Estimation of time-varying autoregressive stochastic volatility models with stable innovations

Gernot Müller[1] · Sebastian Uhl[1]

Abstract

A new time-varying autoregressive stochastic volatility model with α-stable innovations (TVARαSV) is proposed. This new model for time series data combines a time-varying autoregressive component and a stochastic scaling as known from stochastic volatility models with α-stable distributed noise. Hence, the model can cover extreme events better than classical stochastic volatility models. Furthermore, we develop a Gibbs sampling procedure for the estimation of the model parameters. The procedure is based on the estimation strategy by Kim et al. (Rev Econ Stud 65(3): 361–393, 1998) for classical stochastic volatility models, however, the estimation procedure requires a deliberate approximation of α-stable distributions by finite mixtures of normal distributions and the application of a simulation smoother for linear Gaussian state space models. A simulation study for the new estimation procedure illustrates the appealing accuracy. Finally, we apply the model to electricity spot price data.

1 Introduction

Stochastic volatility (SV) models are widely used in financial econometrics and have been extensively studied over many years. These models date back into the 1980s, cf. Hull and White ([1987](#)). Properties were discussed, e.g., in Taylor ([1994](#)), Shephard ([1996](#)) and Ghysels et al. ([1996](#)). The papers by Jacquier et al. ([1994](#)) and Kim et al. ([1998](#)) proposed Bayesian estimation algorithms. The key feature of the SV models is the stochastic scaling of the innovations, so that periods of higher and lower volatility are accurately captured. Usually Gaussian or Student-t innovations were used in the literature. Whereas these distributions seem to be sufficient for many situations of classical financial markets, they can not handle extreme price changes (spikes) which can be observed, e.g., on electricity markets, where extreme

✉ Gernot Müller
    gernot.mueller@math.uni-augsburg.de

[1] Institute of Mathematics, University of Augsburg, Universitätsstraße 14, 86159 Augsburg, Germany

values result from the lack of efficient storage possibilities and a sudden imbalance of supply and demand. Hence, for modeling such phenomenons, distributions with heavy tails, as α-stable distributions, seem to be more appropriate. As a generalization of normal distributions, the α-stable distributions also provide the possibility to account for skewness in the empirical data.

The novel idea of this paper is to combine a time-varying autoregressive component and a stochastic scaling as known from stochastic volatility models with α-stably distributed noise. We will abbreviate this model by TVARαSV. Note that we use the notion stochastic volatility for our model due to its similarity to the classical SV models although for the non-degenerated α-stable distributions the variance as the second centered moment does not exist. Our model is related to Primiceri ([2005](#)) who generalized the autoregressive model to a time-varying structural vector autoregression model, where the time variation derives both from the coefficients and the covariance matrix of the multivariate Gaussian innovations. Casarin ([2004](#)), Lombardi and Calzolari ([2009](#)) and Vankov et al. ([2019](#)) also worked with SV-like models with α-stable innovations, however, they did not include an autoregressive structure in the observation equation.

Our primary goal is the development of an estimation procedure for the TVARαSV model parameters. The challenge for the use of α-stable distributions is that their moments of order greater than or equal to α do not exist, for $\alpha \neq 2$, and their density function cannot be described in closed form. For this reason, maximum likelihood or moment-based estimation procedures are not applicable. In order to estimate parameters of similar models with α-stable distributions, mainly approximate Bayesian computation (Martin et al. 2019; Vankov et al. 2019), likelihood-free Bayesian inference (Peters et al. 2012) or indirect inference methods (Calzolari et al. 2014; Lombardi and Calzolari 2009) were used in the existing literature. Our new proposed estimation procedure for the TVARαSV model parameters, and in general for the class of SV-like models with α-stable distributions, is also carried out in a Bayesian framework, leading to a Markov chain Monte Carlo algorithm (Brooks et al. 2011). The core idea is the approximation of intractable innovation distributions by finite mixtures of normal distributions and the application of a simulation smoother for linear Gaussian state space models. This approach was developed by Shephard (1994), De Jong and Shephard (1995) and Kim et al. (1998). Since we have to approximate α-stable densities, we will use not only classical minimum distance mixture approximations but also additionally employ specific component distributions, which we will call tail components, to improve the approximation of the tails of the α-stable distributions.

describing electricity prices requires an autoregressive model which accounts for time-variation in the parameters and allows for extreme price changes as well.

The paper is organized as follows. In Sect. 2 we introduce the TVARαSV model. A suitable estimation method of the TVARαSV model is presented in Sect. 3. We investigate the quality of the proposed estimation algorithm in a simulation study in Sect. 4. In Sect. 5 we apply the estimation procedure of the TVARαSV model to electricity spot price data.

Regarding notation, we use lower case letters to denote random variables, whereas capital letters are used for their realizations. Also, we shall use $x_{1:n}$ to denote the set $\{x_1, \ldots, x_n\}$. The notation $p(\cdot)$ is used to denote a generic probability density. In case of continuous distributions the reference measure is the Lebesgue measure, whereas the counting measure is used for distributions with discrete support.

2 The time-varying autoregressive stochastic volatility model with stable innovations (TVARαSV)

Following the representation of Samorodnitsky and Taqqu (1994), a random variable x is said to have an α-stable distribution, i.e. $x \sim \mathcal{S}(\alpha, \beta, \gamma, \delta)$, with $\alpha \in (0, 2]$, $\beta \in [-1, 1]$, $\gamma > 0$ and $\delta \in \mathbb{R}$, if it has the characteristic function

$$\mathrm{E}\exp(iux) = \begin{cases} \exp\left(-\gamma^\alpha |u|^\alpha [1 - i\beta(\operatorname{sign} u)\tan\frac{\pi\alpha}{2}] + i\delta u\right), & \text{for } \alpha \neq 1 \\ \exp\left(-\gamma |u|[1 + i\beta\frac{2}{\pi}(\operatorname{sign} u)\log|u|] + i\delta u\right), & \text{for } \alpha = 1 \end{cases}. \tag{1}$$

Benth et al. (2014) and Müller and Seibert (2019) have dealt with modeling electricity prices. They have shown that long-term non-stationarities, i.e. significant price level deviations from the mean price after removing seasonalities and short-term effects, are present in the data. Benth et al. (2014) take this into account by using a Lévy process in addition to the CARMA(2,1) (continuous-time autoregressive moving average) model from García et al. (2011). Indeed, prices show some autoregressive behavior due to the fact that the circumstances leading to higher or lower prices, e.g., weather or power plant failures, usually last for more than one day. The current structural change in energy supply, mainly due to the expansion of renewable energies, requires additional flexibility of the model. Our application of the TVARαSV model to electricity spot price data in Sect. 5 shows that such structural breaks can be modeled well with time-varying parameters. For this reason, classical autoregressive models or autoregressive-moving-average models with constant parameters seem not to be sufficiently flexible. Therefore,

The characteristic parameter α describes the tail thickness. The other three parameters β, γ, and δ are called skewness parameter, scale parameter, and location parameter, respectively. For $\alpha \neq 2$ an α-stable random variable is symmetric about δ if and only if $\beta = 0$. Moreover, for $\alpha \neq 2$ an α-stable distribution is right-skewed if $\beta > 0$ and left-skewed if $\beta < 0$.

The probability densities of α-stable random variables exist and are continuous; however, with a few exceptions, they are not known in closed form. The most famous exception is the normal distribution, $\mathcal{N}\left(\mu, \sigma^2\right) = \mathcal{S}(2, \beta, \frac{\sigma}{\sqrt{2}}, \mu)$, with arbitrary β, since for $\alpha = 2$ it follows that $\tan\left(\frac{\pi\alpha}{2}\right) = 0$ in the characteristic function (1). For $\alpha \neq 2$, α-stable distributions have infinite variance and when $\alpha \leq 1$, they have an infinite mean as well.

A proof of the following property is presented in (Samorodnitsky and Taqqu 1994, Property 1.2.2 and 1.2.3 on page 11): Let $x \sim \mathcal{S}(\alpha, \beta, \gamma, \delta)$, c be a real constant and a be a real non-zero constant. Then

$$ax + c \sim \begin{cases} \mathcal{S}(\alpha, (\text{sign } a)\beta, |a|\gamma, a\delta + c), & \text{for } \alpha \neq 1 \\ \mathcal{S}(\alpha, (\text{sign } a)\beta, |a|\gamma, a\delta - \frac{2}{\pi}a(\log|a|)\gamma\beta + c), & \text{for } \alpha = 1 \end{cases}. \tag{2}$$

In the following, we introduce the time-varying autoregressive stochastic volatility model with α-stable innovations and discuss some important properties.

Definition 1 (TVARαSV) A process $\{y_t\}_{t \in \mathbb{N}_0}$ is called time-varying autoregressive stochastic volatility process with α-stable innovations, if and only if it satisfies for every $t \in \mathbb{N}$ the equation

$$y_t = \phi_t y_{t-1} + e^{\frac{h_t}{2}}\varepsilon_t, \qquad \varepsilon_t \sim \mathcal{S}(\alpha, \beta, 1, 0), \tag{3}$$

such that the dynamic of the time-varying autoregressive coefficient $\phi_t \in \mathbb{R}$ and the time-varying scale parameter $h_t \in \mathbb{R}$ is determined by the equations

$$\phi_t = \phi_{t-1} + \sigma_\phi \xi_{t,\phi}, \qquad \xi_{t,\phi} \sim \mathcal{N}(0, 1), \tag{4}$$
$$h_t = h_{t-1} + \sigma_h \xi_{t,h}, \qquad \xi_{t,h} \sim \mathcal{N}(0, 1), \tag{5}$$

where all innovations ε_t, $\xi_{t,\phi}$ and $\xi_{t,h}$ are mutually independent. The standard deviations σ_ϕ and σ_h are assumed to be positive real numbers.

The essential feature of this model is that both heteroskedasticity and structural changes can be captured. As in the prominent stochastic volatility models the noise ε_t is stochastically scaled via the time-varying parameter h_t. In particular, in view of (2) the conditional distribution of y_t, conditioned on y_{t-1}, ϕ_t, h_t, α and β, is given by

$$y_t | y_{t-1}, \phi_t, h_t, \alpha, \beta$$
$$\sim \begin{cases} \mathcal{S}\left(\alpha, \beta, e^{\frac{h_t}{2}}, \phi_t y_{t-1}\right), & \alpha \neq 1 \\ \mathcal{S}\left(\alpha, \beta, e^{\frac{h_t}{2}}, \phi_t y_{t-1} - \frac{\beta h_t}{\pi}e^{\frac{h_t}{2}}\right), & \alpha = 1 \end{cases}, \tag{6}$$

so that extreme observations can be modeled. Furthermore, we can describe data with asymmetric heavy tails. For identifiability reasons we set the scale and the location parameter of the α-stable distribution in (3) equal to 1 and 0, respectively. Replacing the distribution of ε_t by $\mathcal{S}(\alpha, \beta, \gamma, 0)$, where $\gamma > 0$, or by $\mathcal{S}(\alpha, \beta, 1, \delta)$ (or by the combination of both) would cause an identifiability problem: In the first case define the random variables $h_t^* := h_t + \log \gamma^2$. Together with $\varepsilon_t^* \sim \mathcal{S}(\alpha, \beta, 1, 0)$ this leads to a TVARαSV process which is equivalent to the TVARαSV process where h_t and $\varepsilon_t \sim \mathcal{S}(\alpha, \beta, \gamma, 0)$ is used. In the latter case exploit the recursive definition $y_t^* := y_t + \phi_t(y_{t-1}^* - y_{t-1}) - e^{h_t/2}\delta$. Again,

together with $\varepsilon_t^* \sim \mathcal{S}(\alpha, \beta, 1, 0)$ this leads to a TVARαSV process which is equivalent to the TVARαSV process where y_t and $\varepsilon_t \sim \mathcal{S}(\alpha, \beta, 1, \delta)$ is used. For the remaining parameters it is not straightforward to check identifiability. However, from our simulation studies, with randomly drawn starting parameters from reasonably chosen intervals but posterior mean estimates close to the true values, there is evidence that we are not facing any identifiability problem here.

3 An estimation procedure of the TVARαSV model

3.1 General description

In this section we develop a Gibbs sampling procedure for estimating the unknown parameters $\phi_{1:T}$, $h_{1:T}$, σ_ϕ^2, σ_h^2, α, β given data $y_{0:T}$ of the TVARαSV model. When deriving full conditional distributions we mostly suppress all parameters and variables from which the parameter to update is independent from, for notational convenience. We estimate $\phi_{1:T}$ and $h_{1:T}$ using the simulation smoother of De Jong and Shephard (1995). Since this simulation smoother appears, dependent on the situation at hand, in many different forms, we briefly summarize the corresponding formulas required for our analysis in the Appendix. The simulation smoother requires to approximate the TVARαSV model by a linear Gaussian state space model.

Therefore an important step in our estimation approach is the approximation of α-stable distributions $\mathcal{S}(\alpha, \beta, 1, 0)$ by finite mixtures of normal distributions. Obviously, a standard minimum distance approach to determine the weights and parameters of the component distributions will not lead to satisfying results since it only approximates the center of the distribution accurately. The tails which play an important role for our model are hardly taken into account. Therefore, we use mixture components which are designed specifically to approximate the tails of the α-stable distribution by fitting corresponding extreme quantiles. We call these components tails components. For the choice of the tail components see Sect. 3.2.

