Universität St.Gallen

Regime Switching Stochastic Volatility with Skew, Fat Tails and Leverage using Returns and Realized Volatility Contemporaneously

Sebastian Trojan

August 2014 Discussion Paper no. 2013-41

School of Economics and Political Science,
Department of Economics

University of St. Gallen

Editor:	Martina Flockerzi University of St.Gallen School of Economics and Political Science Department of Economics Bodanstrasse 8 CH-9000 St. Gallen Phone +41 71 224 23 25 Fax +41 71 224 31 35 Email seps@unisg.ch
Publisher:	School of Economics and Political Science Department of Economics University of St.Gallen Bodanstrasse 8 CH-9000 St. Gallen Phone +41 71 224 23 25
Electronic Publication:	Fax +41 71 224 31 35 http://www.seps.unisg.ch

Regime Switching Stochastic Volatility with Skew, Fat Tails and Leverage using Returns and Realized Volatility Contemporaneously[1]

Sebastian Trojan

Author's address: Sebastian Trojan
Institute of Mathematics and Statistics
Bodanstrasse 6
CH-9000 St. Gallen
Email sebastian.trojan@student.unisg.ch
sebastian.trojan@web.de

[1] I thank Prof. Dr. Francesco Audrino, Prof. Dr. Michael Lechner and seminar participants at the University of St. Gallen for many useful comments and suggestions.

Abstract

A very general stochastic volatility (SV) model specification with leverage, heavy tails, skew and switching regimes is proposed, using realized volatility (RV) as an auxiliary time series to improve inference on latent volatility. The information content of the range and of implied volatility using the VIX index is also analyzed. Database is the S&P 500 index. Asymmetry in the observation error is modeled by the generalized hyperbolic skew Student-*t* distribution, whose heavy and light tail enable substantial skewness. Resulting number of regimes and dynamics differ dependent on the auxiliary volatility proxy and are investigated in-sample for the financial crash period 2008/09 in more detail. An out-of-sample study comparing predictive ability of various model variants for a calm and a volatile period yields insights about the gains on forecasting performance from different volatility proxies. Results indicate that including RV or the VIX pays off mostly in more volatile market conditions, whereas in calmer environments SV specifications using no auxiliary series outperform. The range as volatility proxy provides a superior in-sample fit, but its predictive performance is found to be weak.

Keywords

Stochastic volatility, realized volatility, non-Gaussian and nonlinear state space model, Generalized Hyperbolic skew Student-*t* distribution, mixing distribution, regime switching, Markov chain Monte Carlo, particle filter.

JEL Classification

C11, C15, C32, C58.

1 Introduction

Since Jacquier, Polson, and Rossi (1994), and Kim, Shephard, and Chib (1998) (KSC) have proposed Bayesian Markov chain Monte Carlo (McMC) methods for estimating the basic stochastic volatility (SV) model with one factor and Gaussian errors, research in this area has developed considerably. Harvey, Ruiz, and Shephard (1994) propose a Student-t error based SV model. Subsequently, Chib, Nardari, and Shephard (2002) estimate this type of model by McMC, using a scale mixture of normals representation to describe excess kurtosis in S&P 500 index data. Abanto-Valle, Bandyopadhyay, Lachos, and Enriquez (2010) examine the slash and variance-gamma distribution. A heavy tailed error results in a more robust underlying stochastic volatility process, as the latter will not increase before a series of large absolute returns is observed. Jacquier, Polson, and Rossi (2004) provide an illustration. They also incorporate the leverage effect into their proposed SV model, albeit defining leverage in a less widespread, contemporaneous way. More commonly, leverage is defined as an increase in future volatility due to a price drop today (see e.g. Omori, Chib, Shephard, and Nakajima, 2007, or Yu, 2005). Leverage is an important stylized fact of especially stock return indices.

Jumps are another tool to model outliers in the data. Chib et al. (2002) include jumps in the observation equation but find that for the specific dataset at hand, a model with Student-t error outperforms the model with jumps. Moreover, including both jumps and Student-t error does not yield a significant improvement, probably due to overparameterization. Essentially, fat tails and jumps both model outliers. Once accounted for fat tails of the distribution, there may be too few observations left to estimate the jump parameter reliably. Nakajima and Omori (2009) compare various SV models based on heavy tails and/or jumps, further estimating a candidate with correlated jumps in the return and volatility equation. They arrive at the same conclusion.

Alternatively, tail behavior may be modeled by a second volatility factor. Chernov, Gallant, Ghysels, and Tauchen (2003) test various continuous time model specifications with one and two volatility factors and normal errors in an efficient method of moments (EMM) framework. They find the specification with two factors superior. Molina, Han, and Fouque (2010) test one and two factor models with Gaussian errors on various exchange rate series using McMC and conclude that the need for a second factor depends on the dataset at hand. However, as Bouchaud and Potters (2009) note, variance may not be a good description for the tail behavior of a distribution. Results of Durham (2006) support this argument, as he finds the improvement of a second factor to be relatively small and not sufficient to accurately model shape of the conditional return distribution (using simulated maximum likelihood (SML) for estimation). Asai (2008) compares two factor SV models using Gaussian errors with one factor models using the Student-t distribution. His results show empirical superiority of the Student-t specification on various datasets. Liesenfeld and Richard (2003) provide comparable work for the S&P 500, the IBM stock, and the $/DM exchange rate using efficient importance sampling (EIS) for estimation and arrive at similar conclusions. The latter authors also estimate a model with a semi-nonparametric distribution. To this end, Jensen and Maheu (2010) propose to flexibly estimate skewness and kurtosis of the return distribution with nonparametric Bayesian methods.

Yet another flexible approach to model jumps in volatility is by deploying regime switching. A strength of this method is that sudden shifts in volatility can be modeled either as persistent or temporary. For example, during the financial crisis 2008/09, volatility increased significantly for a prolonged period of time. Applications of Markov regime switching to the SV model class can be found in e.g. So, Lam, and Li (1998), Shibata and Watanabe (2005), or Carvalho and Lopes (2007). The latter authors apply sequential Monte Carlo (SMC) for estimation.

In addition to being leptokurtic, return distributions likely exhibit skewness as well. Cappuccio, Lubian, and Raggi (2004) model SV of daily/weekly exchange rates using a skewed generalized error distribution (Skew-GED). They further estimate nested alternatives (neither skew, fat tails, or both), and conclude in favor of the full specification. Abanto-Valle, Lachos, and Dey (2013) use the skew-Student-t (ST) distribution to model SV of NASDAQ daily index returns. In a model selection study similar in structure to that of the former authors they find the skewed heavy tailed specification superior. Nakajima and Omori (2012) model the return innovation as a generalized hyperbolic skew Student-t distribution ($\mathcal{GH}$ skew-t) and show that this specification outperforms both the basic SV model and the model with symmetric Student-t error for different periods of the S&P 500 and TOPIX daily return index. They include leverage in all their models.

An auxiliary time series potentially improves inference on the latent SV process. In a Bayesian framework, Mahieu and Bauer (1998), Watanabe (2000), Abanto-Valle, Migon, and Lopes (2010), and Abanto-Valle, Dey, and Lachos (2012) incorporate trading volume into the SV model as an additional measurement equation. Their primary interest lies in answering the question of how well an underlying latent process (i.e. the information flow) explains both return and volume. Such a complete specification of the joint dynamic distribution of return and related market variables allows for a more structural approach to stochastic volatility modeling and is generally referred to as the mixture of distributions hypothesis (MDH) (Clark, 1972), and building on the latter, the modified mixture hypothesis (MMM) (Andersen, 1996).

Combining SV models with realized volatility (RV) estimators containing intraday information, or with implied volatility containing future expectations, is another possibility to improve on inference and forecasting of latent volatility. Koopman, Jungbacker, and Hol (2005) include realized and implied volatility measures as explanatory variables in the SV transition equation and compare predictive ability with competing GARCH and RV models for the S&P 100. Alizadeh, Brandt, and Diebold (2002), and Brandt and Jones (2005) use the range to extract latent stochastic volatility of daily exchange rates. Takahashi, Omori, and Watanabe (2009) model a bivariate system of RV and returns to estimate stochastic volatility of the TOPIX, testing several RV estimators in-sample, and proposing a Bayesian McMC estimation algorithm. Dobrev and Szerszen (2010) present a related approach. Takahashi, Watanabe and Omori (2014) extend Takahashi et al. (2009) using the $\mathcal{GH}$ skew-t distribution, and conducting volatility and quantile forecasts. Koopman and Scharth (2012) present a two step maximum likelihood estimation method for these kind of models and a detailed empirical study. Jacquier and Miller (2012) (henceforth JM) compare different SV specifications including RV and implied volatility w.r.t. their predictive ability on index returns and foreign exchange rates during the 2008-09 financial crisis, applying sequential Monte Carlo with time varying parameters.

Encouraged by positive results in the literature regarding inference and prediction of latent volatility when including RV into SV models (e.g. JM), this approach is extended in two directions. First, the asymmetric $\mathcal{GH}$ skew-t distribution, which is able to model substantially skewed data, is used for returns. Second, regime switching is introduced. Moreover, the range and implied volatility as auxiliary time series are also investigated. In fact, four distinct regimes can be identified using realized volatility and return contemporaneously. Regime partitions vary according to the auxiliary measure used and are investigated in-sample during the financial crisis 2008/09 in more detail. Extended in- and out-of-sample model selection and forecasting studies provide further insights.

The paper proceeds as follows. Section 2 presents the model and Section 3 details a McMC algorithm to estimate it. Section 4 contains an associated particle filter. Section 5 conducts simulation studies. Empirical applications on S&P 500 index data follow in Section 6, including parameter and regime inference for model variants, model selection, prior sensitivity analysis and out-of-sample predictive density tests. Section 7 concludes. The Appendix contains technical details, model estimates and additional specification tests.

2 General Model Specification

A basic stochastic volatility (SV) model with normal observation error and leverage can be written as

$$y_t = \exp(h_t/2)\epsilon_t, \qquad t = 1, \ldots, n, \tag{1}$$

$$h_{t+1} = \mu + \phi(h_t - \mu) + \eta_{t+1}, \qquad t = 1, \ldots, n-1, \tag{2}$$

$$h_1 \backsim \mathcal{N}(\mu, \sigma^2_{\text{ini}}), \tag{3}$$

$$\begin{pmatrix} \epsilon_t \\ \eta_{t+1} \end{pmatrix} \backsim \mathcal{N}(0, \Sigma), \quad \Sigma = \begin{pmatrix} 1 & \rho\sigma \\ \rho\sigma & \sigma^2 \end{pmatrix}, \tag{4}$$

where y_t is an asset return and h_t the unobserved volatility.[1] In state space terminology, Eq. (1) is called measurement or observation equation and Eq. (2) transition equation, the latter being unobservable (see e.g Durbin and Koopman, 2008, or Harvey, 1989, for a comprehensive treatment). The volatility process is commonly assumed to follow a stationary $\mathrm{AR}(1)$ process by imposing $|\phi| < 1$. The initial value h_1 in Eq. (3) then follows the unconditional distribution with $\sigma^2_{\text{ini}} = \sigma^2/(1 - \phi^2)$. Parameter ρ, Eq. (4), measures the correlation between ϵ_t and η_{t+1}. We have volatility asymmetry if $\rho \neq 0$ and specifically, if $\rho < 0$, we speak of the leverage effect.

