha\theta)}{\partial \alpha}. \quad (\text{D.8})$$

For the β -derivative, the first term in $\mathbb{E}[\tilde{\varepsilon}_t]$ is constant in β , and

$$\frac{\partial \mathbb{E}[\tilde{\varepsilon}_t]}{\partial \beta} = -\frac{1}{\alpha} \frac{\partial \log \cos(\alpha\theta)}{\partial \beta} = \tan(\alpha\theta) \frac{\partial \theta}{\partial \beta}. \quad (\text{D.9})$$

D.2.3 Derivatives of $\text{Var}[\tilde{\varepsilon}_t]$

From (D.6), $\text{Var}[\tilde{\varepsilon}_t] = \pi^2/12 + (\pi^2/6)\alpha^{-2} - \theta^2$. The power rule on α^{-2} gives $-2\alpha^{-3}$, and the chain rule on θ^2 gives $2\theta \partial\theta/\partial\cdot$. The first two terms of $\text{Var}[\tilde{\varepsilon}_t]$ are independent of β . Hence

$$\frac{\partial \text{Var}[\tilde{\varepsilon}_t]}{\partial \alpha} = -\frac{\pi^2}{3\alpha^3} - 2\theta \frac{\partial \theta}{\partial \alpha}, \quad \frac{\partial \text{Var}[\tilde{\varepsilon}_t]}{\partial \beta} = -2\theta \frac{\partial \theta}{\partial \beta}. \quad (\text{D.10})$$

D.2.4 Derivatives of $\sigma^2/(4(1 - \phi^2))$

Write $v(\phi, \sigma) \equiv \sigma^2/(4(1 - \phi^2))$ as a local shorthand. The power rule on σ^2 gives

$$\frac{\partial v}{\partial \sigma} = \frac{2\sigma}{4(1 - \phi^2)} = \frac{\sigma}{2(1 - \phi^2)}.$$

For ϕ , write $v = (\sigma^2/4)(1 - \phi^2)^{-1}$. The chain rule gives

$$\frac{\partial v}{\partial \phi} = \frac{\sigma^2}{4} \cdot (-(1 - \phi^2)^{-2}) \cdot (-2\phi) = \frac{\sigma^2 \phi}{2(1 - \phi^2)^2}.$$

Putting these together,

$$\frac{\partial v}{\partial \phi} = \frac{\sigma^2 \phi}{2(1 - \phi^2)^2}, \quad \frac{\partial v}{\partial \sigma} = \frac{\sigma}{2(1 - \phi^2)}. \quad (\text{D.11})$$

D.2.5 Derivatives of the FLOM Ratio

The FLOM ratio in (D.5) is the fraction $\tan(p\theta)/\tan(p\pi/2)$. The denominator does not depend on ψ , so we differentiate the numerator only. Using $\tan'(u) = \sec^2(u) = 1 + \tan^2(u)$ and the chain rule with inner derivative $p \partial\theta/\partial\cdot$,

$$\frac{\partial}{\partial \alpha} \left(\frac{\tan(p\theta)}{\tan(p\pi/2)} \right) = \frac{p(1 + \tan^2(p\theta))}{\tan(p\pi/2)} \frac{\partial \theta}{\partial \alpha}, \quad \frac{\partial}{\partial \beta} \left(\frac{\tan(p\theta)}{\tan(p\pi/2)} \right) = \frac{p(1 + \tan^2(p\theta))}{\tan(p\pi/2)} \frac{\partial \theta}{\partial \beta}. \quad (\text{D.12})$$

We denote these two derivatives by r_α and r_β , respectively.

D.2.6 Derivatives of the Sign Condition

The theoretical part of the sign condition in (D.4) is $\mathbb{E}[\text{sgn}(y_t)] = 2\theta/\pi$ (Proposition 1), which depends on ψ only through θ . By the chain rule and (D.7),

$$\frac{\partial}{\partial \alpha} \left(\frac{2\theta}{\pi} \right) = \frac{2}{\pi} \frac{\partial \theta}{\partial \alpha}, \quad \frac{\partial}{\partial \beta} \left(\frac{2\theta}{\pi} \right) = \frac{2}{\pi} \frac{\partial \theta}{\partial \beta}. \quad (\text{D.13})$$

D.3 Assembling the Jacobian

The Jacobian $G(\psi)$ has $M = K + 3$ rows and 5 columns $(\mu, \phi, \sigma, \alpha, \beta)$. Since the sample part of each moment is free of ψ , row i of G is the negative gradient of the theoretical part of g_i . We proceed row by row. Substituting the results from Section D.2:

Row g_1 (mean). This part is for $\mu/2 + \mathbb{E}[\tilde{\varepsilon}_t]$:

$$\frac{\partial g_1}{\partial \mu} = -\frac{1}{2}, \quad \frac{\partial g_1}{\partial \phi} = 0, \quad \frac{\partial g_1}{\partial \sigma} = 0, \quad \frac{\partial g_1}{\partial \alpha} = -\frac{\partial \mathbb{E}[\tilde{\varepsilon}_t]}{\partial \alpha}, \quad \frac{\partial g_1}{\partial \beta} = -\frac{\partial \mathbb{E}[\tilde{\varepsilon}_t]}{\partial \beta},$$

with the last two given by (D.8)–(D.9).

Row g_2 (variance). This part is for $v(\phi, \sigma) + \text{Var}[\tilde{\varepsilon}_t]$:

$$\frac{\partial g_2}{\partial \mu} = 0, \quad \frac{\partial g_2}{\partial \phi} = -\frac{\partial v}{\partial \phi}, \quad \frac{\partial g_2}{\partial \sigma} = -\frac{\partial v}{\partial \sigma}, \quad \frac{\partial g_2}{\partial \alpha} = -\frac{\partial \text{Var}[\tilde{\varepsilon}_t]}{\partial \alpha}, \quad \frac{\partial g_2}{\partial \beta} = -\frac{\partial \text{Var}[\tilde{\varepsilon}_t]}{\partial \beta},$$

with the partial derivatives from (D.11) and (D.10).

Row g_{2+k} (autocovariance at lag k). The part is for $v(\phi, \sigma) \phi^k$:

$$\frac{\partial}{\partial \phi} [v \phi^k] = \frac{\partial v}{\partial \phi} \phi^k + v k \phi^{k-1}.$$

The σ derivative is $\partial v / \partial \sigma \cdot \phi^k$. Hence

$$\frac{\partial g_{2+k}}{\partial \mu} = 0, \quad \frac{\partial g_{2+k}}{\partial \phi} = -\frac{\partial v}{\partial \phi} \phi^k - v k \phi^{k-1}, \quad \frac{\partial g_{2+k}}{\partial \sigma} = -\frac{\partial v}{\partial \sigma} \phi^k,$$

and $\partial g_{2+k} / \partial \alpha = \partial g_{2+k} / \partial \beta = 0$ since the theoretical part is independent of α, β .

Row g_M (sign variant). This part is $2\theta/\pi$ and $\text{sgn}(y_t)$ does not depend on ψ :

$$\frac{\partial g_M}{\partial \mu} = \frac{\partial g_M}{\partial \phi} = \frac{\partial g_M}{\partial \sigma} = 0, \quad \frac{\partial g_M}{\partial \alpha} = -\frac{2}{\pi} \frac{\partial \theta}{\partial \alpha}, \quad \frac{\partial g_M}{\partial \beta} = -\frac{2}{\pi} \frac{\partial \theta}{\partial \beta},$$

substituting (D.13).

Row g_M (ratio variant). Averaging (D.5) over t gives $T^{-1} \sum_t y_t^{(p)} - m_p \tan(p\theta) / \tan(p\pi/2)$, where only the FLOM ratio depends on ψ :

$$\frac{\partial g_M}{\partial \mu} = \frac{\partial g_M}{\partial \phi} = \frac{\partial g_M}{\partial \sigma} = 0, \quad \frac{\partial g_M}{\partial \alpha} = -m_p r_\alpha, \quad \frac{\partial g_M}{\partial \beta} = -m_p r_\beta,$$

substituting (D.12).

D.4 The Full Jacobian Matrix

Substituting all derivatives into the row layout of Section D.3, and writing the columns in the order $(\mu, \phi, \sigma, \alpha, \beta)$, the Jacobian $G(\psi)$ is the $(K+3) \times 5$ matrix

$$G(\psi) = \begin{pmatrix} -\frac{1}{2} & 0 & 0 & -\frac{\partial \mathbb{E}[\tilde{\varepsilon}_t]}{\partial \alpha} & -\frac{\partial \mathbb{E}[\tilde{\varepsilon}_t]}{\partial \beta} \\ 0 & -\frac{\partial v}{\partial \phi} & -\frac{\partial v}{\partial \sigma} & -\frac{\partial \text{Var}[\tilde{\varepsilon}_t]}{\partial \alpha} & -\frac{\partial \text{Var}[\tilde{\varepsilon}_t]}{\partial \beta} \\ 0 & -\frac{\partial v}{\partial \phi}\phi - v & -\frac{\partial v}{\partial \sigma}\phi & 0 & 0 \\ 0 & -\frac{\partial v}{\partial \phi}\phi^2 - 2v\phi & -\frac{\partial v}{\partial \sigma}\phi^2 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & -\frac{\partial v}{\partial \phi}\phi^K - Kv\phi^{K-1} & -\frac{\partial v}{\partial \sigma}\phi^K & 0 & 0 \\ 0 & 0 & 0 & -m_p r_\alpha & -m_p r_\beta \end{pmatrix}, \quad (\text{D.14})$$