Since the parameters α and β are updated in each iteration of the Gibbs sampler, the α-stable distribution we have to approximate changes in each iteration. To shorten the computation time, we decided to compute the parameters of the mixture approximations, i.e. weights, means and variances of the components, before starting the algorithm. To this end

we discretize the parameter space of the parameters α and β by defining the set

$$\mathfrak{P} = \big\{(\alpha, \beta) \mid \alpha \in \{1.1, 1.105, 1.110, \ldots, 2\},$$
$$\beta \in \{-1, -0.995, -0.990, \ldots, 1\}\big\}.$$

In this discretized parameter space, the range of α is [1.1, 2], since our developed estimation algorithm in its execution in Sects. 4 and 5 at no time sampled values for α smaller than 1.1. For this reason, we have so far only calculated the parameters for the approximations for α greater than or equal to 1.1 (of course, it is straightforward to calculate parameters for the approximations also for α smaller than 1.1). The parameters of the mixture approximations for these combinations of α and β are stored in a file and can be used for different runs of the Gibbs sampler and different data analyses as well. Of course the accuracy of the discretization can be chosen arbitrarily. However, in our simulation study it turns out that we obtain accurate results using this discretization.

3.2 Estimation of $\phi_{1:T}$

For estimation of the autoregressive parameter ϕ_t we use the state space model given by the equations (3) and (4), whereby we approximate the distribution of the random variable ε_t for fixed α and β, which we write as $(\varepsilon_t|\alpha, \beta)$, by a seven-component mixture of normal distributions:

$$p(\varepsilon_t|\alpha, \beta) \approx \sum_{i=1}^{7} \Pr(\tau_t = i|\alpha, \beta) \cdot p(u_t|\tau_t = i, \alpha, \beta), \quad (7)$$

where $u_t|\tau_t = i, \alpha, \beta$, for $i = 1, \ldots, 7$, denote normally distributed random variables with mean $\mu(\tau_t, \alpha, \beta)$ and variance $\sigma^2(\tau_t, \alpha, \beta)$ and $\Pr(\tau_t = i|\alpha, \beta)$ denote the mixture weights. Depending on the α-stable parameters α and β let $\tau_t \in \{1, \ldots, 7\}$ indicate the component which occurs at time t. Thus the mixture weight $\Pr(\tau_t = i|\alpha, \beta)$ is the prior density of τ_t and reflects the probability that the ith component occurs at time t. Furthermore, we suppose that all normal random variables

$$\big\{u_t|\tau_t, \alpha, \beta \mid t = 1, \ldots, T, \ \tau_t = 1, \ldots, 7, \ (\alpha, \beta) \in \mathfrak{P}\big\}$$

are independent. For every combination $(\alpha, \beta) \in \mathfrak{P}$ the weights and the parameters of the normally distributed components are determined in two steps.

First, we only care about the tail components, which are the sixth and the seventh component. In order to cover the tails of the α-stable distribution $(\varepsilon_t|\alpha, \beta)$ we use the mean and the variance of the sixth component $(u_t|\tau_t = 6, \alpha, \beta)$ such that both the 5×10^{-7}-quantile and $1 - 5 \times 10^{-7}$-quantile of the sixth component and the α-stable distribution coincide.

We use $\Pr(\tau_t = 6|\alpha, \beta) = 0.0009$, and the mean $\mu(\tau_t, \alpha, \beta)$ is chosen as the midpoint between the two quantiles. Let $q_{\varepsilon_t}(\cdot)$ be the quantile function of the α-stable random variable $\varepsilon_t|\alpha, \beta$ and let $q(\cdot)$ be the quantile function of a standard normal variable. Then the required variance can be calculated by

$$\sigma^2(\tau_t = 6, \alpha, \beta) = \left(\frac{q_{\varepsilon_t}(1 - 5 \times 10^{-7}) - q_{\varepsilon_t}(5 \times 10^{-7})}{2 \cdot q\left(\frac{5 \times 10^{-7}}{\Pr(\tau_t = 6|\alpha, \beta)}\right)}\right)^2.$$

Then we repeat the same procedure for the seventh component with weight $\Pr(\tau_t = 7|\alpha, \beta) = 0.0001$ and for the 5×10^{-10}-quantile and the $1 - 5 \times 10^{-10}$-quantile. In our simulation studies we found using two tail components to be sufficient to achieve accurate estimation results in our setup. Of course, in other situations it might be advantageous to add more tail components to cover the tails of the α-stable distribution even more accurately.

Second, the weights and the parameters of the five remaining components are estimated using a minimum distance approach (see Titterington et al. 1985, Section 4.5). Let $f_{\mathcal{S}}(\cdot|\alpha, \beta, 1, 0)$ be the density function of the α-stable random variable $\varepsilon_t|\alpha, \beta$. As support for the minimum distance estimation we choose $\mathcal{X} = \{q_{\varepsilon_t}(x) \mid x \in \{0.002, 0.006, 0.01, \ldots, 0.998\}\}$ and the L_2-norm as distance function. To this end, we have to find a solution for the optimization problem

$$\min_{\psi \in \Psi} \sum_{x \in \mathcal{X}} \left[f_{\mathcal{S}}(x|\alpha, \beta, 1, 0) - \sum_{i=1}^{7} \frac{\Pr(\tau_t = i|\alpha, \beta)}{\sqrt{2\pi\sigma^2(\tau_t = i, \alpha, \beta)}} \right.$$
$$\left. \exp\left(-\frac{(x - \mu(\tau_t = i, \alpha, \beta))^2}{2\sigma^2(\tau_t = i, \alpha, \beta)}\right) \right]^2$$

with parameter space

$$\Psi = \left\{ \left. \Big\{\Pr(\tau_t|\alpha, \beta), \mu(\tau_t, \alpha, \beta), \sigma^2(\tau_t, \alpha, \beta)\Big\}_{\tau_t=1}^{5} \right| \right.$$
$$0 \leq \Pr(\tau_t|\alpha, \beta) \leq 1, \mu(\tau_t, \alpha, \beta) \in \mathbb{R},$$
$$\sigma^2(\tau_t, \alpha, \beta) \in \mathbb{R}^+ \text{ for } \tau_t = 1, \ldots, 5 \text{ and}$$
$$\left. \sum_{i=1}^{5} \Pr(\tau_t = i|\alpha, \beta) = 0.999 \right\}.$$

We have implemented this optimization problem in MATLAB (2020) using the function fmincon with the stopping criterion that the L_2-norm must be less than 0.02. For a more detailed examination of the approximation quality using the L_2-norm as well as the Hellinger distance we refer to Sect. A of the Appendix.

Hence we can simulate $\phi_{1:T}$ from $p(\phi_{1:T}|y_{0:T}, h_{1:T}, \sigma_\phi^2, \tau_{1:T}, \alpha, \beta)$ by using the simulation smoother with respect to the linear Gaussian state space model

$$y_t = e^{\frac{h_t}{2}} \mu(\tau_t, \alpha, \beta) + \phi_t y_{t-1} + e^{\frac{h_t}{2}} \left(\sigma^2(\tau_t, \alpha, \beta)\right)^{\frac{1}{2}} \tilde{u}_t,$$
$$\tilde{u}_t \sim \mathcal{N}(0, 1), \tag{8}$$
$$\phi_t = \phi_{t-1} + \sigma_\phi \xi_{t,\phi}, \xi_{t,\phi} \sim \mathcal{N}(0, 1), \tag{9}$$

using the centered and standardized version $\tilde{u}_{1:T}$ of $u_{1:T}$. The application of the simulation smoother requires an initial value ϕ_0. We use a constant improper prior, $p(\phi_0) = 1$, so that we can draw samples of ϕ_0 from the posterior distribution $\mathcal{N}(\phi_1, \sigma_\phi^2)$, where we use ϕ_1 from the last iteration of the Gibbs sampler (or its starting value at the beginning: this could be any reasonable fixed hyperparameter or a randomly drawn value depending of the specific situation at hand, cf. the simulation study in Sect. 4).

3.3 Estimation of $h_{1:T}$

For estimation of the scale parameter h_t we transform equation (8) which leads to the state space model

$$\log (y_t - \phi_t y_{t-1})^2 = h_t + \log \left(u_t^2\right),$$
$$u_t | \tau_t, \alpha, \beta \sim \mathcal{N}\left(\mu(\tau_t, \alpha, \beta), \sigma^2(\tau_t, \alpha, \beta)\right), \tag{10}$$
$$h_t = h_{t-1} + \sigma_h \xi_{t,h}, \quad \xi_{t,h} \sim \mathcal{N}(0, 1). \tag{11}$$

For fixed t, α and β there are seven different normally distributed random variables u_t indicated by τ_t. Since equations (10) and (11) define a non-Gaussian state space model, we approximate the distribution of the random variable $\log \left(u_t^2\right)$ for fixed τ_t, α and β again by a seven-component mixture of normal distributions, similar to Shephard (1994), where the transformation of a standard normal distribution was approximated:

$$p\left(\log \left(u_t^2\right)\Big| \tau_t, \alpha, \beta\right)$$
$$\approx \sum_{j=1}^{7} \Pr(\omega_t = j | \tau_t, \alpha, \beta) \cdot p(v_t | \omega_t = j, \tau_t, \alpha, \beta). \tag{12}$$

Here $v_t | \omega_t = j, \tau_t, \alpha, \beta$, for $j = 1, \dots, 7$, $t = 1, \dots, T$, $\tau_t = 1, \dots, 7$, $(\alpha, \beta) \in \mathfrak{P}$, are independent and normally distributed with mean $\mu(\omega_t, \tau_t, \alpha, \beta)$ and variance $\sigma^2(\omega_t, \tau_t, \alpha, \beta)$ and $\Pr(\omega_t = j | \tau_t, \alpha, \beta)$ denote the mixture weights. For $t = 1, \dots, T$ the variables ω_t define independently distributed integer valued random variables with prior $\Pr(\omega_t | \tau_t, \alpha, \beta)$. Using again the described minimum distance approach in Sect. 3.2 with the same distance function and stopping criteria we estimate the weights and the

parameters of the normal distributions appearing in equation (12) for each possible combination $(\alpha, \beta, \tau_t) \in \mathfrak{P} \times \{1, \dots, 7\}$. As support for the minimum distance approach the $0.002, 0.006, 0.01, \dots, 0.998$-quantiles of the distribution of $\log \left(u_t^2\right)$ are also used.

By defining $y_t^* = \log (y_t - \phi_t y_{t-1})^2$ we finally obtain the linear Gaussian state space model

$$y_t^* = \mu(\omega_t, \tau_t, \alpha, \beta) + h_t + \left(\sigma^2(\omega_t, \tau_t, \alpha, \beta)\right)^{\frac{1}{2}} \tilde{v}_t,$$
$$\tilde{v}_t \sim \mathcal{N}(0, 1), \tag{13}$$
$$h_t = h_{t-1} + \sigma_h \xi_{t,h}, \xi_{t,h} \sim \mathcal{N}(0, 1). \tag{14}$$

The application of the simulation smoother to the equations (13) and (14) provides samples from $p(h_{1:T}|y_{0:T}, \phi_{1:T}, \sigma_h^2, \omega_{1:T}, \tau_{1:T}, \alpha, \beta)$, where we assume a constant improper prior, $p(h_0) = 1$, so that we can draw samples of h_0 from the posterior distribution $\mathcal{N}(h_1, \sigma_h^2)$, where we use h_1 from the last iteration of the Gibbs sampler (or its starting value at the beginning: this could be any reasonable fixed hyperparameter or a randomly drawn value depending of the specific situation at hand, cf. the simulation study in Sect. 4).

3.4 Estimation of the parameters $\tau_{1:T}$, $\omega_{1:T}$, σ_ϕ^2, σ_h^2, α and β

The unknown mixture indices $\tau_{1:T}$ and $\omega_{1:T}$, the variances σ_ϕ^2 and σ_h^2 and the parameters of the α-stable distribution α and β must also be updated in every iteration of the Gibbs sampler, whereby the general estimation procedure is specified here and technical details leading to the conditional distributions can be found in the Appendix.

Since the mixture indices $\tau_{1:T}$ (resp. $\omega_{1:T}$) are conditionally independent, we can draw τ_t (resp. ω_t) separately. For this purpose we evaluate, for $\tau_t, \omega_t = 1, \dots, 7$, the following densities

$$p(\tau_t | y_{0:T}, \phi_{1:T}, h_{1:T}, \alpha, \beta)$$
$$\propto p(y_t | y_{t-1}, \phi_t, h_t, \tau_t, \alpha, \beta) \cdot \Pr(\tau_t | \alpha, \beta)$$

and

$$p(\omega_t | y_{0:T}, \phi_{1:T}, h_{1:T}, \tau_{1:T}, \alpha, \beta)$$
$$\propto p(y_t^* | h_t, \omega_t, \tau_t, \alpha, \beta) \cdot \Pr(\omega_t | \tau_t, \alpha, \beta),$$

where y_t and y_t^* conditioned on the chosen mixture components are normally distributed according to equation (8) and (13). After standardizing these probabilities we can draw $\tau_t \in \{1, \dots, 7\}$ and $\omega_t \in \{1, \dots, 7\}$.