As briefly mentioned in the introduction, there is a timing issue of the leverage effect. Leverage may be modeled in accordance with the traditional definition of leverage as in Black (1976) and Christie (1982). A negative return today will increase volatility tomorrow. Alternatively, the leverage effect may be modeled as an instantaneous relationship between return and volatility, i.e. a negative return today is associated with a concurrent shift in volatility. This introduces an additional source of non-Gaussianity into the model, as the conditional return distribution gets skewed. Jacquier et al. (2004) argue that this may help to model occasionally occurring extreme negative returns (market crashes). Empirically, Durham (2006)

[1]Strictly speaking, it is the unobserved log-variance, but as in the literature I speak about volatility throughout (similarly for RV related measures). This simplifies terminology and should not cause conceptual confusion as the respective measures are linked by monotonic transformation.

concludes positively for the contemporaneous specification, whereas Yu (2005) for the traditional. However, Yu (2005) shows that a contemporaneous correlation between return and volatility implies a nonzero expected return and makes an interpretation of the correlation parameter difficult. Consequently, leverage is implemented using the traditional specification, which is also favored in the literature.[2] Moreover, as asymmetry is modeled explicitly, the additional effect of conditional skewness via a contemporaneous specification is unnecessary.

Asymmetry in the return distribution is incorporated into the model by replacing measurement error ϵ_t in Eq. (4) with a generalized hyperbolic ($\mathcal{GH}$) random variate w_t following Nakajima and Omori (2012). Specifically, as parameters of this distribution are hard to estimate due to a flat likelihood (e.g. Aas and Haff, 2006), a limiting case, the generalized hyperbolic skew Student-t ($\mathcal{GH}$ skew-t) distribution, is deployed. It is parsimoniously parameterized, more amenable to estimation, and has a normal mean-variance mixture representation,

$$w_t = \mu_w + \gamma z_t + \sqrt{z_t}\epsilon_t, \quad \epsilon_t \backsim \mathcal{N}(0,1), \quad z_t \backsim \mathcal{IG}(\nu/2, \nu/2), \tag{5}$$

with μ_w and γ the location and skew parameter, respectively. Scaling variable z_t is inverse gamma distributed. The structure of Eq. (5) leans itself well to the construction of a McMC algorithm in the context of Bayesian inference. A distinct feature of the $\mathcal{GH}$ skew-t distribution is its heavy and light tail that enable the modeling of substantial skewness. See App. B for a derivation from the $\mathcal{GH}$ distribution and illustrations. To achieve $\mathrm{E}(w_t) = 0$, a common assumption of high frequency returns, $\mu_w = -\gamma\mu_z$ and $\mu_z \equiv \mathrm{E}(z_t) = \nu/(\nu-2)$. The additional constraint $\nu > 4$ ensures existence of the 2^{nd} moment, $\mathrm{E}(w_t^2) < \infty$.

Realized volatility (RV), which has been independently proposed by Andersen and Bollerslev (1998), and Barndorff-Nielsen and Shephard (2001a), provides a consistent and unbiased estimator of the integrated volatility (IV) of a price process under ideal market assumptions, i.e. continuous and frictionless trading. With n intraday returns during day t, $\{r_{t,i}\}_{i=1}^{n}$,

$$RV_t = \sum_{i=1}^{n} r_{t,i}^2, \quad IV_t = \int_t^{t+1} \sigma^2(s)ds, \quad RV_t \to IV_t, \quad n \to \infty. \tag{6}$$

In real markets, however, RV tends to be a biased estimator of IV due to microstructure effects and non-trading. RV provides empirical content to latent volatility state variable h_t (see e.g. Andersen, 2008, for a treatment in continuous time). Consequently, RV is taken as an observable proxy for latent volatility, on which the focus of modeling and prediction lies in this work. Gains in inference on h_t can be expected, as RV contains valuable intraday information. To give a simple illustration, we may have a near zero return on a specific day, but the market went up and down a lot. Then, using return alone, we would falsely infer low volatility.

Various measures for realized volatility have been proposed in the literature. However, a detailed discussion is out of the scope of the current work. To give examples, Hansen and Lunde (2006) scale RV to account for non-trading hours. Zhang, Mykland, and Aït-Sahalia (2005) propose a two scale RV estimator (TSRV), correcting bias due to microstructure noise by combining two RV estimators from different frequencies. Barndorff-Nielsen, Hansen, Lunde, and Shephard (2008) design a realized kernel estimator that explicitly accounts for market frictions while taking

[2] Modeling leverage in the traditional definition has an additional feature: The resulting discrete time model is an Euler approximation to the continuous time SV model with leverage effect.

advantage of the highest frequency available. For reviews on RV, consider e.g. Andersen, Bollerslev and Diebold (2010), or McAleer and Medeiros (2008). See further JM, Takahashi et al. (2009), and the references therein for a more detailed discussion of SV in combination with RV estimators and microstructure effects. As in JM, information content of implied volatility and the range as additional volatility proxies is also investigated.

A complete Markov switching stochastic volatility model with $\mathcal{GH}$ skew-t error and RV as auxiliary measurement variable (MSSVsk$t$-RV in the following) is now formulated as

$$\begin{aligned} y_{1,t} &= \{\gamma_{s_t}(z_t - \mu_z) + \sqrt{z_t}\epsilon_t\} \times \exp(h_t/2), && && (7) \\ y_{2,t} &= \zeta + \xi h_t + u_t, && t = 1, \dots, n, && (8) \\ h_{t+1} &= \mu_{s_{t+1}} + \phi_{s_{t+1}}(h_t - \mu_{s_t}) + \eta_{t+1}, && t = 1, \dots, n-1, && (9) \\ h_1 &\backsim \mathcal{N}(\mu_{s_1}, \sigma^2_{\mathrm{ini}}), \end{aligned}$$

where

$$\begin{aligned} \boldsymbol{\theta}_{s_t} &= \sum_{i=1}^{M} \boldsymbol{\theta}_i \times \mathrm{I}_{\mathrm{s_t=i}}, \quad M \in \mathbb{N}^+, \\ \boldsymbol{\theta}_i &= (\mu_i, \phi_i, \sigma_i, \rho_i, \gamma_i, \sigma_{u,i})', \\ \begin{pmatrix} \epsilon_t \\ \eta_{t+1} \end{pmatrix} &\backsim \mathcal{N}(0, \boldsymbol{\Sigma}_{s_{t+1}}), \quad u_t \backsim \mathcal{N}(0, \sigma^2_{u,s_t}), \quad z_t \backsim \mathcal{IG}\left(\nu/2, \nu/2\right), \\ \boldsymbol{\Sigma}_{s_{t+1}} &= \begin{pmatrix} 1 & \rho_{s_{t+1}}\sigma_{s_{t+1}} \\ \rho_{s_{t+1}}\sigma_{s_{t+1}} & \sigma^2_{s_{t+1}} \end{pmatrix}, \end{aligned} \tag{10}$$

with return $y_{1,t}$, auxiliary time series $y_{2,t}$, and M the number of regimes. As in the basic SV model, Eq. (1)-(4), stationarity of h_t is imposed, $|\phi_{s_t}| < 1$, and accordingly, $\sigma^2_{\mathrm{ini}} = \sigma^2_{s_1}/(1 - \phi^2_{s_1})$. $\mathrm{I}_{\mathrm{s_t=i}}$ is an indicator variable that is one if $s_t = i$ (state i prevailing at time t) and zero otherwise. Vector $\boldsymbol{\theta}_i$, $i = 1, \dots, M$, contains the regime dependent parameters of the i^{th} regime. Note how volatility enters as a linear regression variable into second measurement equation (8). If bias correction term ζ is negative, we may conclude that the effect of non-trading hours dominates market microstructure noise, and vice versa for a positive ζ (Takahashi et al., 2009). Parameter ξ is a scaling parameter. JM show in a simulation exercise that incorporating RV into the SV model via a second measurement equation reduces uncertainty regarding inference of latent volatility in a mean square error sense, both in-sample (smoothing performance) and out-of-sample (predictive ability). The authors further demonstrate that including RV as an exogenous variable directly in Eq. (9) has to be considered inferior according to the above criterion.

Model specification is completed by assuming latent state variable s_t to follow a first-order Markov process (see e.g. Hamilton, 1989). Transition probabilities are time invariant, $\Pr(s_t = j | s_{t-1} = i) = p_{ij}$, $i, j \in \{1, \dots, M\}$, with transition probability matrix having the form

$$\boldsymbol{P} = \begin{pmatrix} p_{11} & \cdots & p_{1M} \\ \vdots & \ddots & \vdots \\ p_{M1} & \cdots & p_{MM} \end{pmatrix}. \tag{11}$$

Generally speaking, in the following empirical applications, as much regimes as possible are extracted out of the employed dataset, without imposing rather informative priors on the identifying variables of the states.

Almost all parameters in the model are made regime dependent, as ex-ante there is little reason for specific parameters to be assumed equal across regimes. Moreover, extensive simulation has shown that stability of equilibria and associated convergence behavior of the McMC chain often depend on the regime switching variable set. Consequently, a very general model specification is proposed. Specifically, volatility mode μ, volatility persistence ϕ, volatility of volatility σ, volatility asymmetry ρ, skew parameter γ, and auxiliary equation measurement noise σ_u are made regime dependent. Degrees of freedom parameter ν is kept constant across regimes, as this parameter is hard to estimate precisely even when the sample size is large. Computational gains can be made with ν regime independent, as the inverse gamma distribution has not to be evaluated in the forward filtering backward sampling (FFBS) algorithm for the latent states, see Sec. 3 Step 3. Note that keeping ν fix imposes equality on the heavier tail decay rate of the regime dependent distributions. However, skew parameter γ may shift large part of the probability mass into the appropriate tail (see App. B for an illustration). Regarding a regime switching RV bias parameter ζ, pilot runs extracted a regime with very low volatility, offset by a large positive value for ζ in the auxiliary equation. This counteracts our motivation to include intraday information, also pointing in the direction of overfitting, and consequently auxiliary regression parameters ζ and ξ are set regime independent as well. To summarize, the proposed regime switching specification is quite exhaustive and beyond that of most models in the literature, which commonly assume a regime dependent volatility mode only. We may miss important insights if we restrict our analysis solely on a regime switching volatility mode.

The proposed model includes various models of the literature as subcomponents. Dropping regime switching and skew ($\gamma = 0$) results in the model of JM. Additionally setting scaling parameters $z_t = 1$, $t = 1, \ldots, n$, and $\xi = 1$ recovers the model proposed by Takahashi et al. (2009). Without regime switching and RV as auxiliary measurement variable, the model reduces to that of Nakajima and Omori (2012). Assuming further a Gaussian measurement error we obtain the model of Yu (2005), or Omori, Chib, and Shephard (2007).