where rows 3 through $K+2$ are the autocovariances at lags $k = 1, \dots, K$. All quantities in this matrix are given in closed form by (D.7)–(D.12), together with the sample average m_p , evaluated at the parameter estimate $\hat{\psi}$. Substituting yields a Jacobian that matches the numerical finite-difference Jacobian to machine precision. The final row corresponds to the ratio variant; for the sign variant it is replaced by

$$\left(0, \quad 0, \quad 0, \quad -\frac{2}{\pi} \frac{\partial \theta}{\partial \alpha}, \quad -\frac{2}{\pi} \frac{\partial \theta}{\partial \beta}\right).$$

For the S-GMM estimator we drop the final row and column, giving a $(K+2) \times 4$ matrix; at $\beta = 0$ the alpha-column of rows 1 and 2 simplifies as shown in Appendix D.5.

D.5 Symmetric Special Case ($\beta = 0$)

At $\beta = 0$ we have $\arctan(0) = 0$, hence $\theta = 0$. In turn $\tan(\alpha\theta) = 0$, $\tan(p\theta) = 0$, and $\log \cos(\alpha\theta) = 0$. The denominator $1 + \beta^2 \tan^2(\pi\alpha/2)$ in the expressions for $\partial A/\partial \alpha$ and $\partial A/\partial \beta$ reduces to 1, so no singularity arises. Evaluating each derivative from Section D.2

at $\beta = 0$:

$$\begin{array}{ll}
\left. \frac{\partial \theta}{\partial \alpha} \right|_{\beta=0} = 0 & (*), & \left. \frac{\partial \theta}{\partial \beta} \right|_{\beta=0} = \frac{\tan(\pi\alpha/2)}{\alpha}, \\
\left. \frac{\partial \mathbb{E}[\tilde{\varepsilon}_t]}{\partial \alpha} \right|_{\beta=0} = -\frac{\gamma_{\text{Euler}}}{\alpha^2}, & & \left. \frac{\partial \mathbb{E}[\tilde{\varepsilon}_t]}{\partial \beta} \right|_{\beta=0} = 0, \\
\left. \frac{\partial \text{Var}[\tilde{\varepsilon}_t]}{\partial \alpha} \right|_{\beta=0} = -\frac{\pi^2}{3\alpha^3}, & & \left. \frac{\partial \text{Var}[\tilde{\varepsilon}_t]}{\partial \beta} \right|_{\beta=0} = 0, \\
\left. \frac{\partial}{\partial \alpha} \left(\frac{\tan(p\theta)}{\tan(p\pi/2)} \right) \right|_{\beta=0} = 0, & & \left. \frac{\partial}{\partial \beta} \left(\frac{\tan(p\theta)}{\tan(p\pi/2)} \right) \right|_{\beta=0} = \frac{p \tan(\pi\alpha/2)}{\alpha \tan(p\pi/2)}, \\
\left. \frac{\partial}{\partial \alpha} \left(\frac{2\theta}{\pi} \right) \right|_{\beta=0} = 0, & & \left. \frac{\partial}{\partial \beta} \left(\frac{2\theta}{\pi} \right) \right|_{\beta=0} = \frac{2 \tan(\pi\alpha/2)}{\pi\alpha}.
\end{array}$$

(*) via the factor β in $\partial A / \partial \alpha$.

The first three columns of the boundary Jacobian (with respect to μ, ϕ, σ) are unchanged. The α -column of rows g_1 and g_2 reduces to the symmetric-case entries $\gamma_{\text{Euler}}/\alpha^2$ and $\pi^2/(3\alpha^3)$. Only the final moment condition (sign or FLOM ratio) carries information about β at $\beta = 0$; for the sign variant, its row reduces to $(0, 0, 0, 0, -2 \tan(\pi\alpha/2)/(\pi\alpha))$, which vanishes as $\alpha \rightarrow 2$.

The implementation in our code switches to these limit expressions when $|\beta|$ falls below 10^{-15} to avoid the form $0/0$ in intermediate quantities; the result agrees with the general formula to machine precision elsewhere.