We assume that the prior distributions for σ_ϕ^2 and σ_h^2 are the inverse gamma distributions $\mathcal{IG}(s_\phi, r_\phi)$ and $\mathcal{IG}(s_h, r_h)$,

where $s_\phi, s_h > 0$ are the shape parameters and $r_\phi, r_h > 0$ the scale parameters. Due to the use of conjugate priors, one obtains the following inverse-gamma distributions as posterior distributions:

$$\sigma_\phi^2 | \phi_{1:T} \sim \mathcal{IG} \left(s_\phi + \frac{T-1}{2} \ , \ r_\phi + \frac{1}{2} \sum_{j=2}^{T} (\phi_j - \phi_{j-1})^2 \right),$$

$$\sigma_h^2 | h_{1:T} \sim \mathcal{IG} \left(s_h + \frac{T-1}{2} \ , r_h + \frac{1}{2} \sum_{j=2}^{T} (h_j - h_{j-1})^2 \right).$$

It turned out that the accuracy of the estimation of σ_ϕ^2 (resp. σ_h^2) can be further improved, when we scale always the new sample of σ_ϕ^2 (resp. σ_h^2) in such a way that the estimated residuals $\hat{\xi}_{1:T,\phi}$ (resp. $\hat{\xi}_{1:T,h}$) are standardized, cf. equation (4) and (5).

For the estimation of α and β the original observation equation (3) can be used without a mixture approximation of the α-stable distributions. We choose flat priors for α and β, i.e. $p(\alpha) = \frac{1}{2} \mathbb{1}_{(0,2]}(\alpha)$ and $p(\beta) = \frac{1}{2} \mathbb{1}_{[-1,1]}(\beta)$. Hence α and β can be drawn from the posterior density

$$p(\alpha, \beta | y_{0:T}, \phi_{1:T}, h_{1:T})$$
$$\propto \prod_{t=1}^{T} p(y_t | y_{t-1}, \phi_t, h_t, \alpha, \beta) \cdot p(\alpha) \cdot p(\beta),$$

where the observation y_t is distributed according to (6). Since we can not directly draw samples from the posterior distribution of α and β, we perform one step of the random walk Metropolis-Hastings algorithm in every iteration of the Gibbs sampler. Therefore we draw a proposal of α and β from a two-dimensional normal distribution with covariance matrix Σ to be chosen by the user. If the new proposed value is rejected, we retain the last sample of α and β. Finally, the drawn samples of α and β must be rounded to the nearest values in the discretized parameter space $\mathfrak{P}$. It should not remain unmentioned that there are of course also alternative approaches to generate candidates for α and β.

3.5 Gibbs sampling procedure

In this subsection we design a suitable Gibbs sampling procedure where all unknown parameters are drawn from the corresponding full conditional distribution.

The Gibbs sampler starts with the estimation of the scale parameters $h_{1:T}$ where the distribution of $\log(u_t^2)$ is approximated by a finite mixture of normal distributions (cf. Sect. 3.3). Next we update the variance σ_h^2 and the α-stable parameters α and β.

We point out that sampling $\phi_{1:T}$ can be performed using equation (8) so that we do not require the approximation of $\log(u_t^2)$; hence, the parameters $\omega_{1:T}$ are not involved here. The same holds for sampling $\tau_{1:T}$. As a consequence, the parameters $\phi_{1:T}, \sigma_\phi^2, \tau_{1:T}$ and $\omega_{1:T}$ need to be simulated in one common block of the Gibbs sampling algorithm. It is therefore advantageous to draw from the joint distribution of $\phi_{1:T}, \sigma_\phi^2$ and $\tau_{1:T}$ and then from the conditional distribution of $\omega_{1:T}$ given $\phi_{1:T}, \sigma_\phi^2$ and $\tau_{1:T}$, using

$$p(\phi_{1:T}, \sigma_\phi^2, \tau_{1:T}, \omega_{1:T}) = p(\omega_{1:T} | \phi_{1:T}, \sigma_\phi^2, \tau_{1:T})$$
$$\cdot p(\phi_{1:T}, \sigma_\phi^2, \tau_{1:T}).$$

Due to the dependence structure of our model it is crucial for a fast convergence of the sampler to update $\phi_{1:T}, \sigma_\phi^2$ and $\tau_{1:T}$ in one block. Since it seems impossible to draw all these parameters from one single high-dimensional distribution, we employ again a Gibbs sampler to draw from the joint distribution $\phi_{1:T}, \sigma_\phi^2$ and $\tau_{1:T}$ conditional on all other parameters. After this inner Gibbs sampler has been executed in one iteration of the overall Gibbs sampler, a sample of $\phi_{1:T}, \sigma_\phi^2$ and $\tau_{1:T}$ is randomly selected, with the first draws discarded as burn-in period. Thus a suitable Gibbs sampling procedure is given as follows:

(I) Simulate $\boldsymbol{h_{1:T}}$ from $p(h_{1:T} | y_{0:T}, \phi_{1:T}, \sigma_h^2, \omega_{1:T}, \tau_{1:T}, \alpha, \beta)$.

(II) Simulate $\boldsymbol{\sigma_h^2}$ from $p(\sigma_h^2 | h_{1:T})$.

(III) Simulate $\boldsymbol{\alpha, \beta}$ from $p(\alpha, \beta | y_{0:T}, \phi_{1:T}, h_{1:T})$.

(IV) Simulate $\boldsymbol{\phi_{1:T}, \sigma_\phi^2, \tau_{1:T}, \omega_{1:T}}$.

    (1) Simulate $\boldsymbol{\phi_{1:T}, \sigma_\phi^2, \tau_{1:T}}$ using Gibbs sampling.

        (i) Simulate $\boldsymbol{\tau_{1:T}}$ from $p(\tau_{1:T} | y_{0:T}, \phi_{1:T}, h_{1:T}, \alpha, \beta)$.

        (ii) Simulate $\boldsymbol{\phi_{1:T}}$ from $p(\phi_{1:T} | y_{0:T}, h_{1:T}, \sigma_\phi^2, \tau_{1:T}, \alpha, \beta)$.

        (iii) Simulate $\boldsymbol{\sigma_\phi^2}$ from $p(\sigma_\phi^2 | \phi_{1:T})$.

    (2) Simulate $\boldsymbol{\omega_{1:T}} \mid \boldsymbol{\phi_{1:T}, \sigma_\phi^2, \tau_{1:T}}$ from $p(\omega_{1:T} | y_{0:T}, \phi_{1:T}, h_{1:T}, \tau_{1:T}, \alpha, \beta)$.

The precise algorithm is outlined in Algorithm 1, which we have completely implemented in R (R Core Team 2020). Similarly to Del Negro and Primiceri (2015) it is possible to add a Metropolis-Hastings step for sampling $\phi_{1:T}$ and $h_{1:T}$ in order to remove the mixture approximation error. However, this is practically not desirable here due to the enormous additional computing time (an estimated doubling of the computing time) caused by the α-stable density functions. Furthermore, Del Negro and Primiceri (2015) conclude that the results of the original and the extended algorithm are indistinguishable and the mixture approximation error is negligible.

Algorithm 1 Estimation of the TVARαSV model

Initialization: For $i = 0$.

- Initialize $\phi_{1:T}^{(i)}, h_1^{(i)}, \sigma_\phi^{2(i)}, \sigma_h^{2(i)}, \omega_{1:T}^{(i)}, \tau_{1:T}^{(i)}, \alpha^{(i)}, \beta^{(i)}$ and determine prior hyperparameters $s_h, r_h, s_\phi, r_\phi, \Sigma$.

Gibbs Sampling: For $i = 1, \ldots, N$.

- Simulate $\boldsymbol{h_{1:T}^{(i)}}$ from $p\left(h_{1:T} \big| y_{0:T}, \phi_{1:T}^{(i-1)}, \sigma_h^{2(i-1)}, \omega_{1:T}^{(i-1)}, \tau_{1:T}^{(i-1)}, \alpha^{(i-1)}, \beta^{(i-1)}\right)$ using the simulation smoother with respect to the equations (13) and (14). The start value $h_0^{(i)}$ for Kalman filtering is drawn from $\mathcal{N}\left(h_1^{(i-1)}, \sigma_h^{2(i-1)}\right)$.

- Simulate $\boldsymbol{\sigma_h^{2(i)}} | h_{1:T}^{(i)} \sim \mathcal{IG}\left(s_h + \frac{T-1}{2}, r_h + \frac{1}{2}\sum_{j=2}^{T}\left(h_j^{(i)} - h_{j-1}^{(i)}\right)^2\right)$.

- Simulate $\left(\boldsymbol{\alpha^{(i)}}, \boldsymbol{\beta^{(i)}}\right)' \sim \mathcal{N}_2\left(\left(\alpha^{(i-1)}, \beta^{(i-1)}\right)', \Sigma\right)$ and accept with probability (A3). Otherwise $\left(\alpha^{(i)}, \beta^{(i)}\right) = \left(\alpha^{(i-1)}, \beta^{(i-1)}\right)$.

- Initialize $\tilde{\phi}_{1:T}^{(0)} = \phi_{1:T}^{(i-1)}, \tilde{\sigma}_\phi^{2(0)} = \sigma_\phi^{2(i-1)}$ and do Gibbs sampling for $j = 1, \ldots, G$:

  - For $t = 1, \ldots, T$ draw $\boldsymbol{\tilde{\tau}_t^{(j)}} \in \{1, \ldots, 7\}$ according to the probabilities (A1) given $y_{0:T}, \tilde{\phi}_{1:T}^{(j-1)}, h_{1:T}^{(i)}, \alpha^{(i)}, \beta^{(i)}$.

  - Simulate $\boldsymbol{\tilde{\phi}_{1:T}^{(j)}}$ from $p\left(\phi_{1:T} \big| y_{0:T}, h_{1:T}^{(i)}, \tilde{\sigma}_\phi^{2(j-1)}, \tilde{\tau}_{1:T}^{(j)}, \alpha^{(i)}, \beta^{(i)}\right)$ using the simulation smoother with respect to the equations (8) and (9). The start value $\tilde{\phi}_0^{(j)}$ for Kalman filtering is drawn from $\mathcal{N}\left(\tilde{\phi}_1^{(j-1)}, \tilde{\sigma}_\phi^{2(j-1)}\right)$.

  - Simulate $\boldsymbol{\tilde{\sigma}_\phi^{2(j)}} | \tilde{\phi}_{1:T}^{(j)} \sim \mathcal{IG}\left(s_\phi + \frac{T-1}{2}, r_\phi + \frac{1}{2}\sum_{l=2}^{T}\left(\tilde{\phi}_l^{(j)} - \tilde{\phi}_{l-1}^{(j)}\right)^2\right)$.

- Draw $k \in \{\lfloor 0.7 \cdot G \rfloor, \ldots, G\}$ and take $\tau_{1:T}^{(i)} = \tilde{\tau}_{1:T}^{(k)}, \phi_{1:T}^{(i)} = \tilde{\phi}_{1:T}^{(k)}$ and $\sigma_\phi^{2(i)} = \tilde{\sigma}_\phi^{2(k)}$.

- For $t = 1, \ldots, T$ draw $\boldsymbol{\omega_t^{(i)}} \in \{1, \ldots, 7\}$ according to the probabilities (A2) given $y_{0:T}, \phi_{1:T}^{(i)}, h_{1:T}^{(i)}, \tau_{1:T}^{(i)}, \alpha^{(i)}, \beta^{(i)}$.

---

4 Simulation study

For assessing the accuracy of the proposed estimation procedure described in Sect. 3 we investigate 1000 data sets each consisting of $Y_{0:2500}, \phi_{1:2500}$ and $h_{1:2500}$ ($T = 2500$) of the TVARαSV model according to Definition 1 with $\sigma_\phi^2 = 0.00058$, $\sigma_h^2 = 0.011$, $\phi_0 = 0.5$, $h_0 = 2.5$, $\alpha = 1.73$, $\beta = -0.14$ and $y_0 \sim \mathcal{S}(\alpha, \beta, e^{h_0/2}, 0)$. These parameters were set equal to those found in the application of the TVARαSV model to electricity spot price data, cf. Sect. 5.

The behavior of the observed realization from the TVARαSV process depends crucially on the range of the autoregressive parameters $\phi_{1:T}$. If these leave the interval $(-1, 1)$ substantially long and/or significantly this usually leads to an explosion of the observable time series $Y_{0:T}$ (even if the process returns to the initial level again later). Since

we do not observe such an extreme behavior for electricity spot price data, we focus here in assessing the quality of our Bayesian inference procedure for realizations from the TVARαSV process, where ϕ_t takes on only values in the interval $(-1, 1)$. In Sect. D of the Appendix we investigate additionally TVARαSV processes, where ϕ_t takes values outside the interval $(-1, 1)$, too.

Since in most cases in practice one has a priori no information about the range of ϕ_t, our estimation procedure here does not use this knowledge, that ϕ_t, for all $t = 1, \ldots, T$, takes on only values in the interval (-1,1), and we apply Algorithm 1 unchanged for the 1000 simulated time series. On the other hand, the systematic consideration of TVARαSV processes with underlying $|\phi_t| < 1$, for all $t = 1, \ldots, T$, corresponds to introducing an indicator function to the prior of $\phi_{1:T}$. Hence, also the posterior of $\phi_{1:T}$ must reflect this specification. For this reason, assuming that this additional knowledge about the prior of $\phi_{1:T}$ is available, we have applied an adapted version of Algorithm 1 to the same 1000 simulated time series, which is presented in Sect. E of the Appendix. At the end, the question about the truncated posterior boils down to which prior information the practitioner wants to use. If one is convinced that the data set at hand is based on a truncated vector $\phi_{1:T}$, one must use the second version of the Gibbs sampler (cf. Sect. E of the Appendix), which includes a resampling step for $\phi_{1:T}$. If one does not want to put such a prior information on $\phi_{1:T}$, Algorithm 1 is appropriate.