3 McMC Algorithm

Let $\boldsymbol{\theta} \equiv \{\boldsymbol{\mu}, \boldsymbol{\phi}, \boldsymbol{\sigma}, \boldsymbol{\rho}, \boldsymbol{\gamma}, \nu, \zeta, \xi, \boldsymbol{\sigma}_u, \boldsymbol{P}\}$ and $\boldsymbol{y}_t = (y_{1,t}, y_{2,t})'$, $\boldsymbol{y}_{1:n} = (\boldsymbol{y}_1, \ldots, \boldsymbol{y}_n)'$, $\boldsymbol{h}_{1:n} = (h_1, \ldots, h_n)'$, $\boldsymbol{z}_{1:n} = (z_1, \ldots, z_n)'$, $\boldsymbol{s}_{0:n} = (s_0, s_1, \ldots, s_n)'$. All parameters in $\boldsymbol{\theta}$ except ν and $\boldsymbol{\vartheta} \equiv (\zeta, \xi)'$ are regime dependent, e.g. $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_M)'$, $M \in \mathbb{N}^+$. Denote by $\boldsymbol{\theta}_{-r}$ all parameters in $\boldsymbol{\theta}$ without r. Further, introduce the reparameterization $\{\boldsymbol{\sigma}, \boldsymbol{\rho}\} \mapsto \{\boldsymbol{\lambda}, \boldsymbol{a}\}$ leading to a Cholesky-type decomposition of $\boldsymbol{\Sigma}$, Eq. (10), to sample $\{\boldsymbol{\sigma}, \boldsymbol{\rho}\}$ in analytical form, see Step 11 below. Prior distributions are assumed to be independent and generically denoted by $\pi(\cdot)$, with $\pi(\boldsymbol{\mu})$, $\pi(\boldsymbol{\gamma})$, $\pi(\boldsymbol{a})$, $\pi(\boldsymbol{\vartheta})$ conjugate multivariate normal,

$$\begin{aligned}\boldsymbol{\mu} &\backsim \mathcal{N}(\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_{\mu_0}), \quad \boldsymbol{\gamma} \backsim \mathcal{N}(\boldsymbol{\gamma}_0, \boldsymbol{\Sigma}_{\gamma_0}),\\ \boldsymbol{a} &\backsim \mathcal{N}(\boldsymbol{a}_0, \boldsymbol{\Sigma}_{a_0}), \quad \boldsymbol{\vartheta} \backsim \mathcal{N}(\boldsymbol{\vartheta}_0, \boldsymbol{\Sigma}_{\vartheta_0}),\end{aligned}$$

where $\boldsymbol{\Sigma}_{\mu_0}$, $\boldsymbol{\Sigma}_{\gamma_0}$, $\boldsymbol{\Sigma}_{a_0}$, $\boldsymbol{\Sigma}_{\vartheta_0}$ are diagonal. Further, priors for $\boldsymbol{\lambda}$, $\boldsymbol{\sigma}_u$ are conjugate inverse gamma,

$$\lambda_i \backsim \mathcal{IG}\left(a_{0,i}/2, b_{0,i}/2\right), \quad \sigma^2_{u,i} \backsim \mathcal{IG}\left(c_{0,i}/2, d_{0,i}/2\right), \tag{12}$$

and those for $\boldsymbol{P}$ conjugate Dirichlet,

$$\boldsymbol{P}_{i\cdot} \backsim \mathcal{D}(\varrho_{i1}, \ldots, \varrho_{iM}), \quad \varrho_{ij} \neq 0, \quad i, j = 1, \ldots, M,$$

where $\boldsymbol{P}_{i\cdot}$ denotes the i^{th} (independent) row of $\boldsymbol{P}$. Random samples from the full posterior distribution of $\{\boldsymbol{\theta}, \boldsymbol{h}_{1:n}, \boldsymbol{z}_{1:n}, \boldsymbol{s}_{0:n}\}$ given $\boldsymbol{y}_{1:n}$ are then drawn using the McMC method (see e.g. Robert and Casella, 2004).

The algorithm for the MSSVsk$t$-RV model is now structured as follows:[3]

1. Initialize $\boldsymbol{\theta}$, $\boldsymbol{h}_{1:n}$, $\boldsymbol{z}_{1:n}$, $\boldsymbol{s}_{0:n}$.
2. Generate $\boldsymbol{h}_{1:n} \,|\, \boldsymbol{\theta}, \boldsymbol{z}_{1:n}, \boldsymbol{s}_{0:n}, \boldsymbol{y}_{1:n}$.
3. Generate $\boldsymbol{s}_{0:n} \,|\, \boldsymbol{\theta}, \boldsymbol{h}_{1:n}, \boldsymbol{z}_{1:n}, \boldsymbol{y}_{1:n}$.
4. Generate $\boldsymbol{P} \,|\, \boldsymbol{s}_{0:n}$.
5. Generate $\boldsymbol{z}_{1:n} \,|\, \boldsymbol{\theta}, \boldsymbol{h}_{1:n}, \boldsymbol{s}_{0:n}, \boldsymbol{y}_{1,1:n}$.
6. Generate $\nu \,|\, \boldsymbol{\theta}_{-\nu}, \boldsymbol{h}_{1:n}, \boldsymbol{z}_{1:n}, \boldsymbol{s}_{0:n}, \boldsymbol{y}_{1,1:n}$.
7. Generate $\gamma \,|\, \boldsymbol{\theta}_{-\gamma}, \boldsymbol{h}_{1:n}, \boldsymbol{z}_{1:n}, \boldsymbol{s}_{0:n}, \boldsymbol{y}_{1,1:n}$.
8. Generate $(\zeta, \xi)' \,|\, \boldsymbol{\theta}_{-(\zeta,\xi)'}, \boldsymbol{h}_{1:n}, \boldsymbol{s}_{0:n}, \boldsymbol{y}_{2,1:n}$.
9. Generate $\boldsymbol{\sigma}_u \,|\, \boldsymbol{\theta}_{-\sigma_u}, \boldsymbol{h}_{1:n}, \boldsymbol{s}_{0:n}, \boldsymbol{y}_{2,1:n}$.
10. Generate $\boldsymbol{\mu} \,|\, \boldsymbol{\theta}_{-\mu}, \boldsymbol{h}_{1:n}, \boldsymbol{z}_{1:n}, \boldsymbol{s}_{0:n}, \boldsymbol{y}_{1,1:n}$.
11. Generate $\{\boldsymbol{\sigma}, \boldsymbol{\rho}\} \,|\, \boldsymbol{\theta}_{-(\sigma,\rho)'}, \boldsymbol{h}_{1:n}, \boldsymbol{z}_{1:n}, \boldsymbol{s}_{0:n}, \boldsymbol{y}_{1,1:n}$.
12. Generate $\boldsymbol{\phi} \,|\, \boldsymbol{\theta}_{-\phi}, \boldsymbol{h}_{1:n}, \boldsymbol{z}_{1:n}, \boldsymbol{s}_{0:n}, \boldsymbol{y}_{1,1:n}$.
13. Go to 2.

Note that volatility innovation η_{t+1} can be written as

$$\eta_{t+1} = \sigma_{s_{t+1}} \rho_{s_{t+1}} \epsilon_t + \sigma_{s_{t+1}} \sqrt{1 - \rho_{s_{t+1}}^2} \kappa_{t+1}, \quad \begin{pmatrix} \epsilon_t \\ \kappa_{t+1} \end{pmatrix} \backsim \mathcal{N}(0, \boldsymbol{I}_2). \tag{13}$$

This is a regression of η_{t+1} on ϵ_t, with slope coefficient $\sigma_{s_{t+1}} \rho_{s_{t+1}}$ and error variance $\sigma_{s_{t+1}}^2 (1 - \rho_{s_{t+1}}^2)$. Further, define auxiliary variables

$$\bar{h}_t = h_t - \mu_{s_t}, \quad \bar{z}_t = z_t - \mu_z, \quad \text{and} \quad \bar{y}_{1,t} = \sigma_{s_{t+1}} \rho_{s_{t+1}} (y_{1,t} \mathrm{e}^{-h_t/2} - \gamma_{s_t} \bar{z}_t)/\sqrt{z_t}.$$

In the sequel each updating step of the McMC algorithm is presented in detail.

Generation of latent states $\boldsymbol{h}_{1:n}$, $\boldsymbol{s}_{0:n}$, and transition probability matrix $\boldsymbol{P}$

Step 2. The multi-move or block sampler proposed by Shepard and Pitt (1997) and modified by Watanabe and Omori (2004), which samples from the true conditional distribution, is applied to extract latent volatility $\boldsymbol{h}_{1:n}$. Another efficient method would be the mixture sampler (e.g. KSC and Omori et al., 2007), which

[3]Implementation is in MATLAB® code developed by the author.

samples from an approximate distribution and error-corrects afterwards. Although computationally less demanding, the required linearization can not be applied to measurement equation (7). The single-move sampler proposed by Jacquier, Polson, and Rossi (1994, 2004) or JM is inherently less inefficient but computationally fast. Surveys comparing Bayesian and non-Bayesian methods in the context of SV model estimation are given by Broto and Ruiz (2004), and Bos (2011). The implemented sampler is detailed in App. A.

Step 3. Sampling of the discrete states is done with a forward filtering backward sampling (FFBS) algorithm as outlined in Chib (1996), or Kim and Nelson (1998, 1999). The forward pass calculates the filtered probabilities recursively by iterating between step 1 and 2, see below. Further define $\Psi_t \equiv \{\boldsymbol{y}_{1:t}, \boldsymbol{h}_{1:t}, \boldsymbol{z}_{1:t}, \boldsymbol{\theta}\}$, the information set available at time t to draw inference on s_t.

Forward Filter - Step 1: Given $\pi(s_{t-1}|\Psi_{t-1})$, $s_t \in \{1, \dots, M\}$, at the beginning of the t^{th} iteration, weighting terms are calculated as

$$\pi(s_t, s_{t-1}|\Psi_{t-1}) = \pi(s_t|s_{t-1})\pi(s_{t-1}|\Psi_{t-1}),$$

with $\pi(s_t|s_{t-1})$ the transition probabilities.

Forward Filter - Step 2: Once $\boldsymbol{y}_t$ is observed at the end of the t^{th} iteration, the probabilities are updated,

$$\pi(s_t|\Psi_t) = \sum_{s_{t-1}} \frac{f(\boldsymbol{y}_t, \bar{h}_t|s_t, s_{t-1}, \Psi_t)\pi(s_t, s_{t-1}|\Psi_{t-1})}{f(\boldsymbol{y}_t, \bar{h}_t|\Psi_t)},$$

with $f(\boldsymbol{y}_t, \bar{h}_t|s_t, s_{t-1}, \Psi_t)$ the joint conditional likelihood and

$$f(\boldsymbol{y}_t, \bar{h}_t|\Psi_t) = \sum_{s_t} \sum_{s_{t-1}} f(\boldsymbol{y}_t, \bar{h}_t|s_t, s_{t-1}, \Psi_t)\pi(s_t, s_{t-1}|\Psi_{t-1}).$$

The joint conditional likelihood at time t depends also on time $t-1$, due to the leverage effect. We have

$$\begin{aligned} f(\boldsymbol{y}_t, \bar{h}_t|s_t, s_{t-1}, \Psi_t) \propto \sigma_{u,s_t}^{-1} \exp\left[-\frac{(y_{2,t} - \zeta - \xi h_t)^2}{2\sigma_{u,s_t}^2} - \frac{h_t}{2} - \frac{(y_{1,t} - \gamma_{s_t}\bar{z}_t \mathrm{e}^{h_t/2})^2}{2z_t \mathrm{e}^{h_t}}\right] \\ \times \begin{cases} \frac{(1-\phi_{s_1}^2)^{1/2}}{\sigma_{s_1}} \exp\left[-\frac{(1-\phi_{s_1}^2)\bar{h}_1^2}{2\sigma_{s_1}^2}\right] & t = 1 \\ \frac{1}{\sigma_{s_t}(1-\rho_{s_t}^2)^{1/2}} \exp\left[-\frac{(\bar{h}_t - \phi_{s_t}\bar{h}_{t-1} - \bar{y}_{t-1})^2}{2\sigma_{s_t}^2(1-\rho_{s_t}^2)}\right] & t > 1. \end{cases} \end{aligned}$$

Initialization of the filter is with the unconditional state probabilities,

$$\pi(s_0) = (\boldsymbol{A}'\boldsymbol{A})^{-1}\boldsymbol{A}' \begin{pmatrix} \boldsymbol{0}_M \\ 1 \end{pmatrix}, \quad \boldsymbol{A} = \begin{pmatrix} \boldsymbol{I}_M - \boldsymbol{P}' \\ \boldsymbol{i}_M' \end{pmatrix},$$

where $\boldsymbol{0}_M$ is a $M \times 1$ vector of zeros, $\boldsymbol{I}_M$ the identity matrix of dimension M, and $\boldsymbol{i}_M$ a $M \times 1$ vector of ones.