For the estimation of α and β we use a normal proposal density with a diagonal matrix with 0.001 on the main diagonal as covariance matrix Σ. Moreover, we choose $s_h = r_h = s_\phi = r_\phi = 0.001$ such that we obtain quite noninformative priors. As starting values we use for each data set randomly drawn values from uniform distributions: $\sigma_\phi^{2(0)} \sim \mathcal{U}[0.00001, 0.1]$, $\phi_1^{(0)} = \phi_2^{(0)} = \cdots = \phi_T^{(0)} \sim \mathcal{U}[0.1, 0.9]$, $\alpha^{(0)} \sim \mathcal{U}[1.4, 1.9]$, $\sigma_h^{2(0)} \sim \mathcal{U}[0.00001, 0.1]$, $h_1^{(0)} \sim \mathcal{U}[-2, 7]$, $\beta^{(0)} \sim \mathcal{U}[-0.6, 0.6]$. The boundaries of the uniform distributions were determined according to the following considerations: Looking at the observations of the

Table 1 Results of the simulation study using 6000 iterations after a burn-in period of 4000 iterations

	Mean	SD	2.5% quantile	97.5% quantile
Statistics of average biases of 1000 simulations				
$\phi_{1:T}$	−0.0010	0.0166	−0.0313	0.0305
$h_{1:T}$	−0.0607	0.0417	−0.1423	0.0186
Statistics of RMSEs of 1000 simulations				
$\phi_{1:T}$	0.0984	0.0087	0.0828	0.1181
$h_{1:T}$	0.3220	0.0213	0.2821	0.3670

simulation study, it can be checked a priori whether or not the signs of the first observations are systematically alternating and whether or not there are explosive periods, which leads in our case to fixing the underlying interval of the uniform distribution for the starting values of $\phi_{1:T}$ to [0.1, 0.9]. If the time series initially shows values with alternating signs we recommend to choose negative initial values of $\phi_{1:T}$ instead. From degenerated submodels we can also guess the magnitude of the other parameters. For instance, assuming one constant value for $\phi_{1:T}$ and one constant value h for $h_{1:T}$, and $\alpha = 2$, one can estimate the magnitude of h by fitting the degenerated submodel to the truncated time series, where the truncation is used to exclude extreme observations caused by the α-stable distribution. Moreover, from recent publications using α-stable processes for modelling electricity prices (cf. Müller and Seibert 2019) we learn that even for daily peak prices the parameter α is usually greater then 1.4, so that the interval given above seems to be a reasonable choice for the starting values of α. Assuming a quite smooth behavior of $\phi_{1:T}$ and $h_{1:T}$ the chosen intervals for σ_ϕ^2 and σ_h^2 have a very large right boundary in view of the two-sigma rule. We avoid values of β close to 1 or -1, since those values would imply an extreme behavior of the noise process $\varepsilon_{1:T}$. Next we need to initialize the mixture indices $\tau_{1:T}^{(0)}$ and $\omega_{1:T}^{(0)}$. First, the mixture indices $\tau_{1:T}^{(0)}$ are independently drawn from $\{1, \ldots, 7\}$ according to the probabilities $\Pr(\tau_t | \alpha^{(0)}, \beta^{(0)})$. Afterwards, we independently draw the mixture indices $\omega_{1:T}^{(0)}$ from $\{1, \ldots, 7\}$ according to the prior probabilities $\Pr(\omega_t | \tau_t^{(0)}, \alpha^{(0)}, \beta^{(0)})$.

The Gibbs sampler in Sect. 3.5 was carried out through 10000 iterations for each of the 1000 time series. Furthermore, the inner Gibbs sampler for the simulation of $\tau_{1:T}$, $\phi_{1:T}$ and σ_ϕ^2 was carried out through 50 iterations. We kept the last 6000 draws for inference. To measure the accuracy of the posterior mean estimates of $\phi_{1:T}$ we calculated 1000 root mean square errors (RMSEs) of the form $(\sum_{t=1}^{2500} (\hat{\phi}_t - \phi_t)^2/2500)^{1/2}$ and 1000 average biases of the form $\sum_{t=1}^{2500} (\hat{\phi}_t - \phi_t)/2500$ where $\hat{\phi}_t$ denotes the posterior mean estimate of ϕ_t (analogously for $h_{1:T}$). The results are summarized in Table 1 and Table 2.

On average the RMSEs of $\phi_{1:T}$ is quite concentrated around 0.09, which is remarkably accurate given the range of

$|\phi_t| < 1$ for the simulated parameters. Similarly, the average RMSE of $h_{1:T}$ can be considered satisfactory, as the magnitude of $h_{1:T}$ is not limited and the simulated paths of $h_{1:T}$ range from -14.45 to 20.46. In view of $\sigma_h^2 \approx 19 \cdot \sigma_\phi^2$ and the correspondingly higher range of h_t also h_t is estimated quite accurately on average. From Table 2 it is easy to see that all parameters σ_ϕ^2, σ_h^2, α and β are estimated very precisely.

Additionally, we plotted estimation and convergence results of one simulation for illustrative purpose in Fig. 1. This figure supports that all parameters including the time-varying parameters ϕ_t and h_t are estimated very accurately. The 95% credibility corridors cover almost all true values. Moreover, from the last four plots of Fig. 1 it is obvious that the chains converge very quickly to the area around the true values and that the mixing behavior of the chains is very satisfying.

Finally we note that running the Algorithm in Sect. 3.5 with 10000 iterations for one data set of length 2500 takes about 28 hours on an Intel Core i7-3635QM (2.40 GHz) processor.

5 Application of the TVARαSV model to electricity spot price data

## 5.1 Description of the electricity spot prices

We consider the daily base prices for power traded at the European Power Exchange (EPEX SPOT SE). The base price is the average of the hourly market prices of one day. The price index for the German and Austrian market is called the Physical Electricity Index (Phelix). The data are available from EEX (European Energy Exchange) and cover the period from July 1, 2002 to November 18, 2017. Hence, our data set has length 5620 and is shown in the first row of Fig. 2. We summarize some key features of daily electricity spot prices in the following list, without any intention to be exhaustive in our presentation. We refer to Klüppelberg, Meyer-Brandis and Schmidt (2010), where the stylized facts of electricity spot prices are presented in detail.

(i) *Seasonality* The plot of the empirical autocorrelation function demonstrates seasonal behavior in weekly

Table 2 Results of the simulation study using 6000 iterations after a burn-in period of 4000 iterations

	True value	Statistics of posterior means of 1000 simulations				
		Mean	SD	2.5% quantile	97.5% quantile	Bias
σ_ϕ^2	0.00058	0.000528	0.000181	0.000221	0.000898	-0.000052
σ_h^2	0.011	0.010635	0.002678	0.005737	0.016151	-0.000365
α	1.73	1.704864	0.031869	1.642219	1.764899	-0.025136
β	-0.14	-0.131964	0.072848	-0.286128	-0.001879	0.008036

Fig. 1 Application of the estimation procedure presented in Sect. 3. The time series $Y_{0:2500}$ was generated according to Definition 1 (first row). In the second and third row the smooth dashed line constitutes the posterior mean of the 6000 used Gibbs sampling estimations and the solid line constitutes the true values of $\phi_{1:2500}$ and $h_{1:2500}$. The 95% confidence bound is plotted gray. In the fourth and fifth row the solid (resp. horizontal dashed) line corresponds to the estimated (resp. true) values of σ_ϕ, σ_h, α and β
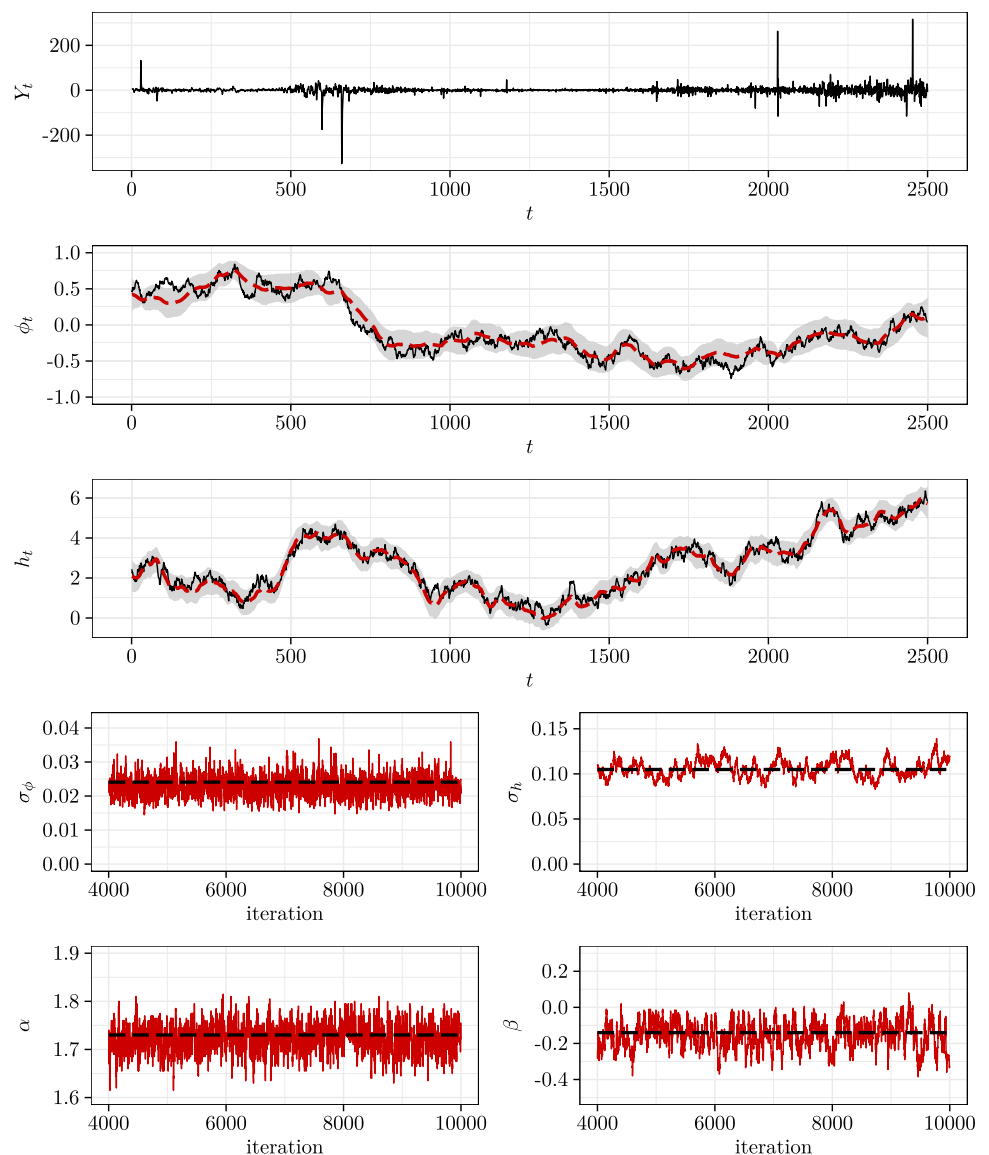
cycles (see Fig. 8 in the Appendix). Moreover, a seasonal behavior in yearly cycles can be observed (cf. Sect. F in the Appendix).

(ii) *Spikes, Skewness, Non-Gaussianity* The electricity spot prices exhibit large spikes (see Fig. 2) and skewness which cannot be modeled by a Gaussian distribution.

(iii) *Mean reversion* After attaining extreme spikes, electricity spot prices are mean reverting. The price index does not jump directly from the extreme value to the trend, but decreases step by step.

(iv) *Heteroscedasticity* A rolling window analysis of the Phelix Day Base prices shows that the standard deviation significantly changes over time (see Fig. 3).

According to the argumentation in Sect. 2 the TVARαSV model covers the stylized facts (ii), (iii) and (iv) and,

hence, seems to be appropriate for modeling these data after accounting for (i) by removing seasonal effects.

5.2 Estimation results

After estimation and elimination of trend and (weekly and yearly) seasonal components, cf. Sect. F in the Appendix, we applied our estimation procedure for the TVARαSV model (Algorithm 1) to the deseasonalized and trend adjusted electricity spot prices $\hat{Y}_t$, which are shown in the second row of Fig. 2. The Gibbs sampler in Sect. 3.5 was carried out through 10000 iterations, with the last 6000 draws being kept for inference. Furthermore, the inner Gibbs sampler for the simulation of $\tau_{1:T}$, $\phi_{1:T}$ and σ_ϕ^2 was carried out for 50 iterations each. We suppose the same prior hyperparameters as in

Fig. 2 Daily EEX Phelix Base electricity price index from July 1, 2002 to November 18, 2017 and deseasonalized and trend adjusted data $\hat{Y}_t$ (cf. Sect. F of the Appendix)
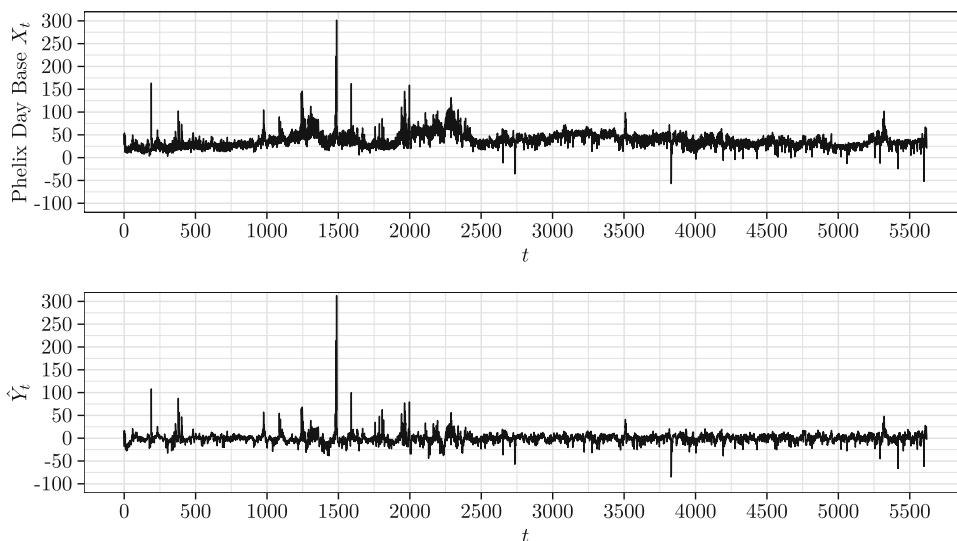

Fig. 3 Estimated standard deviations of the deseasonalized Phelix Day Base prices using a rolling window of size 101
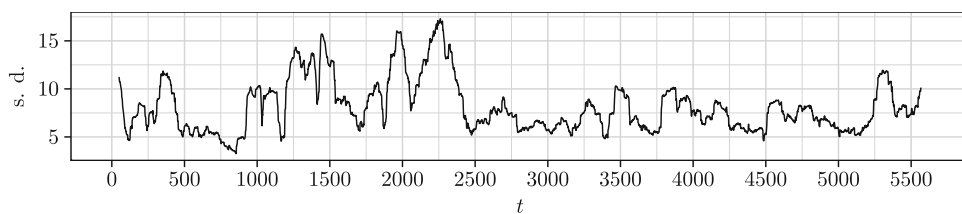

Table 3 Means and standard deviations of the marginal posterior distributions using 6000 iterations after a burn-in period of 4000 iterations

	σ_ϕ^2	σ_h^2	α	β
Mean	0.000579	0.011019	1.732964	-0.138908
SD	0.000112	0.001628	0.021264	0.058420

the simulation study of Sect. 4 and $\sigma_\phi^{2\,(0)} = 0.01$, $\phi_{1:T}^{(0)} = 0.6$, $\alpha^{(0)} = 1.5$, $\sigma_h^{2\,(0)} = 0.01$, $h_1^{(0)} = 0$, $\beta^{(0)} = 0.25$ as starting values. The results are shown in Table 3 and Fig. 4.

The marginal posterior distributions of σ_ϕ^2 and σ_h^2 support our assumption of time-varying parameters ϕ_t and h_t. The stability parameter α has a posterior mean of about 1.733 with a small standard deviation of 0.021, which indicates clearly that the distribution of the innovations ε_t has heavier tails than the Gaussian distribution. From the estimate of β we learn that the innovations are slightly negatively skewed. Figure 4 shows that there is evidence of a significant autoregressive structure in the electricity spot price data. Furthermore, rises of the scale parameter h_t coincide with periods of higher fluctuations of $\hat{Y}_t$. Investigating extreme values of the data set, the extreme spikes are created by large values of the innovations, but not by high values of the time-varying scale parameter h_t.

5.3 Model verification

In this subsection we briefly check whether some basic model assumptions are satisfied in our empirical data analysis. In particular, we look at the assumptions of normality, homoscedasticity and serial independence of $\xi_{t,\phi}$ and $\xi_{t,h}$ appearing in equations (4) and (5). To this end, we check the null of normality of the series using Jarque-Bera tests (Jarque and Bera 1980), the null of residual homoscedasticity using ARCH tests (Engle 1982), and the null of independence of the series by Ljung-Box tests (Ljung and Box 1978). In each of the 6000 iterations after the burn-in period, we first derived estimates $\hat{\xi}_{t,\phi}$ and $\hat{\xi}_{t,h}$ using the estimates of $\phi_{1:T}$, $h_{1:T}$, σ_ϕ^2 and σ_h^2, and then calculated all corresponding statistics. Each line of Table 4 evaluates the 6000 values of the test statistics by reporting the percentage which leads to a rejection of the null on the levels of 2.5%, 5%, and 10%. The table indicates that the model assumptions under investigation are satisfied.

These results are also strongly supported by Fig. 5: It shows the the autocorrelation and quantile plots for the estimated residuals $\hat{\xi}_{2:T,\phi}$ and $\hat{\xi}_{2:T,h}$ of one randomly chosen iteration in the Gibbs sampling estimation. The plots shown in Fig. 5 are a further indication that the assumption of normality and serial independence of $\xi_{t,\phi}$ and $\xi_{t,h}$ have been met.

Fig. 4 Application of the estimation procedure presented in Sect. 3 to the Phelix Day Base price $\hat{Y}_t$ after removing the trend and the seasonal component. In the first and second row the solid line constitutes the posterior mean of the 6000 used Gibbs sampling estimations of $\phi_{1:T}$ and $h_{1:T}$. The 95% confidence bound is plotted in gray
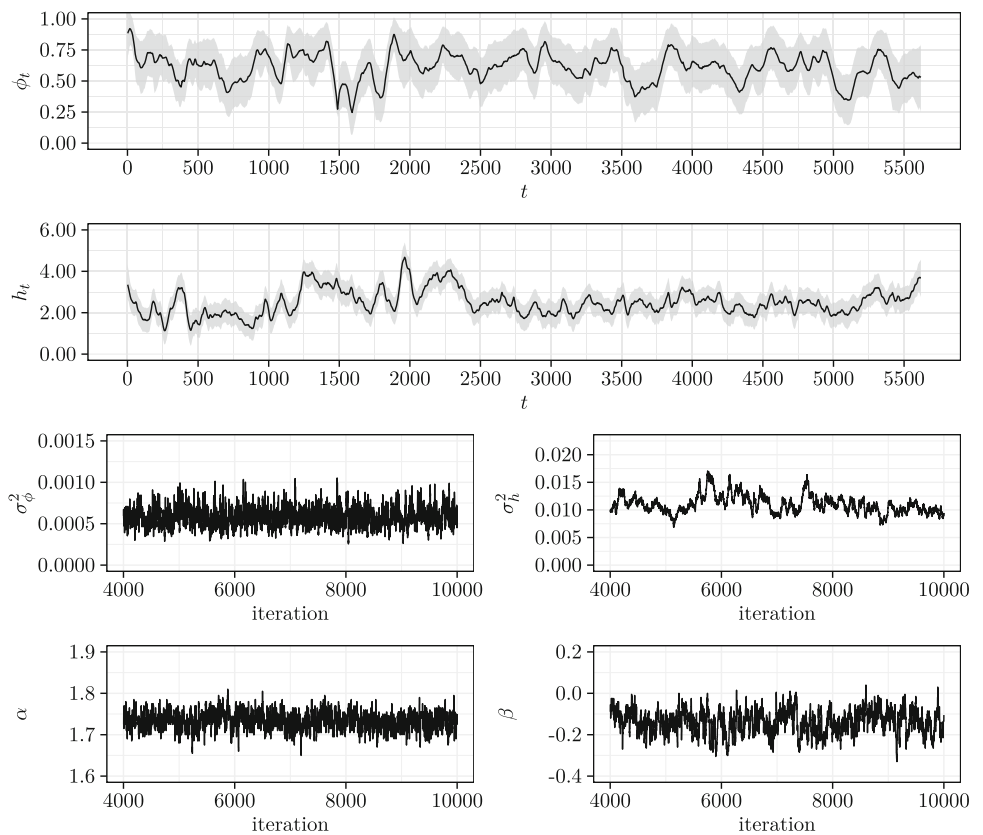
Table 4 Test results for the 6000 iterations after a burn-in period of 4000 iterations

| Test	Residuals	Percentage of rejections for a level of ...		
		2.5%	5%	10%
Jarque-Bera test	$\hat{\xi}_{t,\phi}$	2.70	5.02	9.65
	$\hat{\xi}_{t,h}$	2.33	4.80	9.93
ARCH test	$\hat{\xi}_{t,\phi}$	2.95	5.22	10.23
	$\hat{\xi}_{t,h}$	2.75	5.30	10.03
Ljung-Box test	$\hat{\xi}_{t,\phi}$	3.12	5.57	10.25
	$\hat{\xi}_{t,h}$	3.80	7.05	13.13

6 Conclusion and outlook

Motivated by the stylized facts of electricity spot prices, we introduced the time-varying autoregressive stochastic volatility model with α-stable innovations (TVARαSV). The novel idea is to combine a time-varying autoregressive component and a stochastic scaling as known from stochastic volatility models with stably distributed noise. Furthermore, we developed a Gibbs sampling procedure for the estimation of the model parameters. A key step of this estimation procedure is the approximation of α-stable densities and transformed mixture components by finite mixtures of normal distributions,

which enable the application of the simulation smoother of De Jong and Shephard (1995). In a simulation study we showed that the algorithm provides very accurate estimates. In an empirical application the TVARαSV model was fitted to the daily base prices of PHELIX. We observed that the spikes in the data are associated to the α-stable innovations, whereas the time-varying scaling parameter accounts for periods of larger fluctuations.

It is work in progress to extend the TVARαSV model in the way that it is able to discover a (classical or inverse) leverage effect, as it has been discussed, e.g., in Benth and Vos (2013) or Kristoufek (2014). However, there is evidence that the leverage (or inverse leverage) effect is quite limited in Germany (cf. Erdogdu 2016), so that this extension seems to be irrelevant for the data analyzed here. Nevertheless, adapting our model in that direction is an interesting challenge, since we use stable distributions, where introducing a leverage effect is not as straightforward as in the Gaussian case: Due to the lack of finite second moments, a correlation in its original sense does not exist for stable distributions. Instead, various modified notions of "covariance" have been proposed in the literature (see, e.g., Samorodnitsky and Taqqu 1994). In particular, the correlation structure of a multivariate stable distribution is usually determined by the spectral measure. Besides the theoretical efforts of incorporating a leverage effect into our model, also the estimation of the extended

Fig. 5 Empirical
autocorrelation functions for the
estimated residuals $\hat{\xi}_{t,\phi}$ (first
row left) and $\hat{\xi}_{t,h}$ (first row
right). QQ-plots of the estimated
residuals $\hat{\xi}_{t,\phi}$ and $\hat{\xi}_{t,h}$ against
standard normal distributions
can be seen in the second row
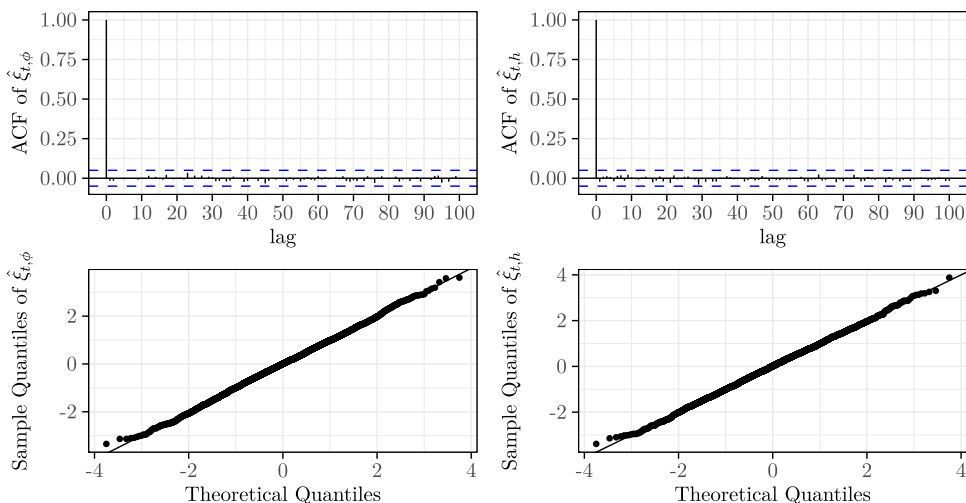
model including the spectral measure (cf. Pivato and Seco
2003) seems to be a new challenge which we defer to a future
paper.

Appendix

A Investigation of the approximation accuracy of α-stable distributions by finite mixtures of normal distributions

In our estimation procedure, the approximations of α-stable
distributions as well as of their squared and logarith-
mized normally distributed components by finite mixtures
of normal distributions are a key step for successful param-
eter estimation (cf. Sects. 3.2 and 3.3). For this reason,
the approximation accuracy will be examined in this sec-
tion. We used the L_2-norm as distance function for both
approximations with a corresponding stopping criteria and
the 0.002, 0.006, 0.01, …, 0.998-quantiles of the respective
approximated distribution as support. In addition to the L_2-
norm, the mean absolute percentage error (MAPE) and the
continuous form of the Hellinger distance is evaluated on
the support used for the optimization in order to quantify
the similarity between the true distribution and its approx-
imation: Let $f_{\mathcal{S}}(\cdot)$ be the density function of the α-stable
random variable $\varepsilon_t | \alpha, \beta$ and $f_{\mathcal{A}}(\cdot)$ the density function
of the approximation defined in equation (7). Using the
0.002, 0.006, 0.01, …, 0.998-quantiles of the distribution
of $\varepsilon_t | \alpha, \beta$ as support points $\mathcal{X} = \{x_1, x_2, x_3, \dots, x_{250}\}$, the
MAPE is defined by $\frac{1}{|\mathcal{X}|} \sum_{x \in \mathcal{X}} |f_{\mathcal{S}}(x) - f_{\mathcal{A}}(x)| / |f_{\mathcal{S}}(x)|$.