Backward Sampling: States $\boldsymbol{s}_{0:n}$ are drawn as a block from the following conditional distribution

$$\pi(\boldsymbol{s}_{0:n}|\Psi_n) = \pi(s_n|\Psi_n) \prod_{t=0}^{n-1} \pi(s_t|s_{t+1}, \Psi_t),$$

where

$$\pi(s_t|s_{t+1}, \Psi_t) \propto \pi(s_{t+1}|s_t)\pi(s_t|\Psi_t).$$

Thus, we can sample s_t from posterior mass function

$$\pi(s_t = i|s_{t+1}, \Psi_t) = \frac{\pi(s_{t+1}|s_t = i)\pi(s_t = i|\Psi_t)}{\sum_{j=1}^{M} \pi(s_{t+1}|s_t = j)\pi(s_t = j|\Psi_t)}, \quad i = 1, \ldots, M.$$

The backward sampler is initialized with a draw from $\pi(s_n|\Psi_n)$.

Step 4. Conditional on $\boldsymbol{s}_{0:n}$, the transition probability matrix $\boldsymbol{P}$ is independent of all other variables in the model. Further, the rows $\boldsymbol{P}_{i\cdot}$ of $\boldsymbol{P}$ are independent a posteriori and are drawn from the following Dirichlet distribution,

$$\boldsymbol{P}_{i\cdot} \backsim \mathcal{D}(\varrho_{i1} + n_{i1}(\boldsymbol{s}_{0:n}), \ldots, \varrho_{iM} + n_{iM}(\boldsymbol{s}_{0:n})), \quad i = 1, \ldots, M,$$

where $n_{ij}(\boldsymbol{s}_{0:n}) = \#\{s_{t-1} = i, s_t = j\}$ counts the transitions from i to j.

Generation of $\mathcal{GH}$ skew-t parameters $\boldsymbol{z}_{1:n}$, ν, γ

Step 5. Latent variables z_t, $t = 1, \ldots, n$, are sampled one at a time, having conditional posterior

$$\pi(z_t|\cdot) \propto g(z_t) \times z_t^{-(\frac{\nu+1}{2}+1)} \exp\Big(-\frac{\nu}{2z_t}\Big),$$
$$g(z_t) = \exp\Big\{-\frac{(y_{1,t} - \gamma_{s_t}\bar{z}_t \mathrm{e}^{h_t/2})^2}{2z_t \mathrm{e}^{h_t}} - \frac{(\bar{h}_{t+1} - \phi_{s_{t+1}}\bar{h}_t - \bar{y}_{1,t})^2}{2\sigma^2_{s_{t+1}}(1-\rho^2_{s_{t+1}})}\mathrm{I}_{t<n}\Big\}.$$

Using the Metropolis-Hastings (MH) algorithm (see e.g. Chib and Greenberg, 1995), a candidate $z_t^* \backsim \mathcal{IG}((\nu+1)/2, \nu/2)$ is generated and accepted with probability $\min\{g(z_t^*)/g(z_t), 1\}$, where z_t is the current draw.

Step 6. The conditional posterior probability density of ν,

$$\pi(\nu|\cdot) \propto \pi(\nu) \times \prod_{t=1}^{n} \frac{(\nu/2)^{\nu/2}}{\Gamma(\nu/2)} z_t^{-\nu/2} \exp\Big\{-\frac{\nu}{2z_t}\Big\} \times$$
$$\exp\Big\{-\sum_{t=1}^{n} \frac{(y_{1,t} - \gamma_{s_t}\bar{z}_t \mathrm{e}^{h_t/2})^2}{2z_t \mathrm{e}^{h_t}} - \sum_{t=1}^{n-1} \frac{(\bar{h}_{t+1} - \phi_{s_{t+1}}\bar{h}_t - \bar{y}_{1,t})^2}{2\sigma^2_{s_{t+1}}(1-\rho^2_{s_{t+1}})}\Big\},$$

does not permit direct sampling. A MH algorithm is deployed, based on a normal approximation of the posterior probability density around the mode, truncated on $]4, \infty[$. Moments are calculated by constraint numerical optimization,

$$\mu_\nu = \arg\max_\nu \log(\pi(\nu|\cdot)), \quad \sigma^2_\nu = \Big\{-\frac{\partial^2 \log(\pi(\nu|\cdot))}{\partial \nu^2}\Big\}^{-1}\Big|_{\nu=\mu_\nu},$$

with proposal variance σ_ν^2 the negative inverse of the Hessian at μ_ν. Starting value for optimization is the current draw of ν. Then a proposal $\nu^* \backsim \mathcal{N}(\mu_\nu, \sigma_\nu^2)$ is drawn and accepted with probability

$$\min\left\{\frac{\pi(\nu^*|\cdot) \times f_{\mathcal{N}}(\nu|\mu_\nu, \sigma_\nu^2)}{\pi(\nu|\cdot) \times f_{\mathcal{N}}(\nu^*|\mu_\nu, \sigma_\nu^2)}, 1\right\},$$

where $f_{\mathcal{N}}(x|\mu, \sigma^2)$ denotes the normal density with mean μ and variance σ^2.

Step 7. Sampling of $\boldsymbol{\gamma}$ is straightforward using results of linear regression theory. The conditional posterior probability density of $\pi(\boldsymbol{\gamma}|\cdot)$ can be written as

$$\begin{aligned}
\pi(\boldsymbol{\gamma}|\cdot) &\propto \pi(\boldsymbol{\gamma}) \times \exp\Bigg\{-\sum_{t=1}^{n} \frac{(y_{1,t} - \gamma_{s_t}\bar{z}_t \mathrm{e}^{h_t/2})^2}{2z_t \mathrm{e}^{h_t}} - \sum_{t=1}^{n-1} \frac{(\bar{h}_{t+1} - \phi_{s_{t+1}}\bar{h}_t - \bar{y}_{1,t})^2}{2\sigma_{s_{t+1}}^2(1-\rho_{s_{t+1}}^2)}\Bigg\} \\
&= \exp\Bigg\{-\frac{1}{2}\Big[(\boldsymbol{\gamma} - \boldsymbol{\gamma}_0)'\boldsymbol{\Sigma}_{\gamma_0}^{-1}(\boldsymbol{\gamma} - \boldsymbol{\gamma}_0) + \\
&\quad (\boldsymbol{Y}_{\gamma_1} - \boldsymbol{X}_{\gamma_1}\boldsymbol{\gamma})'(\boldsymbol{Y}_{\gamma_1} - \boldsymbol{X}_{\gamma_1}\boldsymbol{\gamma}) + (\boldsymbol{Y}_{\gamma_2} - \boldsymbol{X}_{\gamma_2}\boldsymbol{\gamma})'(\boldsymbol{Y}_{\gamma_2} - \boldsymbol{X}_{\gamma_2}\boldsymbol{\gamma})\Big]\Bigg\},
\end{aligned}$$

with

$$\begin{aligned}
\boldsymbol{x}_{\gamma_{1,t}} &= \bar{z}_t \begin{bmatrix} \mathrm{I}_{s_t=1} & \cdots & \mathrm{I}_{s_t=M} \end{bmatrix} / \sqrt{z_t}, \quad y_{\gamma_{1,t}} = y_{1,t}/\sqrt{z_t \mathrm{e}^{h_t}}, \quad t = 1, \ldots, n, \\
\boldsymbol{x}_{\gamma_{2,t}} &= -\rho_{s_{t+1}}\bar{z}_t \begin{bmatrix} \mathrm{I}_{s_t=1} & \cdots & \mathrm{I}_{s_t=M} \end{bmatrix} / \sqrt{(1-\rho_{s_{t+1}}^2)z_t}, \\
y_{\gamma_{2,t}} &= \left(\frac{\bar{h}_{t+1} - \phi_{s_{t+1}}\bar{h}_t}{\sigma_{s_{t+1}}} - \frac{\rho_{s_{t+1}} y_{1,t}}{(z_t \mathrm{e}^{h_t})^{1/2}}\right) / \sqrt{1-\rho_{s_{t+1}}^2}, \qquad t = 1, \ldots, n-1,
\end{aligned}$$

the row elements of $\boldsymbol{X}_{\gamma_1}$, $\boldsymbol{Y}_{\gamma_1}$, $\boldsymbol{X}_{\gamma_2}$, and $\boldsymbol{Y}_{\gamma_2}$, respectively. Hence $\boldsymbol{\gamma}|\cdot \backsim \mathcal{N}(\boldsymbol{\mu}_\gamma, \boldsymbol{\Sigma}_\gamma)$, with

$$\boldsymbol{\mu}_\gamma = \boldsymbol{\Sigma}_\gamma \times (\boldsymbol{\Sigma}_{\gamma_0}^{-1}\boldsymbol{\gamma}_0 + \boldsymbol{X}_{\gamma_1}'\boldsymbol{Y}_{\gamma_1} + \boldsymbol{X}_{\gamma_2}'\boldsymbol{Y}_{\gamma_2}), \tag{14}$$

$$\boldsymbol{\Sigma}_\gamma = (\boldsymbol{\Sigma}_{\gamma_0}^{-1} + \boldsymbol{X}_{\gamma_1}'\boldsymbol{X}_{\gamma_1} + \boldsymbol{X}_{\gamma_2}'\boldsymbol{X}_{\gamma_2})^{-1}. \tag{15}$$

Generation of RV regression parameters ζ, ξ, $\boldsymbol{\sigma}_u$

Step 8. Sampling of $\boldsymbol{\vartheta} \equiv (\zeta, \xi)'$ is straightforward using results of linear regression theory. The conditional posterior probability is

$$\begin{aligned}
\pi(\boldsymbol{\vartheta}|\cdot) &\propto \pi(\boldsymbol{\vartheta}) \times \exp\Bigg\{-\sum_{t=1}^{n} \frac{(y_{2,t} - \zeta - \xi h_t)^2}{2\sigma_{u,s_t}^2}\Bigg\} \\
&= \exp\Bigg\{-\frac{1}{2}(\boldsymbol{\vartheta} - \boldsymbol{\vartheta}_0)'\boldsymbol{\Sigma}_{\vartheta_0}^{-1}(\boldsymbol{\vartheta} - \boldsymbol{\vartheta}_0) - \frac{1}{2}(\boldsymbol{Y}_\vartheta - \boldsymbol{X}_\vartheta\boldsymbol{\vartheta})'(\boldsymbol{Y}_\vartheta - \boldsymbol{X}_\vartheta\boldsymbol{\vartheta})\Bigg\},
\end{aligned}$$

with

$$\boldsymbol{x}_{\vartheta_t} = \sigma_{u,s_t}^{-1}\begin{bmatrix} 1 & h_t \end{bmatrix}, \quad y_{\vartheta_t} = \sigma_{u,s_t}^{-1} y_{2,t}, \quad t = 1, \ldots, n,$$

the row elements of $\boldsymbol{X}_\xi$ and $\boldsymbol{Y}_\xi$, respectively. Hence $\boldsymbol{\vartheta}|\cdot \backsim \mathcal{N}(\boldsymbol{\mu}_\vartheta, \boldsymbol{\Sigma}_\vartheta)$, where $\boldsymbol{\mu}_\vartheta$ and $\boldsymbol{\Sigma}_\vartheta$ are calculated in analogy to Eq. (14)-(15).

Step 9. Sampling of $\boldsymbol{\sigma}_u^2$ is straightforward using results of linear regression theory. The conditional posterior probability is

$$\begin{aligned}\pi(\boldsymbol{\sigma}_u^2|\cdot) &\propto \pi(\boldsymbol{\sigma}_u^2) \times \prod_{i=1}^{n} (\sigma_{u,s_t}^2)^{-1/2} \exp\Big\{ \frac{(y_{2,t} - \zeta - \xi h_t)^2}{2\sigma_{u,s_t}^2} \Big\} \\ &= \prod_{i=1}^{M} (\sigma_{u,i}^2)^{-c_i/2-1} \exp\Big\{ -\frac{d_i}{2\sigma_{u,i}^2} \Big\},\end{aligned}$$

with

$$c_i = c_{0,i} + n_{s_i}, \quad d_i = d_{0,i} + \sum_{t=1}^{n} (y_{2,t} - \zeta - \xi h_t)^2 \mathrm{I}_{s_t=i}, \quad i = 1, \ldots, M,$$

where n_{s_i} denotes the observations belonging to state i. Hence, the elements of $\boldsymbol{\sigma}_u^2$ are independent and sampled state-by-state, $\sigma_{u,i}^2|\cdot \backsim \mathcal{IG}(c_i/2, d_i/2)$.

Generation of volatility process parameters $\boldsymbol{\mu}$, $\{\boldsymbol{\sigma}, \boldsymbol{\rho}\}$, $\boldsymbol{\phi}$

Step 10. Sampling of $\boldsymbol{\mu}$ is straightforward using results of linear regression theory. The conditional posterior probability is

$$\begin{aligned}\pi(\boldsymbol{\mu}|\cdot) &\propto \pi(\boldsymbol{\mu}) \times \exp\Big\{ -\frac{(1-\phi_{s_1}^2)\bar{h}_1^2}{2\sigma_{s_1}^2} - \sum_{t=1}^{n-1} \frac{(\bar{h}_{t+1} - \phi_{s_{t+1}}\bar{h}_t - \bar{y}_{1,t})^2}{2\sigma_{s_{t+1}}^2(1-\rho_{s_{t+1}}^2)} \Big\} \\ &= \exp\Big\{ -\frac{1}{2}(\boldsymbol{\mu} - \boldsymbol{\mu}_0)'\boldsymbol{\Sigma}_{\mu_0}^{-1}(\boldsymbol{\mu} - \boldsymbol{\mu}_0) - \frac{1}{2}(\boldsymbol{Y}_\mu - \boldsymbol{X}_\mu\boldsymbol{\mu})'(\boldsymbol{Y}_\mu - \boldsymbol{X}_\mu\boldsymbol{\mu}) \Big\},\end{aligned}$$

with

$$\begin{aligned}\boldsymbol{x}_{\mu_{t+1}} &= \big[\mathrm{I}_{s_{t+1}=1} - \phi_{s_{t+1}}\mathrm{I}_{s_t=1} \quad \cdots \quad \mathrm{I}_{s_{t+1}=M} - \phi_{s_{t+1}}\mathrm{I}_{s_t=M}\big] \big/ \sqrt{\sigma_{s_{t+1}}^2(1-\rho_{s_{t+1}}^2)}, \\ y_{\mu_{t+1}} &= (h_{t+1} - \phi_{s_{t+1}}h_t - \bar{y}_{1,t}) \big/ \sqrt{\sigma_{s_{t+1}}^2(1-\rho_{s_{t+1}}^2)}, \quad t = 1, \ldots, n-1, \\ \boldsymbol{x}_{\mu_1} &= \big[\mathrm{I}_{s_1=1} \quad \cdots \quad \mathrm{I}_{s_1=M}\big] \sqrt{1-\phi_1^2}\sigma_{s_1}^{-1}, \quad y_{\mu_1} = h_1\sqrt{1-\phi_1^2}\sigma_{s_1}^{-1}\end{aligned}$$

the row elements of $\boldsymbol{X}_\mu$ and $\boldsymbol{Y}_\mu$, respectively. Hence $\boldsymbol{\mu}|\cdot \backsim \mathcal{N}(\hat{\boldsymbol{\mu}}, \boldsymbol{\Sigma}_\mu)$, where $\hat{\boldsymbol{\mu}}$ and $\boldsymbol{\Sigma}_\mu$ are calculated in analogy to Eq. (14)-(15).

Label switching in mixture models may occur when using McMC for estimation, due to the invariance of the likelihood to relabeling the components. This issue has been well investigated (see e.g. Frühwirth-Schnatter, 2001, 2006) and is dealt with by introducing an identification constraint on mode parameter $\boldsymbol{\mu}$, see Sec. 5 and 6.2-6.5.

Step 11. Sampling of $\{\boldsymbol{\sigma}, \boldsymbol{\rho}\}$ is by Cholesky decomposition as proposed in Chan and Jeliazkov (2009). Let

$$\begin{aligned}\boldsymbol{w}_{t:t+1} &= \begin{pmatrix} [y_{1,t}\mathrm{e}^{-h_t/2} - \gamma_{s_t}\bar{z}_t]/\sqrt{z_t} \\ \bar{h}_{t+1} - \phi_{s_{t+1}}\bar{h}_t \end{pmatrix} \backsim \mathcal{N}(0, \boldsymbol{\Sigma}_{s_{t+1}}), \quad t = 1, \ldots, n-1, \qquad (16) \\ w_1 &= \bar{h}_1\sqrt{1-\phi_{s_1}^2} \backsim \mathcal{N}(0, \sigma_{s_1}^2),\end{aligned}$$

with covariance $\mathbf{\Sigma}_{s_{t+1}}$ defined in Eq. (10). Applying Cholesky-type decomposition $\mathbf{\Sigma}_{s_{t+1}}^{-1} = \boldsymbol{A}'_{s_{t+1}}\boldsymbol{D}_{s_{t+1}}^{-1}\boldsymbol{A}_{s_{t+1}}$, rewrite Eq. (16) as $\boldsymbol{A}_{s_{t+1}}\boldsymbol{w}_{t:t+1} \backsim \mathcal{N}(0, \boldsymbol{D}_{s_{t+1}})$, with

$$\boldsymbol{D}_{s_{t+1}} = \begin{pmatrix} 1 & 0 \\ 0 & \lambda_{s_{t+1}} \end{pmatrix}, \quad \boldsymbol{A}_{s_{t+1}} = \begin{pmatrix} 1 & 0 \\ a_{s_{t+1}} & 1 \end{pmatrix}.$$

The reparameterization in terms of $\boldsymbol{\lambda}$ and $\boldsymbol{a}$ allows for a conjugate representation with respect to the multivariate Gaussian likelihood. Define $\boldsymbol{\omega}_{t:t+1} = \boldsymbol{A}_{s_{t+1}}\boldsymbol{w}_{t:t+1}$, and let $\boldsymbol{w} = \{\boldsymbol{w}_{1:2}, \ldots, \boldsymbol{w}_{n-1:n}\}$. Observe that w_1 has been dropped from $\boldsymbol{w}$. A state specific MH step must be applied due to the initial condition, see below. Note further that $|\mathbf{\Sigma}_{s_{t+1}}^{-1}| = |\boldsymbol{A}'_{s_{t+1}}||\boldsymbol{D}_{s_{t+1}}^{-1}||\boldsymbol{A}_{s_{t+1}}| = |\boldsymbol{D}_{s_{t+1}}|^{-1} = \lambda_{s_{t+1}}^{-1}$. The conditional likelihood without initial condition can now be expressed as

$$\begin{aligned} f(\boldsymbol{w}|\cdot) &\propto \prod_{t=1}^{n-1} |\mathbf{\Sigma}_{s_{t+1}}|^{-1/2} \exp\Big\{-\frac{1}{2}\boldsymbol{w}'_{t:t+1}\mathbf{\Sigma}_{s_{t+1}}^{-1}\boldsymbol{w}_{t:t+1}\Big\} \\ &= \Big(\prod_{t=1}^{n-1} \lambda_{s_{t+1}}^{-1/2}\Big) \exp\Big\{-\frac{1}{2}\sum_{t=1}^{n-1}\boldsymbol{\omega}'_{t:t+1}\boldsymbol{D}_{s_{t+1}}^{-1}\boldsymbol{\omega}_{t:t+1}\Big\}. \end{aligned} \tag{17}$$

Rewriting the above expression in terms of λ_i, $i = 1, \ldots, M$ yields

$$\begin{aligned} f(\boldsymbol{w}|\cdot) &\propto \prod_{i=1}^{M} \lambda_i^{-n_i/2} \exp\Big\{-\frac{1}{2}\mathrm{tr}\Big(\boldsymbol{D}_i^{-1}\sum_{t=1}^{n-1}\boldsymbol{\omega}_{t:t+1}\boldsymbol{\omega}'_{t:t+1}\mathrm{I}_{s_{t+1}=i}\Big)\Big\} \\ &\propto \prod_{i=1}^{M} \lambda_i^{-n_i/2} \exp\Big\{-\frac{s_i}{2\lambda_i}\Big\}, \end{aligned}$$

with $n_i = \#\{s_t = i,\ t = 2, \ldots, n\}$ and $s_i = \sum_{t=1}^{n-1} \omega_{2,t:t+1}^2 \mathrm{I}_{s_{t+1}=i}$, where $\omega_{j,t:t+1}$ denotes the j^{th} row of $\boldsymbol{\omega}_{t:t+1}$. Hence, the conditional posterior of $\boldsymbol{\lambda}$ is

$$\pi(\boldsymbol{\lambda}|\cdot) \propto \pi(\boldsymbol{\lambda}) \times \sigma_{s_1}^{-1} \exp\Big\{-\frac{w_1^2}{2\sigma_{s_1}^2}\Big\} \prod_{i=1}^{M} \lambda_i^{-n_i/2} \exp\Big\{-\frac{s_i}{2\lambda_i}\Big\}.$$

Under the inverse gamma prior in Eq. (12), sampling of λ_i is state-by-state,

$$\lambda_i|\cdot \propto \mathcal{IG}\left(\frac{a_{0,i} + n_i}{2}, \frac{b_{0,i} + s_i}{2}\right).$$

Note that for state $i|s_1 = i$ a MH step must be deployed, see below.