Table 5 Evaluated distance functions for the approximations to the
distributions of the random variables $(\varepsilon_t | \alpha, \beta)$ and $\left(\log \left(u_t^2 \right) \Big| \tau_t, \alpha, \beta \right)$
for $(\alpha, \beta, \tau_t) \in \mathfrak{P} \times \{1, \dots, 7\}$

	$(\varepsilon_t | \alpha, \beta)$		$\left(\log \left(u_t^2 \right) \Big| \tau_t, \alpha, \beta \right)$	
	Mean	SD	Mean	SD
L_2-norm	0.0001	0.0006	0.0005	0.0031
MAPE (%)	2.33	3.42	1.24	0.43
Hellinger distance	0.0399	0.0394	0.0129	0.0053

With the trapezoidal rule, the continuous form of the
Hellinger distance can be approximated by

$$\left(\frac{1}{2} \int_{\mathbb{R}} \left(\sqrt{f_{\mathcal{S}}(x)} - \sqrt{f_{\mathcal{A}}(x)} \right)^2 \mathrm{d}x \right)^{\frac{1}{2}} \approx \left(\frac{1}{4} \sum_{k=2}^{250} (x_k - x_{k-1}) \right.$$
$$\left. \left(\left(\sqrt{f_{\mathcal{S}}(x_{k-1})} - \sqrt{f_{\mathcal{A}}(x_{k-1})} \right)^2 + \left(\sqrt{f_{\mathcal{S}}(x_k)} - \sqrt{f_{\mathcal{A}}(x_k)} \right)^2 \right) \right)^{\frac{1}{2}}.$$

An advantage of considering Hellinger distances is that
they only take on distance values between 0 and 1, and their
interpretability: The closer the distance value is to 0, the
more the two probability distributions match. This evalua-
tion of the mentioned distance functions was also performed
equivalently for the approximation to the distribution of the
random variables $\left(\log \left(u_t^2 \right) \Big| \tau_t, \alpha, \beta \right)$ in Sect. 3.3. The results
are shown in Table 5.

The evaluated distance functions reveal an appealing
accuracy for both approximations, which is additionally
supported by the successful parameter estimation in the sim-
ulation study in Sect. 4. The two components created for the
approximation of the tails provide additional certainty about
the approximation quality at the tails of the α-stable distri-
butions, since these two components are constructed in such
a way that they correspond to the alpha-stable distribution in
the 5×10^{-7} (resp. 5×10^{-10}) and $1 - 5 \times 10^{-7}$-quantile
(resp. $1 - 5 \times 10^{-10}$-quantile).

Table 6 Parameters of the mixture approximation for $\alpha = 1.73$ and $\beta = -0.14$

	ω_t	1	2	3	4	5	6	7
$\tau_{t=1}$	$\Pr(\tau_t\|\alpha,\beta) = 0.291$	$\mu(\tau_t,\alpha,\beta) = 0.137$	$\sigma^2(\tau_t,\alpha,\beta) = 1.221$					
	$\Pr(\omega_t\|\tau_t,\alpha,\beta)$	0.076	0.082	0.159	0.202	0.273	0.018	0.190
	$\mu(\omega_t,\tau_t,\alpha,\beta)$	-4.432	-3.097	-1.615	-1.337	-0.139	0.931	0.949
	$\sigma^2(\omega_t,\tau_t,\alpha,\beta)$	5.623	6.943	2.845	1.729	0.742	7.959	0.432
$\tau_{t=2}$	$\Pr(\tau_t\|\alpha,\beta) = 0.212$	$\mu(\tau_t,\alpha,\beta) = 0.041$	$\sigma^2(\tau_t,\alpha,\beta) = 1.835$					
	$\Pr(\omega_t\|\tau_t,\alpha,\beta)$	0.106	0.210	0.070	0.081	0.387	0.140	0.006
	$\mu(\omega_t,\tau_t,\alpha,\beta)$	-4.229	-1.899	-1.058	-1.002	0.461	1.500	2.899
	$\sigma^2(\omega_t,\tau_t,\alpha,\beta)$	5.897	2.404	1.095	0.686	0.731	0.383	5.304
$\tau_{t=3}$	$\Pr(\tau_t\|\alpha,\beta) = 0.209$	$\mu(\tau_t,\alpha,\beta) = 0.042$	$\sigma^2(\tau_t,\alpha,\beta) = 1.861$					
	$\Pr(\omega_t\|\tau_t,\alpha,\beta)$	0.085	0.101	0.175	0.181	0.264	0.177	0.018
	$\mu(\omega_t,\tau_t,\alpha,\beta)$	-3.710	-2.949	-0.961	-0.768	0.365	1.403	1.834
	$\sigma^2(\omega_t,\tau_t,\alpha,\beta)$	6.657	4.139	2.037	1.668	0.707	0.423	6.830
$\tau_{t=4}$	$\Pr(\tau_t\|\alpha,\beta) = 0.201$	$\mu(\tau_t,\alpha,\beta) = -0.118$	$\sigma^2(\tau_t,\alpha,\beta) = 2.982$					
	$\Pr(\omega_t\|\tau_t,\alpha,\beta)$	0.078	0.221	0.117	0.125	0.314	0.139	0.006
	$\mu(\omega_t,\tau_t,\alpha,\beta)$	-4.227	-1.536	-0.352	-0.225	1.039	2.004	3.222
	$\sigma^2(\omega_t,\tau_t,\alpha,\beta)$	6.269	2.983	1.417	0.915	0.659	0.376	5.683
$\tau_{t=5}$	$\Pr(\tau_t\|\alpha,\beta) = 0.086$	$\mu(\tau_t,\alpha,\beta) = -0.219$	$\sigma^2(\tau_t,\alpha,\beta) = 14.384$					
	$\Pr(\omega_t\|\tau_t,\alpha,\beta)$	0.090	0.148	0.158	0.168	0.258	0.165	0.013
	$\mu(\omega_t,\tau_t,\alpha,\beta)$	-2.370	-0.318	1.304	1.468	2.511	3.501	5.015
	$\sigma^2(\omega_t,\tau_t,\alpha,\beta)$	5.989	2.612	1.609	1.377	0.663	0.404	5.481
$\tau_{t=6}$	$\Pr(\tau_t\|\alpha,\beta) = 0.0009$	$\mu(\tau_t,\alpha,\beta) = -104.349$	$\sigma^2(\tau_t,\alpha,\beta) = 155018.415$					
	$\Pr(\omega_t\|\tau_t,\alpha,\beta)$	0.042	0.029	0.026	0.216	0.289	0.271	0.126
	$\mu(\omega_t,\tau_t,\alpha,\beta)$	7.254	7.297	7.322	9.304	10.885	12.055	12.973
	$\sigma^2(\omega_t,\tau_t,\alpha,\beta)$	9.069	10.007	10.351	2.755	1.161	0.605	0.354
$\tau_{t=7}$	$\Pr(\tau_t\|\alpha,\beta) = 0.0001$	$\mu(\tau_t,\alpha,\beta) = -5657.136$	$\sigma^2(\tau_t,\alpha,\beta) = 248276285.709$					
	$\Pr(\omega_t\|\tau_t,\alpha,\beta)$	0.091	0.129	0.144	0.134	0.156	0.158	0.188
	$\mu(\omega_t,\tau_t,\alpha,\beta)$	14.298	16.802	17.245	18.943	19.015	19.021	20.120
	$\sigma^2(\omega_t,\tau_t,\alpha,\beta)$	5.652	2.796	1.873	1.294	1.117	1.088	0.485

For illustration, the parameters of the mixture approximation for the α-stable distribution with $\alpha = 1.73$ and $\beta = -0.14$, for which the Hellinger distance is equal to 0.0143, can be found in Table 6.

B Computation of conditional distributions

B.1 Estimation of $\tau_{1:T}$

Since the mixture indices $\tau_{1:T}$ are conditionally independent, we can draw τ_t from

$$p(\tau_t|y_{0:T}, \phi_{1:T}, h_{1:T}, \alpha, \beta)$$
$$= p(\tau_t|y_t, y_{t-1}, \phi_t, h_t, \alpha, \beta)$$
$$\propto p(y_t|y_{t-1}, \phi_t, h_t, \tau_t, \alpha, \beta) \cdot p(y_{t-1}, \phi_t, h_t, \tau_t, \alpha, \beta)$$
$$\propto p(y_t|y_{t-1}, \phi_t, h_t, \tau_t, \alpha, \beta) \cdot \Pr(\tau_t|\alpha, \beta),$$

where the last proportionality follows from the fact that the mixture index τ_t for the approximation of ε_t does not depend on the product $\phi_t y_{t-1}$ and the scale parameter h_t in the observation equation (3). Since y_t conditioned on the chosen mixture component is normally distributed according to equation (8), the likelihood is given by

$$r^\tau_{i,t} := p(y_t|y_{t-1}, \phi_t, h_t, \tau_t = i, \alpha, \beta)$$
$$= \left(2\pi e^{h_t} \sigma^2(\tau_t = i, \alpha, \beta)\right)^{-\frac{1}{2}}$$
$$\exp\left(-\frac{\left(y_t - e^{\frac{h_t}{2}}\mu(\tau_t = i, \alpha, \beta) - \phi_t y_{t-1}\right)^2}{2e^{h_t}\sigma^2(\tau_t = i, \alpha, \beta)}\right),$$

for $i = 1, \ldots, 7$ and $t = 1, \ldots, T$. Therefore we draw τ_t from $\{1, \ldots, 7\}$ according to the posterior probabilities

$$p(\tau_t = i | y_{0:T}, \phi_{1:T}, h_{1:T}, \alpha, \beta)$$
$$= \frac{r_{i,t}^{\tau} \cdot \Pr(\tau_t = i | \alpha, \beta)}{\sum_{j=1}^{7} r_{j,t}^{\tau} \cdot \Pr(\tau_t = j | \alpha, \beta)}, \quad i = 1, \ldots, 7. \quad \text{(A1)}$$

B.2 Estimation of $\omega_{1:T}$

We have to draw ω_t in each iteration of the Gibbs sampler from

$$p(\omega_t | y_{0:T}, \phi_{1:T}, h_{1:T}, \tau_{1:T}, \alpha, \beta)$$
$$= p(\omega_t | y_t, y_{t-1}, \phi_t, h_t, \tau_t, \alpha, \beta)$$
$$\propto p(y_t^* | h_t, \omega_t, \tau_t, \alpha, \beta) \cdot p(h_t, \omega_t, \tau_t, \alpha, \beta)$$
$$\propto p(y_t^* | h_t, \omega_t, \tau_t, \alpha, \beta) \cdot \Pr(\omega_t | \tau_t, \alpha, \beta),$$

where y_t^* is normally distributed, cf. equation (13). In order to update ω_t we calculate the value of the conditional distribution

$$r_{i,t}^{\omega} := p(y_t^* | h_t, \omega_t = i, \tau_t, \alpha, \beta)$$
$$= \left(2\pi \sigma^2(\omega_t = i, \tau_t, \alpha, \beta)\right)^{-\frac{1}{2}}$$
$$\exp\left(-\frac{\left(y_t^* - \mu(\omega_t = i, \tau_t, \alpha, \beta) - h_t\right)^2}{2\sigma^2(\omega_t = i, \tau_t, \alpha, \beta)}\right)$$

for $i = 1, \ldots, 7$ and then we draw $\omega_t \in \{1, \ldots, 7\}$ according to the posterior probabilities

$$p(\omega_t = i | y_{0:T}, \phi_{1:T}, h_{1:T}, \tau_{1:T}, \alpha, \beta)$$
$$= \frac{r_{i,t}^{\omega} \cdot \Pr(\omega_t = i | \tau_t, \alpha, \beta)}{\sum_{j=1}^{7} r_{j,t}^{\omega} \cdot \Pr(\omega_t = j | \tau_t, \alpha, \beta)}. \quad \text{(A2)}$$

B.3 Estimation of σ_{ϕ}^2 and σ_h^2

We assume that the prior distribution for σ_{ϕ}^2 is the inverse gamma distribution $\mathcal{IG}(s_{\phi}, r_{\phi})$ with density $r_{\phi}^{s_{\phi}} x^{-s_{\phi}-1}$ $\exp(-r_{\phi}/x)/\Gamma(s_{\phi})$, $x > 0$, where $s_{\phi} > 0$ is the shape parameter and $r_{\phi} > 0$ the scale parameter. From the state equation (4) we derive

$$p\left(\phi_{1:T} \Big| \sigma_{\phi}^2\right) = \prod_{j=2}^{T} p\left(\phi_j \Big| \phi_{j-1}, \sigma_{\phi}^2\right)$$
$$\propto \sigma_{\phi}^{-(T-1)} \exp\left(-\frac{\sum_{j=2}^{T} (\phi_j - \phi_{j-1})^2}{2\sigma_{\phi}^2}\right)$$

so that

$$p\left(\sigma_{\phi}^2 \Big| \phi_{1:T}\right) \propto \left(\sigma_{\phi}^2\right)^{-\left(s_{\phi} + \frac{T-1}{2}\right)-1}$$
$$\exp\left(-\frac{r_{\phi} + \frac{1}{2} \sum_{j=2}^{T} (\phi_j - \phi_{j-1})^2}{\sigma_{\phi}^2}\right).$$

Hence we can simulate σ_{ϕ}^2 from the posterior distribution

$$\sigma_{\phi}^2 | \phi_{1:T}$$
$$\sim \mathcal{IG}\left(s_{\phi} + \frac{T-1}{2}, \; r_{\phi} + \frac{1}{2} \sum_{j=2}^{T} (\phi_j - \phi_{j-1})^2\right).$$

It turned out that the accuracy of the estimation of σ_{ϕ}^2 can be further improved, when we scale always the new sample of σ_{ϕ}^2 in such a way that the estimated residuals $\hat{\xi}_{1:T,\phi}$ are standardized. An analogous procedure was also chosen for the estimation of σ_h^2.

B.4 Estimation of α and β

We choose flat priors for α and β, i.e. $p(\alpha) = \frac{1}{2}\mathbb{1}_{(0,2]}(\alpha)$ and $p(\beta) = \frac{1}{2}\mathbb{1}_{[-1,1]}(\beta)$. For the estimation of α and β the original observation equation (3) can be used without a mixture approximation of the stable distributions. We draw α and β from the posterior density

$$p(\alpha, \beta | y_{0:T}, \phi_{1:T}, h_{1:T})$$
$$\propto p(y_{0:T} | \phi_{1:T}, h_{1:T}, \alpha, \beta) \cdot p(\alpha) \cdot p(\beta)$$
$$= \prod_{t=1}^{T} p(y_t | y_{t-1}, \phi_t, h_t, \alpha, \beta) \cdot p(\alpha) \cdot p(\beta),$$

where the observation y_t is distributed according to equation (6). As described in Sect. 3.4, we perform one step of the random walk Metropolis-Hastings algorithm in every iteration of the Gibbs sampler, i.e. we draw a proposal of α and β from a two-dimensional normal distribution with covariance matrix Σ to be chosen by the user, and accept the proposal with probability

$$\min\left\{\frac{\prod_{t=1}^{T} f_{\mathcal{S}}\left(y_t|\alpha^{new}, \beta^{new}, e^{\frac{h_t}{2}}, \mu_{\alpha^{new}, \beta^{new}}\right) \cdot \mathbb{1}_{(0,2]}(\alpha^{new}) \cdot \mathbb{1}_{[-1,1]}(\beta^{new})}{\prod_{t=1}^{T} f_{\mathcal{S}}\left(y_t|\alpha^{old}, \beta^{old}, e^{\frac{h_t}{2}}, \mu_{\alpha^{old}, \beta^{old}}\right)}, 1\right\}, \tag{A3}$$

where

$$\mu_{\alpha,\beta} = \begin{cases} \phi_t y_{t-1}, & \alpha \neq 1 \\ \phi_t y_{t-1} - \frac{\beta h_t}{\pi} e^{\frac{h_t}{2}}, & \alpha = 1 \end{cases}.$$

If the new proposed value is rejected, we retain the last sample of α and β. Finally, the drawn samples of α and β must be rounded to the nearest values in the discretized parameter space $\mathfrak{P}$.

C Formulas of the simulation smoother in our analysis

We briefly summarize here the formulas of the simulation smoother by De Jong and Shephard (1995) in the form required for our analysis. We consider the linear Gaussian state space model defined by the equations

$$y_t = c_t + z_t x_t + g_t \varepsilon_t, \qquad \varepsilon_t \sim \mathcal{N}(0, 1),$$
$$x_t = d_t + s_t x_{t-1} + h_t \xi_t, \qquad \xi_t \sim \mathcal{N}(0, 1),$$

where the innovations ε_t and ξ_t are mutually independent for $t = 1, \ldots, T$. The parameter vector $\theta = (c_{1:T}, z_{1:T}, g_{1:T}, d_{1:T}, s_{1:T}, h_{1:T})$ is assumed to be known. The simulation smoother provides a sample from the distribution $p(\varepsilon_{1:T}|y_{1:T}, \theta)$.

Algorithm 2 Simulation smoother
---

    0. Initialisation of $x_{1|0}$ and $p_{1|0} = h_1^2$.
    1. Kalman filtering: For $t = 1, \ldots, T$.

$$v_t = y_t - c_t - z_t x_{t|t-1}, \qquad x_{t+1|t} = d_{t+1} + s_{t+1} x_{t|t-1} + k_t v_t,$$
$$f_t = z_t^2 p_{t|t-1} + g_t^2, \qquad l_t = s_{t+1} - z_t k_t,$$
$$k_t = s_{t+1} z_t p_{t|t-1} f_t^{-1}, \qquad p_{t+1|t} = s_{t+1} p_{t|t-1} l_t + h_{t+1}^2.$$

    2. Saving of $v_{1:T}$, $f_{1:T}$, $k_{1:T}$ and $l_{1:T}$.
    3. Simulation smoothing: For $t = T, \ldots, 1$, with $u_T = 0$ and $r_T = 0$.

$$\psi_t = 1 - g_t^2 f_t^{-1} - g_t^2 k_t^2 u_t, \qquad r_{t-1} = z_t f_t^{-1} v_t + l_t r_t - w_t \psi_t^{-1} \eta_t,$$
$$\eta_t \sim \mathcal{N}(0, \psi_t), \qquad u_{t-1} = z_t^2 f_t^{-1} + l_t^2 u_t + w_t^2 \psi_t^{-1},$$
$$w_t = g_t f_t^{-1} z_t - g_t k_t u_t l_t, \qquad \tilde{\varepsilon}_t = g_t f_t^{-1} v_t - g_t k_t r_t + \eta_t.$$

It can be shown that $\tilde{\varepsilon}_t$ in Algorithm 2 is a sample from $p(\varepsilon_t|\varepsilon_{t+1:T}, y_{1:T}, \theta)$. Since

$$p(\varepsilon_{1:T}|y_{1:T}, \theta) = \prod_{t=1}^{T} p(\varepsilon_t|\varepsilon_{t+1:T}, y_{1:T}, \theta)$$

this implies directly that $\tilde{\varepsilon}_{1:T}$ is a sample from $p(\varepsilon_{1:T}|y_{1:T}, \theta)$.

D TVARαSV processes with $|\phi_t| \geq 1$

In the simulation study of Sect. 4, we have only investigated TVARαSV processes where the process of the autoregressive parameter ϕ_t moves exclusively in the interval $(-1, 1)$. In this section also the TVARαSV processes are analyzed, for which ϕ_t takes values outside the interval $(-1, 1)$ too. These processes were simulated with the same parameters as in the simulation study of Sect. 4. When analyzing these time series, it is noticeable that if the absolute value of ϕ_t is greater than 1 over a certain period of time, the time series very quickly reaches extreme values that are not caused by the α-stable innovations. If this behavior occurs, one can differentiate between the two scenarios that the time series either returns to values of the initial magnitude or continues to increase (or decrease) within the time period under consideration.

Figure 6 shows the first scenario where the underlying Gaussian random walk of ϕ_t causes an extremely strong decrease and subsequent increase of the observable time series $Y_{0:2500}$ in a certain time period. In this time frame, ϕ_t is estimated very accurately by the posterior mean, whereas the posterior mean of h_t differs clearly from the true values. This can be explained by the fact that in this situation in the observation equation (3), due to the achieved order of magnitude of the last observations y_{t-1}, the stochastically scaled innovations are negligible in contrast to the autoregressive part.

The second scenario, in which the observable time series $Y_{0:2500}$ does not return to the initial magnitude, is illustrated in Fig. 7. The last line in this figure also clarifies that this development is not caused by the α-stable innovations $\varepsilon_{1:2500}$. According to our experience for the reason already mentioned above, h_t cannot be estimated at the end of this time series at all and endangers the overall stability of the estima-

Fig. 6 Application of the estimation procedure presented in Sect. 3. The time series $Y_{0:2500}$ was generated according to Definition 1 (first row). In the second and third row the smooth dashed line constitutes the posterior mean of the 6000 used Gibbs sampling estimations and the solid line constitutes the true values of $\phi_{1:2500}$ and $h_{1:2500}$. The 95% confidence bound is plotted gray. In the fourth and fifth row the solid (resp. horizontal dashed) line corresponds to the estimated (resp. true) values of σ_ϕ, σ_h, α and β
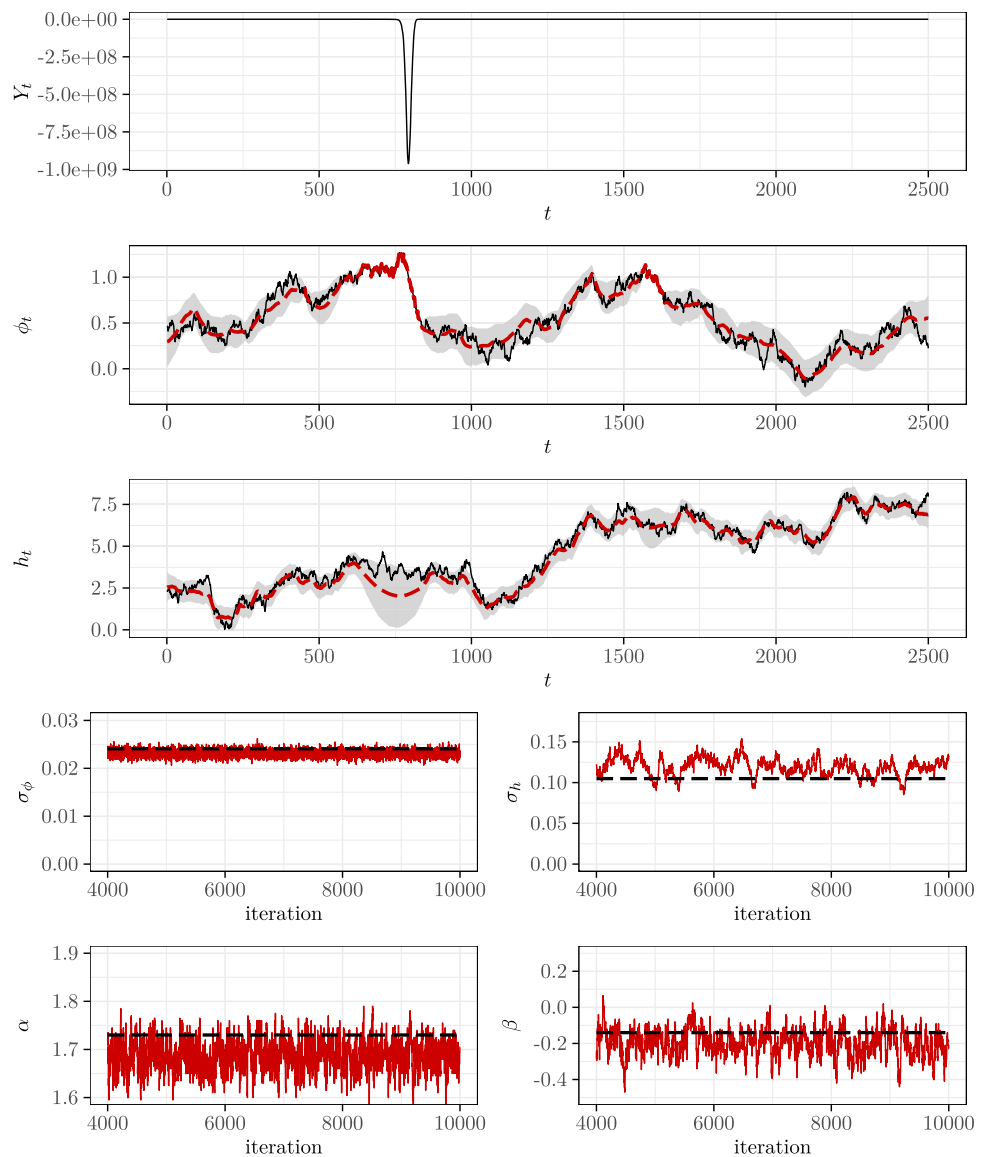
tion algorithm, which is why the estimation procedure was not applied to this TVARαSV realization.

E Simulation study with additional prior knowledge of $\phi_{1:T}$

In the simulation study of Sect. 4, we truncated the prior of $\phi_{1:T}$ by systematically selecting only TVARαSV processes with $|\phi_t| < 1$, for $t = 1, \ldots, T$, which actually requires a reflection in the posterior of $\phi_{1:T}$. In this section we assume that this information about the prior of $\phi_{1:T}$ is available and take the changed posterior into account by resampling $\phi_{1:T}$. This means that Algorithm 1 changes in the way that now in

each iteration of the Gibbs sampler the vector $\tilde{\phi}_{1:T}^{(j)}$ is sampled repeatedly using the same simulation smoother until $|\tilde{\phi}_t^{(j)}| < 1$, for $t = 1, \ldots, T$, is achieved. We applied this adjusted Gibbs sampler to the same 1000 simulated time series from Sect. 4, using the same random starting values and hyperparameters. Furthermore, resampling of $\phi_{1:T}$ was only used after 3500 iterations so that the algorithm can conduct more flexible moves in a certain part of the burn-in phase. In a next step we performed the identical evaluation of the estimation results as in Sect. 4, which is summarized in Table 7 and Table 8. When comparing these tables with Table 1 and Table 2 in Sect. 4, one can see that the estimation results do not differ essentially.