Starting with the likelihood of Eq. (17), the conditional posterior of $\boldsymbol{a}$ can be written as

$$\begin{aligned} \pi(\boldsymbol{a}|\cdot) &\propto \pi(\boldsymbol{a}) \times \sigma_{s_1}^{-1} \exp\Big\{-\frac{w_1^2}{2\sigma_{s_1}^2} - \sum_{t=1}^{n-1}\frac{(w_{2,t:t+1} + a_{s_{t+1}}w_{1,t:t+1})^2}{2\lambda_{s_{t+1}}}\Big\} \\ &= \sigma_{s_1}^{-1} \exp\Big\{-\frac{w_1^2}{2\sigma_{s_1}^2}\Big\} \times \\ &\quad \exp\Big\{-\frac{1}{2}(\boldsymbol{a} - \boldsymbol{a}_0)'\mathbf{\Sigma}_{a_0}^{-1}(\boldsymbol{a} - \boldsymbol{a}_0) - \frac{1}{2}(\boldsymbol{Y}_a - \boldsymbol{X}_a\boldsymbol{a})'(\boldsymbol{Y}_a - \boldsymbol{X}_a\boldsymbol{a})\Big\}, \end{aligned}$$

with

$$x_{a_t} = -w_{1,t:t+1}\lambda_{s_{t+1}}^{-1/2}, \quad y_{a_t} = w_{2,t:t+1}\lambda_{s_{t+1}}^{-1/2}, \quad t = 1, \ldots, n-1,$$

the row elements of $\boldsymbol{X}_a$ and $\boldsymbol{Y}_a$, respectively. Hence $\boldsymbol{a}|\cdot \backsim \mathcal{N}(\boldsymbol{\mu}_a, \boldsymbol{\Sigma}_a)$, where $\boldsymbol{\mu}_a$ and $\boldsymbol{\Sigma}_a$ are calculated in analogy to Eq. (14)-(15). However, as $\boldsymbol{\Sigma}_a$ is diagonal in the current implementation, the elements of $\boldsymbol{a}$ are sampled state-by-state right after the corresponding elements of $\boldsymbol{\lambda}$.

For state $i|s_1 = i$ a MH step must be applied. Corresponding proposal $\{\lambda_i^*, a_i^*\}$ is then accepted with probability

$$\min\left\{\frac{\sigma_{s_1}\exp\{-0.5w_1^2/\sigma_{s_1}^{2*}\}}{\sigma_{s_1}^*\exp\{-0.5w_1^2/\sigma_{s_1}^2\}}, 1\right\},$$

where $\sigma_{s_1}^2$ is the current state dependent variance prevailing at $t = 1$.

Step 12. The conditional posterior probability distribution of $\boldsymbol{\phi}$ can be written as

$$\begin{aligned}\pi(\boldsymbol{\phi}|\cdot) &\propto \pi(\boldsymbol{\phi}) \times \sqrt{1-\phi_{s_1}^2}\exp\left\{-\frac{(1-\phi_{s_1}^2)\bar{h}_1^2}{2\sigma_{s_1}^2} - \sum_{t=1}^{n-1}\frac{(\bar{h}_{t+1} - \phi_{s_{t+1}}\bar{h}_t - \bar{y}_{1,t})^2}{2\sigma_{s_{t+1}}^2(1-\rho_{s_{t+1}}^2)}\right\} \\ &= \pi(\boldsymbol{\phi}) \times \mathrm{c}(\boldsymbol{\phi}) \times \exp\left\{-\frac{1}{2}(\boldsymbol{Y}_\phi - \boldsymbol{X}_\phi\boldsymbol{\phi})'(\boldsymbol{Y}_\phi - \boldsymbol{X}_\phi\boldsymbol{\phi})\right\},\end{aligned}$$

where

$$\mathrm{c}(\boldsymbol{\phi}) = \sqrt{1-\phi_{s_1}^2}\exp\left\{-\frac{(1-\phi_{s_1}^2)\bar{h}_1^2}{2\sigma_{s_1}^2}\right\}$$

and

$$\begin{aligned}\boldsymbol{x}_{\phi_{t+1}} &= \bar{h}_t\left[\mathrm{I}_{s_{t+1}=1} \quad \cdots \quad \mathrm{I}_{s_{t+1}=M}\right]/\sqrt{\sigma_{s_{t+1}}^2(1-\rho_{s_{t+1}}^2)}, \\ y_{\phi_{t+1}} &= (\bar{h}_{t+1} - \bar{y}_{1,t})/\sqrt{\sigma_{s_{t+1}}^2(1-\rho_{s_{t+1}}^2)}, \quad t = 1, \ldots, n-1,\end{aligned}$$

the row elements of $\boldsymbol{X}_\phi$ and $\boldsymbol{Y}_\phi$, respectively. The MH algorithm is used to sample from this density. A candidate $\boldsymbol{\phi}^* \backsim \mathcal{TN}_{(-1,1)}(\boldsymbol{\mu}_\phi, \boldsymbol{\Sigma}_\phi)$ is generated, where $\mathcal{TN}_{(a,b)}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ denotes the truncated normal distribution on (a, b) and $\boldsymbol{\mu}_\phi$, $\boldsymbol{\Sigma}_\phi$ are calculated in analogy to Eq. (14)-(15). The proposal is then accepted with probability

$$\min\left\{\frac{\pi(\boldsymbol{\phi}^*)\,\mathrm{c}(\boldsymbol{\phi}^*)}{\pi(\boldsymbol{\phi})\,\mathrm{c}(\boldsymbol{\phi})}, 1\right\},$$

where $\boldsymbol{\phi}$ is the current draw. Experimentation has shown that when the ϕ_i's are sampled jointly and vary more significantly in value, the acceptance rate may drop. As priors $\pi(\phi_i)$, $i = 1, \ldots, M$, are independent, a more robust state-by-state sampling strategy is implemented in the current work.

4 Particle Filter

A stratified auxiliary particle filter is developed that recursively delivers draws of the latent variables $\{\alpha_t = h_t - \mu_{s_t}, z_t, s_t\}$ given parameter values $\boldsymbol{\theta}$ and observables up to time t, $\boldsymbol{y}_{1:t}$. These draws are needed to evaluate the conditional likelihood, calculate goodness-of-fit statistics, and for forecasting. Pitt and Shephard (1999a) propose the auxiliary particle filter, a sequential Monte Carlo (SMC) technique that increases efficiency in the propagation process by weighting particles according to an importance function dependent on the subsequent observation. Further, multinomial resampling in particle filters is known to increase the Monte Carlo variance of the associated estimators. A stratified approach is able to significantly reduce that variance (see e.g. Douc, Cappé, and Moulines, 2005, or Hol, Schön, and Gustafsson, 2006). For more information on particle filtering, consider Doucet, de Freitas, and Gordon (2001), or Whiteley and Johansen (2011).

Define extended state vector $\boldsymbol{x}_t \equiv (\alpha_t, s_t, z_t)'$ (for better readability, time invariant parameter set $\boldsymbol{\theta}$ may be dropped from the conditioning arguments in this section). The proposed model is a nonlinear and non-Gaussian state space model with measurement density

$$\begin{aligned} f(\boldsymbol{y}_t|\boldsymbol{x}_t) = &\mathcal{N}(y_{1,t}|\gamma_{s_t}\bar{z}_t \exp[(\mu_{s_t}+\alpha_t)/2], z_t \exp[\mu_{s_t}+\alpha_t])\times \\ &\mathcal{N}(y_{2,t}|\zeta+\xi(\mu_{s_t}+\alpha_t), \sigma_u^2), \end{aligned}$$

transition densities

$$\begin{aligned} f(\alpha_{t+1}|s_{t+1}, \boldsymbol{x}_t, y_{1,t}) &= \mathcal{N}(\phi_{s_{t+1}}\alpha_t + \bar{y}_{1,t}, \sigma^2_{s_{t+1}}(1-\rho^2_{s_{t+1}})), \qquad (18) \\ \Pr(s_{t+1}|s_t) &= \mathrm{Multinomial}(\boldsymbol{p}_{s_t}), \end{aligned}$$

and scaling variable

$$f(z_t) = \mathcal{IG}(\nu/2, \nu/2), \qquad (19)$$

where $\boldsymbol{p}_{s_t=i} = (p_{i1}, \ldots, p_{iM})'$, $i \in M$. Applying Bayes Theorem we obtain the target posterior density

$$f(\boldsymbol{x}_{t+1}, \boldsymbol{x}_t|\boldsymbol{y}_{1:t+1}) \propto f(\boldsymbol{y}_{t+1}|\boldsymbol{x}_{t+1}) f(\boldsymbol{x}_{t+1}|\boldsymbol{x}_t, \boldsymbol{y}_{1:t}) f(\boldsymbol{x}_t|\boldsymbol{y}_{1:t}), \qquad (20)$$

with

$$f(\boldsymbol{x}_{t+1}|\boldsymbol{x}_t, \boldsymbol{y}_{1:t}) \propto f(z_{t+1}) f(\alpha_{t+1}|s_{t+1}, \boldsymbol{x}_t, \boldsymbol{y}_{1:t}) \Pr(s_{t+1}|s_t),$$

where we assume that we have samples (particles) from $f(\boldsymbol{x}_t|\boldsymbol{y}_{1:t})$ and a discrete uniform approximation $\hat{f}(\boldsymbol{x}_t|\boldsymbol{y}_{1:t})$ to $f(\boldsymbol{x}_t|\boldsymbol{y}_{1:t})$.

Samples from Eq. (20) are generated with the help of the following importance probability density function involving $\boldsymbol{y}_{t+1}$,

$$\begin{aligned} &g(\boldsymbol{x}_{t+1}, \boldsymbol{x}_t^{(i)}|\boldsymbol{y}_{1:t+1}) \\ &\quad \propto f(\boldsymbol{y}_{t+1}|\tilde{\alpha}_{t+1}^{(i,k)}, s_{t+1}, \tilde{z}) f(z_{t+1}) f(\alpha_{t+1}|s_{t+1}, \boldsymbol{x}_t^{(i)}, \boldsymbol{y}_{1:t}) \Pr(s_{t+1}|s_t^{(i)}) \hat{f}(\boldsymbol{x}_t^{(i)}|\boldsymbol{y}_{1:t}) \\ &\quad \propto f(z_{t+1}) f(\alpha_{t+1}|s_{t+1}, \boldsymbol{x}_t^{(i)}, \boldsymbol{y}_{1:t}) g(\boldsymbol{x}_t^{(i)}, s_{t+1}|\boldsymbol{y}_{1:t+1}, \tilde{\alpha}_{t+1}^{(i,k)}, \tilde{z}), \end{aligned}$$

with

$$\begin{aligned} &g(\boldsymbol{x}_t^{(i)}, s_{t+1}|\boldsymbol{y}_{1:t+1}, \tilde{\alpha}_{t+1}^{(i,k)}, \tilde{z}) \\ &\quad \propto f(\boldsymbol{y}_{t+1}|\tilde{\alpha}_{t+1}^{(i,k)}, s_{t+1}, \tilde{z}) \Pr(s_{t+1}|s_t^{(i)}) \hat{f}(\boldsymbol{x}_t^{(i)}|\boldsymbol{y}_{1:t}) \end{aligned} \qquad (21)$$

and "best" guesses

$$\tilde{\alpha}_{t+1}^{(i,k)} = \phi_{s_{t+1}}\alpha_t^{(i)} + \bar{y}_{1,t}^{(i,k)}, \quad \bar{y}_{1,t}^{(i,k)} = \bar{y}_{1,t}\big|_{\alpha_t=\alpha_t^{(i)}, s_t=s_t^{(i)}, s_{t+1}=k},$$
$$\tilde{z} = 1,$$

where superscripts $i = 1, \ldots, I$ denote particles and $k = 1, \ldots, M$ states. By including $\boldsymbol{y}_{t+1}$ in the importance function more weight is given to particles with larger predictive values. Observe that no best guess is made for s_{t+1}. Instead, $g(\boldsymbol{x}_t^{(i)}, s_{t+1}|\boldsymbol{y}_{1:t+1}, \tilde{\alpha}_{t+1}^{(i,k)}, \tilde{z})$, Eq. (21), is evaluated for every possible value of s_{t+1} (strata), resulting in a collection of $I \times M$ weighted sample points. Then, $\{\boldsymbol{x}_t, s_{t+1}\}$ are drawn jointly from the respective distribution. Accordingly, there is no direct sampling from $\Pr(s_{t+1}|s_t^{(i)})$. This leads to the following particle filter:[4]

1. Initialization, $t = 1$ $(i = 1, \ldots, I)$:

   (a) Generate $s_1^{(i)} \backsim \text{Multinomial}(\boldsymbol{p}_1)$, with $\boldsymbol{p}_1$ the initial distribution of $\boldsymbol{P}$.