Fig. 7 The time series $Y_{0:2500}$ and the underlying α-stable innovations $\varepsilon_{1:2500}$ were generated according to Definition 1
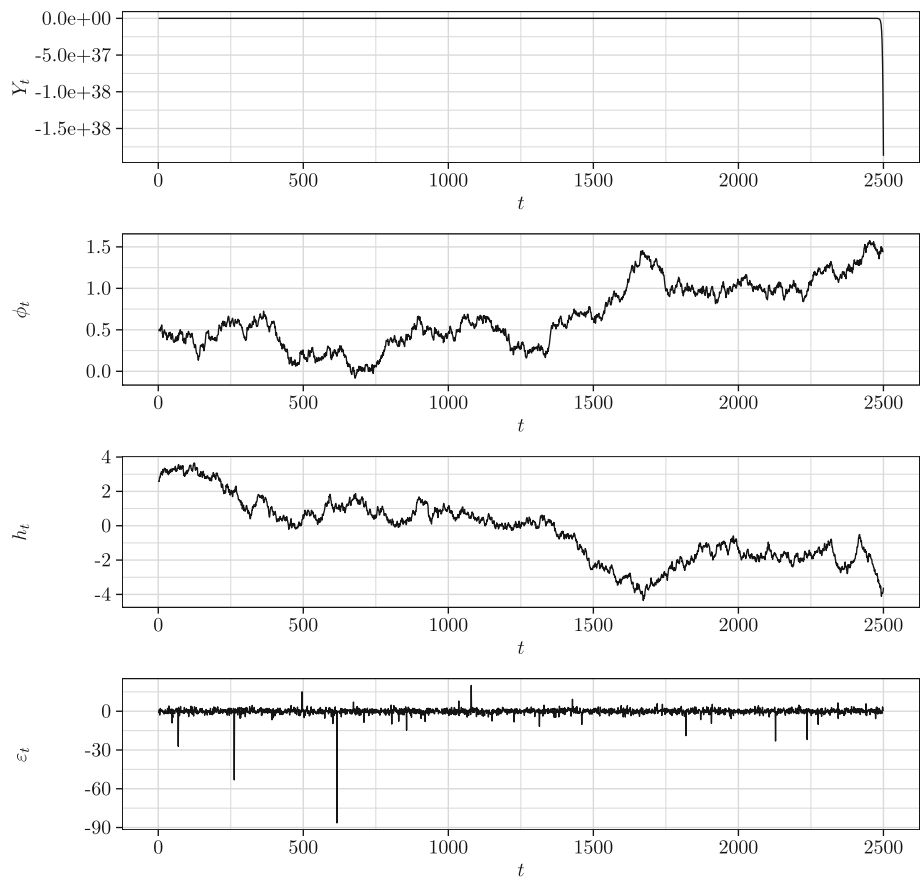
Table 7 Results of the simulation study with resampled $\phi_{1:T}$ using 6000 iterations after a burn-in period of 4000 iterations

	Mean	SD	2.5% quantile	97.5% quantile
Statistics of average biases of 1000 simulations				
$\phi_{1:T}$	-0.0010	0.0165	-0.0315	0.0305
$h_{1:T}$	-0.0604	0.0418	-0.1400	0.0194
Statistics of RMSEs of 1000 simulations				
$\phi_{1:T}$	0.0983	0.0087	0.0823	0.1178
$h_{1:T}$	0.3220	0.0212	0.2813	0.3657

F Estimation and elimination of trend and seasonal components

When we plot the daily EEX Phelix Base electricity price index, we observe a slowly changing local trend component, cf. Fig. 2. Moreover, considering the empirical autocorrelation function, cf. Fig. 8, we see that the data contain a weekly seasonal component. To estimate and eliminate the trend and seasonal component we apply a method based on moving average estimation; see Brockwell and Davis (1991).

We assume that the data x_t are the sum of a trend component m_t, a weekly seasonal component s_t and a TVARαSV

Table 8 Results of the simulation study with resampled $\phi_{1:T}$ using 6000 iterations after a burn-in period of 4000 iterations

	True value	Statistics of posterior means of 1000 simulations				
		Mean	SD	2.5% quantile	97.5% quantile	*Bias*
σ_ϕ^2	0.00058	0.000512	0.000170	0.000223	0.000872	-0.000068
σ_h^2	0.011	0.010617	0.002677	0.005764	0.016146	-0.000383
α	1.73	1.704942	0.031771	1.643511	1.765120	-0.025058
β	-0.14	-0.132107	0.072290	-0.283576	-0.002801	0.007893

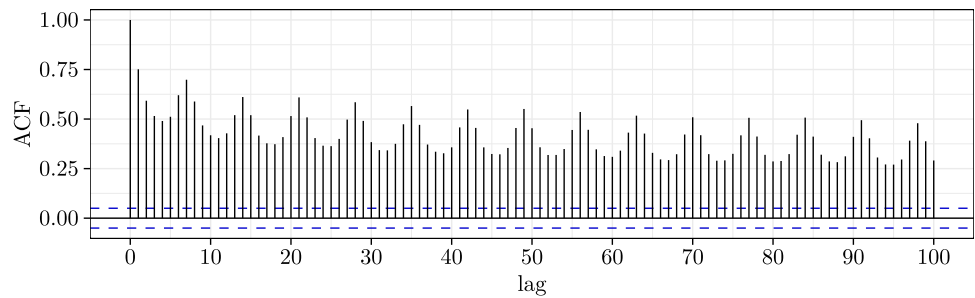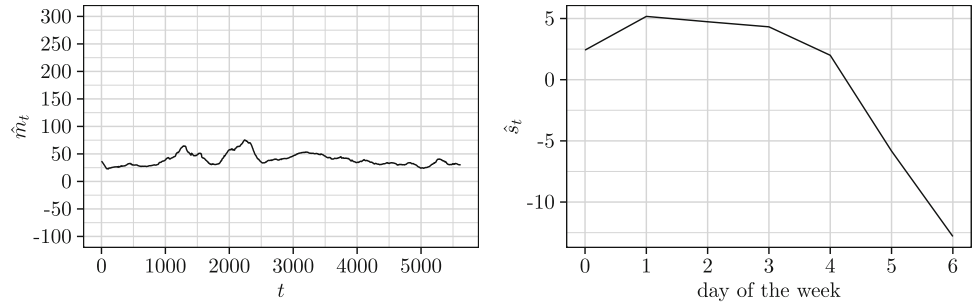**Fig. 8** Empirical autocorrelation function for the Phelix Day Base prices

Fig. 9 Estimation of trend and seasonal components according to the approach described in Sect. 1. Day 0 reflects Monday

process y_t which is scaled by a yearly periodical component ζ_t^*:

$$x_t = m_t + s_t + \zeta_t^* y_t,$$

where $\mathrm{E} y_t = 0$, $s_{t+7} = s_t$, for $t = 0, \ldots, T$, $\sum_{j=0}^{6} s_j = 0$. The exact form and meaning of ζ_t^* is explained later. First, we estimate the trend m_t using a moving average with a window of length 161, that is $\hat{m}_t = \sum_{j=-80}^{80} x_{t+j}/161$, where $x_{-j} = x_0$ and $x_{T+j} = x_T$ for $j = 1, \ldots, 80$. Second, we compute the weekly seasonal component s_t by calculating $w_k = \frac{1}{w} \sum_{j=0}^{w-1} (x_{k+7j} - \hat{m}_{k+7j})$ for $k = 0, \ldots, 6$, where $w = \frac{T}{7}$ is the number of weeks covered by the data, and then setting

$$\hat{s}_t := \begin{cases} w_t - \frac{1}{7} \sum_{j=0}^{6} w_j, & \text{for } t = 0, \ldots, 6 \\ \hat{s}_{t-7}, & \text{for } t > 6 \end{cases}.$$

As in Brockwell and Davis (1991), Section 1.4 (Method S2) we finally reestimate the trend component $\hat{m}_t$ from the deseasonalized data $x_t - \hat{s}_t$. For removing a seasonal pattern in the scaling, we first define

$$\zeta_t := \begin{cases} -\cos\left(\frac{2\pi t}{366}\right), & \text{if } t \text{ lies in a leap year} \\ -\cos\left(\frac{2\pi t}{365}\right), & \text{otherwise} \end{cases}.$$

Due to the extreme spikes the commonly used variance estimate is very sensitive. Hence, we work here with the modulus, i.e. we assume a linear regression of the form

$$|x_t - \hat{m}_t - \hat{s}_t| = \mu + \beta \zeta_t + \kappa_t.$$

The parameter μ reflects the mean level of the scaling, and β is the amplitude of the cosine oscillation modeled by ζ_t. The random variables κ_t reflect the noise in the regression equation. As estimates we obtain $\hat{\mu} = 6.753$ with a standard error of 0.117 and $\hat{\beta} = 1.537$ with a standard error of 0.166 which clearly indicates the presence of a yearly seasonal scaling factor. To keep the magnitude of values in the original data set, we do not divide $x_t - \hat{m}_t - \hat{s}_t$ by $\hat{\mu} + \hat{\beta} \zeta_t$, but by $\zeta_t^* = 1 + (\hat{\beta}/\hat{\mu}) \zeta_t$, i.e.

$$\hat{y}_t = \frac{x_t - \hat{m}_t - \hat{s}_t}{1 + \frac{\hat{\beta}}{\hat{\mu}} \zeta_t}, \quad t = 0, \ldots, T.$$

The results of the described method are shown in Fig. 9.

References

Benth, F.E., Klüppelberg, C., Müller, G., Vos, L.: Futures pricing in electricity markets based on stable CARMA spot models. Energy Econ. **44**, 392–406 (2014)

Benth, F.E., Vos, L.: Cross-commodity spot price modeling with stochastic volatility and leverage for energy markets. Adv. Appl. Probab. **45**(2), 545–571 (2013)

Brockwell, P.J., Davis, R.A.: Time Series: Theory and Methods, 2nd edn. Springer, New York (1991)

Brooks, S., Gelman, A., Jones, G., Meng, X.-L. (eds.): Handbook of Markov Chain Monte Carlo. Chapman and Hall/CRC, New York (2011)

Calzolari, G., Halbleib, R., Parrini, A.: Estimating GARCH-type models with symmetric stable innovations: Indirect inference versus maximum likelihood. Comput. Stat. Data Anal. **76**, 158–171. CFEnetwork: The Annals of Computational and Financial Econometrics (2014)

Casarin, R.: Bayesian inference for generalised Markov switching stochastic volatility models. Cahier du CEREMADE (0414) (2004)

De Jong, P., Shephard, N.: The simulation smoother for time series models. Biometrika **82**(2), 339–350 (1995)

Del Negro, M., Primiceri, G.E.: Time varying structural vector autoregressions and monetary policy: a corrigendum. Rev. Econ. Stud. **82**(4), 1342–1345 (2015)

Engle, R.F.: Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. Econometrica **50**(4), 987–1007 (1982)

Erdogdu, E.: Asymmetric volatility in European day-ahead power markets: a comparative microeconomic analysis. Energy Econ. **56**, 398–409 (2016)

García, I., Klüppelberg, C., Müller, G.: Estimation of stable CARMA models with an application to electricity spot prices. Stat. Model. **11**(5), 447–470 (2011)

Ghysels, E., Harvey, A.C., Renault, E.: Stochastic volatility. In: Maddala, G., Rao, C. (Eds.), Statistical Methods in Finance, Volume 14 of Handbook of Statistics, pp. 119–191. Elsevier (1996)

Hull, J., White, A.: The pricing of options on assets with stochastic volatilities. J. Finance **42**(2), 281–300 (1987)

Jacquier, E., Polson, N.G., Rossi, P.E.: Bayesian analysis of stochastic volatility models. J. Bus. Econ. Stat. **12**(4), 371–389 (1994)

Jarque, C.M., Bera, A.K.: Efficient tests for normality, homoscedasticity and serial independence of regression residuals. Econ. Lett. **6**(3), 255–259 (1980)

Kim, S., Shephard, N., Chib, S.: Stochastic volatility: likelihood inference and comparison with ARCH models. Rev. Econ. Stud. **65**(3), 361–393 (1998)

Klüppelberg, C., Meyer-Brandis, T., Schmidt, A.: Electricity spot price modelling with a view towards extreme spike risk. Quant. Finance **10**(9), 963–974 (2010)

Kristoufek, L.: Leverage effect in energy futures. Energy Econ. **45**, 1–9 (2014)

Ljung, G.M., Box, G.E.P.: On a measure of lack of fit in time series models. Biometrika **65**(2), 297–303 (1978)

Lombardi, M.J., Calzolari, G.: Indirect estimation of α-stable stochastic volatility models. Comput. Stat. Data Anal. *53*(6), 2298–2308. (2009) The Fourth Special Issue on Computational Econometrics

Martin, G.M., McCabe, B.P.M., Frazier, D.T., Maneesoonthorn, W., Robert, C.P.: Auxiliary likelihood-based approximate Bayesian computation in state space models. J. Comput. Gr. Stat. **28**(3), 508–522 (2019)

MATLAB (2020). version 9.8 (R2020a). Natick, Massachusetts: The MathWorks Inc

Müller, G., Seibert, A.: Bayesian estimation of stable CARMA spot models for electricity prices. Energy Econ. **78**, 267–277 (2019)

Peters, G.W., Sisson, S.A., Fan, Y.: Likelihood-free Bayesian inference for α-stable models. Computational Statistics & Data Analysis *56*(11), 3743–3756. (2012) 1st issue of the Annals of Computational and Financial Econometrics Sixth Special Issue on Computational Econometrics

Pivato, M., Seco, L.: Estimating the spectral measure of a multivariate stable distribution via spherical harmonic analysis. J. Multivar. Anal. **87**(2), 219–240 (2003)

Primiceri, G.E.: Time varying structural vector autoregressions and monetary policy. Rev. Econ. Stud. **72**(3), 821–852 (2005)

R Core Team: R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria (2020)

Samorodnitsky, G., Taqqu, M.S.: Stable Non-Gaussian Random Processes: Stochastic Models with Infinite Variance. Chapman and Hall/CRC, New York (1994)

Shephard, N.: Partial non-Gaussian state space. Biometrika **81**(1), 115–131 (1994)

Shephard, N.: Statistical aspects of ARCH and stochastic volatility. In: Cox, D., Hinkley, D.V., Barndorff-Nielsen, O.E. (eds.) Time Series Models in Econometrics, Finance and Other Fields, pp. 1–67. Chapman & Hall, London (1996)

Taylor, S.J.: Modeling stochastic volatility: a review and comparative study. Math. Finance **4**(2), 183–204 (1994)

Titterington, D.M., Smith, A.F.M., Makov, U.E.: Statistical Analysis of Finite Mixture Distributions, 1st edn. Wiley, Chichester (1985)

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

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.