   (b) Generate $\alpha_1^{(i)} \backsim \mathcal{N}(0, \sigma^2_{s_1^{(i)}}/(1 - \phi^2_{s_1^{(i)}}))$ from the initial distribution.

   (c) Generate $z_1^{(i)} \backsim \mathcal{IG}(\nu/2, \nu/2)$.

   (d) Compute $w_1^{(i)} = f(\boldsymbol{y}_1|\boldsymbol{x}_1^{(i)})$, and let $\hat{f}(\boldsymbol{x}_1^{(i)}|\boldsymbol{y}_1) = w_1^{(i)}/\sum_{j=1}^{I} w_1^{(j)}$.

2. Iterate, $t = 1, \ldots, n-1$ $(i, j = 1, \ldots, I)$:

   (a) Generate $\{\boldsymbol{x}_{t+1}^{(i)}, \boldsymbol{x}_t^{(i)}\}$ from $g(\boldsymbol{x}_{t+1}, \boldsymbol{x}_t^{(j)}|\boldsymbol{y}_{1:t+1})$, with $\boldsymbol{x}_t^{(j)}$ the current particle set at time t, as follows. First, evaluate importance weights for all possible values of $s_{t+1} = 1, \ldots, M$ (pre-weighting),

   $$w_t^{(j,k)} \propto g(\boldsymbol{x}_t^{(j)}, s_{t+1} = k|\boldsymbol{y}_{1:t+1}, \tilde{\alpha}_{t+1}^{(j,k)}, \tilde{z}),$$

   and resample $\{\boldsymbol{x}_t^{(j)}, k, w_t^{(j,k)}\}_{(j,k)\in\{1,\ldots,I\}\times\{1,\ldots,M\}}$ to obtain equally weighted $\{\boldsymbol{x}_t^{(i)}, s_{t+1}^{(i)}, I^{-1}\}$. Then generate $\alpha_{t+1}^{(i)}$ and $z_{t+1}^{(i)}$ from the densities given in Eq. (18) and (19), respectively.

   (b) Compute weights (correcting for the pre-weighting)

   $$w_{t+1}^{(i)} = \frac{f(\boldsymbol{y}_{t+1}|\boldsymbol{x}_{t+1}^{(i)})}{f(\boldsymbol{y}_{t+1}|\tilde{\alpha}_{t+1}^{(i)}, s_{t+1}^{(i)}, \tilde{z})},$$

   and let $\hat{f}(\boldsymbol{x}_{t+1}^{(i)}|\boldsymbol{y}_{1:t+1}) = w_{t+1}^{(i)}/\sum_{j=1}^{I} w_{t+1}^{(j)}$.

Draws $\boldsymbol{x}_{t|t}^{(i)}$ obtained from Step 1 below are used to calculate posterior log-likelihood ordinate $\log f_{\text{post}}(\boldsymbol{y}_{1:n}|\boldsymbol{\theta}) = \sum_{t=1}^{n} \log \hat{f}(\boldsymbol{y}_t|\boldsymbol{y}_{1:t}, \boldsymbol{\theta})$. Draws $\{\alpha_{t+1|t}^{(i,k)}, z_{t+1}^{(i)}\}$ are used to simulate one-step-ahead prediction density $f(\boldsymbol{y}_{t+1}|\boldsymbol{y}_{1:t}, \boldsymbol{\theta})$, which is needed for forecasting, model diagnostics and calculation of predictive ordinate $\log f_{\text{prior}}(\boldsymbol{y}_{1:n}|\boldsymbol{\theta}) = \sum_{t=1}^{n} \log \hat{f}(\boldsymbol{y}_t|\boldsymbol{y}_{1:t-1}, \boldsymbol{\theta})$. Simulation of the one-step-ahead prediction density is now summarized by the following steps $(i, j = 1, \ldots, I)$:

[4] Implementation is in stand alone C++ code developed by the author using the Scythe statistical library (Pemstein, Quinn, and Martin, 2011).

1. Resample from $\hat{f}(\boldsymbol{x}_t^{(j)}|\boldsymbol{y}_{1:t})$ after Step 1.(d) and 2.(b) to obtain $\boldsymbol{x}_{t|t}^{(i)}$.
2. Draw a state independent random variate $\mathcal{N}(0,1)$ to calculate $\{\alpha_{t+1|t}^{(i,k)}\}_{k=1,\dots,M}$, see Eq. (18).
3. Draw $z_{t+1}^{(i)} \backsim \mathcal{IG}(\nu/2, \nu/2)$.
4. Estimate the one-step-ahead prediction density by the mixture
$$\hat{f}(\boldsymbol{y}_{t+1}|\boldsymbol{y}_{1:t}) = I^{-1}\sum_{i=1}^{I}\sum_{k=1}^{M}\Pr(s_{t+1}=k|s_t^{(i)})f(\boldsymbol{y}_{t+1}|\alpha_{t+1|t}^{(i,k)}, z_{t+1}^{(i)}, s_{t+1}=k).$$

Diebold, Gunther, and Tay (1998) show that the correct predictive density is weakly superior to all other forecasts, i.e. will be preferred, in terms of expected loss, by all forecast users regardless of their loss function. In the current context, however, the predictive ability of the model w.r.t. realized volatility is not of interest. Instead, RV is treated as an auxiliary variable that is supposed to improve prediction of the return distribution, on which the focus in modeling lies.

Testing whether the forecasting densities are correct can be done by exploiting the properties of the probability integral transform. The probability that $y_{1,t+1}^2$ will be less than observed $y_{1,t+1}^{o2}$ conditional on $\boldsymbol{y}_{1:t}$ can be estimated as

$$\Pr(y_{1,t+1}^2 \leq y_{1,t+1}^{o2}|\boldsymbol{y}_{1:t}) \cong \hat{u}_{t+1} = \\ I^{-1}\sum_{i=1}^{I}\sum_{k=1}^{M}\Pr(s_{t+1}=k|s_t^{(i)})\Pr(y_{1,t+1}^2 \leq y_{1,t+1}^{o2}|\alpha_{t+1|t}^{(i,k)}, z_{t+1}^{(i)}, s_{t+1}=k).$$

Under the null hypothesis of a correctly specified model, observed returns $y_{1,t}^o$ are random draws from observation density $\mathcal{N}(y_{1,t}|\cdot)$, compare Eq. (7), and $\hat{u}_t$ converges in distribution to independent and identically distributed uniform random variables as $I \to \infty$ (Rosenblatt 1952). These variables can be mapped into the normal distribution, by using the inverse of the normal distribution function, to give a standard sequence $\hat{n}_t$ of independently and identically distributed normal random variables, which are then transformed one-step-ahead forecasts normed by their correct standard errors. This provides a valid basis for diagnostic checking and was popularized by KSC, and in another context by Diebold et al. (1998), among others.

5 Simulation Study

Regime switching data is simulated using parameters reflecting empirical relevant values to show that the proposed sampler works well. Moreover, sensitivity of skewness and kurtosis estimates regarding prior choice for degrees of freedom parameter ν is investigated. Specifically, two regimes are simulated, one showing higher volatility and a negative skew, the other lower volatility and a positive skew. True parameter values are

$$\boldsymbol{\mu} = (1,-1)', \quad \phi = 0.95, \quad \sigma = 0.15, \quad \rho = -0.5,$$
$$\boldsymbol{\gamma} = (-1,1)', \quad \nu = 30,$$
$$\zeta = 0, \quad \xi = 1, \quad \sigma_u = 0.4,$$
$$\boldsymbol{P} = \begin{pmatrix} 0.50 & 0.50 \\ 0.20 & 0.80 \end{pmatrix}.$$

Regime switching is restricted to volatility mode μ_i and skew parameter γ_i, $i = \{1, 2\}$, for simplicity. Regime 1 is rather transitory, whereas regime 2 shows moderate persistence, resulting in approximately one third of the data belonging to regime 1 and two third to regime 2. The underlying volatility process exhibits common high persistence and leverage is present. A rather extreme choice of the different volatility modes (they are modeled on the log-scale) keeps false state assignment of observations low, demonstrating most clearly well-functioning of the sampler. Of course, less distinct regime characteristic parameters make a correct state assignment of observations increasingly difficult and hence, inference on true parameter values. Consider in this context Lo and Müller (2010) for a reflective discussion on limits of system identification and uncertainty in general.

The following independent prior distributions are assumed

$$\begin{gathered}
\boldsymbol{\mu} \backsim \mathcal{N}((0.5, -0.5)', \boldsymbol{I}_2), \quad \frac{\phi+1}{2} \backsim \mathcal{B}(20, 1.5), \\
\lambda \backsim \mathcal{IG}\left(\frac{5}{2}, \frac{0.05}{2}\right), \quad a \backsim \mathcal{N}(0, 10^{12}), \\
\boldsymbol{\gamma} \backsim \mathcal{N}(\mathbf{0}, 10^{12}\boldsymbol{I}_2), \quad \nu \backsim \mathcal{G}(16, 0.8)\mathrm{I}_{\nu>4}, \\
(\zeta, \xi)' \backsim \mathcal{N}((0,1)', \mathrm{diag}(10^2, 1)), \quad \sigma_u^2 \backsim \mathcal{IG}\left(\frac{5}{2}, \frac{0.3}{2}\right), \\
\boldsymbol{P}_{i\cdot} \backsim \mathcal{D}(1, 1),
\end{gathered}$$

where $\mathcal{B}(\cdot)$ denotes the beta and $\mathcal{G}(\cdot)$ the gamma distribution. Operator $\mathrm{diag}(\cdot)$ creates a diagonal matrix. The mildly informative priors on $\boldsymbol{\mu}$ reflect the regime identification constraint into high and low volatility state ($\mu_1 > \mu_2$) to prevent label switching (see Sec. 3, Step 10). The priors for ϕ and λ are identical to those used by KSC, the latter in the case of zero correlation. Then, volatility of volatility σ has implied mean 0.12 and standard deviation 0.05 (moments obtained by MC simulation). Accordingly, the prior on Cholesky reparameterization parameter a implies a diffuse belief in zero correlation. The prior on ϕ mirrors a belief in moderate volatility persistence with mean 0.86 and standard deviation 0.11. For skew related parameter vector $\boldsymbol{\gamma}$ a diffuse prior centered around zero is chosen. A possible motivation for such an uninformative prior would be the desire to 'let the data speak for itself'. Moreover, as will be demonstrated below, prior choice of degrees of freedom parameter ν influences higher moment estimates. Hence the choice of a very weak prior for $\boldsymbol{\gamma}$ to isolate this effect. Then, the prior for degrees of freedom parameter ν is mildly informative, having mean 20.0 and standard deviation 5.0, reflecting a conservative belief in the existence of tail risk. Priors on ζ and ξ assume no bias correction for RV and unity scaling, respectively. The prior on σ_u^2 implies a σ_u with mean 0.29 and standard deviation 0.12 (moments obtained by MC simulation), reflecting empirical relevant values. Finally, priors for each row of $\boldsymbol{P}$ are uniform over the unit simplex, expressing no specific prior belief in regime persistence.

50 datasets are simulated, each containing $2{,}000$ observations. McMC is iterated for $15{,}000$ iterations, discarding the first $5{,}000$ as burn-in, and collecting the next $10{,}000$ draws for summary statistics. In the multi-move sampler, average block size is set to 100 and number of iterations to achieve convergence to 3.

Tab. 1 reports grand averages, standard deviations, minima, and maxima of the posterior mean estimates, and average inefficiency factors (IF) (see App. C for the deployed formulae). Except for higher moment parameters $\boldsymbol{\gamma}$ and ν, grand averages

Table 1: Summary Output of Simulated Data from the MS2SVsk$t$-RV Model

	True	Mean	Stdev	Min	Max	Avg. IF
μ_1	1	0.9731	0.1010	0.7346	1.1865	16.9
μ_2	-1	-1.0045	0.0908	-1.2300	-0.7400	8.8
ϕ	0.95	0.9474	0.0084	0.9287	0.9620	7.3
σ	0.15	0.1490	0.0107	0.1232	0.1680	70.5
ρ	-0.5	-0.4822	0.0564	-0.6038	-0.3333	21.4
γ_1	-1	-0.7353	0.3873	-1.5266	0.1868	54.2
γ_2	1	0.7498	0.2873	0.1150	1.3941	73.4
ν	30	23.825	1.9319	18.786	27.844	141.0
$\mathrm{skewness}_1$	-0.2374	-0.2346	0.1231	-0.4662	0.0555	40.2
$\mathrm{skewness}_2$	0.2374	0.2352	0.0832	0.0517	0.4587	43.4
$\mathrm{kurtosis}_1$	0.3545	0.5206	0.1413	0.3179	0.8434	52.0
$\mathrm{kurtosis}_2$	0.3545	0.4834	0.0905	0.3216	0.8445	64.9
ζ	0	0.0247	0.0596	-0.1159	0.1425	91.9
ξ	1	1.0152	0.0438	0.9366	1.1033	135.7
σ_u	0.4	0.4007	0.0089	0.3814	0.4213	4.5
p_{11}	0.5	0.5027	0.0208	0.4633	0.5479	1.2
p_{12}	0.5	0.4973	0.0208	0.4521	0.5367	1.2
p_{21}	0.2	0.1981	0.0098	0.1772	0.2221	1.2
p_{22}	0.8	0.8019	0.0098	0.7779	0.8228	1.2

True value, mean, standard deviation, minimum, maximum of the posterior mean estimates, and average inefficiency factor across 50 replications.

are very close to their true values. Observe a higher standard deviation of volatility mode μ_1 compared to μ_2 and especially of skew parameter γ_1 to γ_2. Of course these are a due to less observations in state 1 (on average, $569.1/1{,}430.9$ data points are assigned to regime $1/2$), but it is clearly visible how an increase in observations leads to a stronger improvement in precision for more difficult to estimate skew parameter γ. Note also that the maximum estimated value of γ_1 is actually positive. More importantly, however, higher moment parameters $\boldsymbol{\gamma}$ and ν are clearly biased downwards in absolute value.

Prior sensitivity of degrees of freedom parameter ν for the SV model with Student-t error is known, see e.g. Nakajima and Omori (2009). For the SV model with $\mathcal{GH}$ skew-t distributed error, Nakajima and Omori (2012) show that skew parameter γ is also affected by the prior choice of ν. As skewness and kurtosis are determined jointly by parameters γ and ν, if the posterior mean of ν decreases, then γ has to decrease in absolute value as to maintain skew and heavy tailedness of the empirical return distribution. This is also observed here. Moreover, if we consider the average implied skewness and kurtosis of the estimated distribution, we infer that whereas true skewness is precisely extracted, kurtosis is overestimated. This justifies the conservative perception of the prior choice for ν in this stylized example. To investigate the point further, an alternative prior for ν is chosen,

$$\text{Prior \#2: } \nu \backsim \mathcal{G}(18, 0.6)\mathrm{I}_{\nu>4},$$

which is centered at the true value having mean 30.0 and standard deviation 7.1. Again 50 datasets are generated, as above. Results are reported in Tab. 2. Both

Table 2: Output of Simulated Data from the MS2SVsk$t$-RV Model, ν Prior #2

	True	Mean	Stdev	Min	Max	Avg. IF
γ_1	-1	-1.0315	0.4967	-2.2501	0.2560	70.5
γ_2	1	1.0424	0.4052	0.1100	1.8217	95.0
ν	30	33.031	2.7021	26.000	39.284	144.3
skewness_1	-0.2374	-0.2185	0.1024	-0.4241	0.0514	51.9
skewness_2	0.2374	0.2188	0.0786	0.0314	0.3905	59.2
kurtosis_1	0.3545	0.3726	0.0977	0.2340	0.6027	58.7
kurtosis_2	0.3545	0.3508	0.0722	0.2236	0.5594	72.2

True value, mean, standard deviation, minimum, maximum of the posterior mean estimates, and average inefficiency factor across 50 replications.

skew and kurtosis parameters are now located satisfactorily close to their true values. Specifically, kurtosis is recovered with adequate precision.

To conclude, in real world applications, some subjectivity remains as we have to impose a belief on what true kurtosis is.[5] Handling this additional freedom in a responsible way, the researcher may conservatively set the prior on ν sufficiently low to guarantee some excess kurtosis in the data.

Referring back to the results in Tab. 1, average inefficiencies (IF) are highest for degrees of freedom parameter ν and auxiliary equation bias and scale parameters ξ and ζ. Average AR/MH acceptance rates in the multi-move sampler are high regarding the rather large block size used in this study, 90.8% and 90.5%, respectively. This can be attributed to a positive effect of the auxiliary time series for inference on latent volatility. Acceptance probabilities of the MH steps sampling volatility persistence ϕ, Cholesky reparameterization parameters $\{\lambda, a\}$, degrees of freedom parameter ν, and scaling variable z_t are on average 95.6%, 99.1%, 98.0% and 86.4%, respectively. The above results suggest good mixing properties of the proposed McMC algorithm.

6 Empirical Application

6.1 Data

Proposed models are differentiated along the attributes fat tails, skew, regimes and auxiliary observation variable. S&P 500 daily return data from Thomson Reuters Datastream is used, spanning the period from Jan-02-1990 to Aug-06-2010, a total of $5{,}164$ observations or roughly 20 years. As auxiliary series, realized volatility, the range, and the VIX are deployed.[6] The latter dictates the start date of the sample and has been downloaded from the CBOE (Chicago Board of Exchange) web page. Realized volatility is calculated from tick-by-tick prices of S&P 500 futures traded "open outcry" on the CBOE. I gratefully acknowledge provision of the dataset by Francesco Audrino, Faculty of Mathematics and Statistics, University of St. Gallen.

[5] For the SV model with symmetric Student-t error and leverage, Jacquier et al. (2004) sample ν over a discrete grid, applying an uniform prior. They draw $\pi(\boldsymbol{z}_{1:n}, \nu|\cdot)$ as a block, $\pi(\boldsymbol{z}_{1:n}|\nu,\cdot)\pi(\nu|\cdot)$, rather than with a Gibbs cycle $[\pi(\boldsymbol{z}_{1:n}|\nu,\cdot), \pi(\nu|\boldsymbol{z}_{1:n},\cdot)]$, increasing efficiency. However, a marginalization over $\boldsymbol{z}_{1:n}$ is not applicable in the current asymmetric context.

[6] The CBOE Volatility Index© measures market expectations of near-term volatility conveyed by S&P 500 stock index option prices. For more information, see www.cboe.com.

Table 3: Descriptive Statistics of the S&P 500 (90/01/02-10/08/06, 5,164 obs.)

	Return	Return^2	RV	log-RV
Mean	0.0210	1.3880	1.0545	-0.5323
Stdev	1.1781	4.5657	2.1640	0.9713
Skew	-0.1970	12.728	10.904	0.5662
Kurtosis	11.847	239.89	198.59	3.6168
Min	-9.4695	0.0000	0.0331	-3.4087
Max	10.957	120.06	62.460	4.1345
$LB(15)^\dagger$	78.7	5,637.1	20,065.6	37,842.9

	Range	log-Range	VIX	log-VIX
Mean	1.3518	0.1103	1.2837	0.1818
Stdev	1.0028	0.5999	0.5248	0.3570
Skew	3.3743	0.2792	2.0115	0.5372
Kurtosis	23.067	3.1681	10.104	3.2175
Min	0.1776	-1.7283	0.5865	-0.5336
Max	11.5208	2.4442	5.0937	1.6280
$LB(15)^\dagger$	22,816.8	20,462.9	66,370.9	67,826.0

† Critical value for the $LB(15)$ test is 30.6 at the 1% significance level.

Percentage returns are continuously compounded, $100 \times (\log P_t - \log P_{t-1})$, with P_t the closing price on day t, and demeaned. Realized volatility is calculated as in Eq. (6), using 5-min returns of the contract closest to expiry minimum 1 month, applying the previous tick rule, and discarding observations prior to the open (8.30 AM) and after the close (15.15 PM).[7] The logarithmized range is calculated as $y_{2,t} = \log(P_{\text{high},t}/P_{\text{low},t} - 1)$.[8] Finally, the VIX is scaled by $252^{-1/2}$ to approximately express expected daily movement of the S&P 500 in $\%$ over the next 30-day period and logarithmized.

Descriptive statistics of the datasets are reported in Tab. 3. Considering returns, we observe common negative skew and fat tails. Note that the largest absolute return in the analyzed period is actually positive. RV, the range and the VIX all exhibit positive skew and large kurtosis, with RV the most pronounced. After logarithmizing, however, these series get much closer to normal. RV fluctuates the most and the VIX the least, as indicated by respective standard deviations, minimum and maximum values. By inspection of Fig 1, the range appears to be the least persistent (or most erratic) and the VIX the smoothest measure, with RV in-between. The period of significantly increased volatility during the financial crisis 2008/09 is remarkable.

6.2 The MS3SVsk$t$ Model

This section covers the SVsk$t$ model with skewed observation error but no auxiliary measurement equation and extends it to a regime switching MS3SVsk$t$ model featuring three states. Variants with Student-t and Gaussian error are reported in App. D, as for the models treated in Sec. 6.3-6.5. Priors are those of Sec. 5, except the following. Regarding the volatility mode of the SVsk$t$ model, $\mu_0 = -0.5$,

[7]The basic 5-min RV estimator can be taken as a baseline, against which the range and VIX are measured, having in mind that more elaborated estimators like e.g. TSRV exist.

[8]A cut-off of 10^{-3} is applied before logarithmizing to prevent a negative numerical overflow. However, this constraint is never binding for the current dataset.

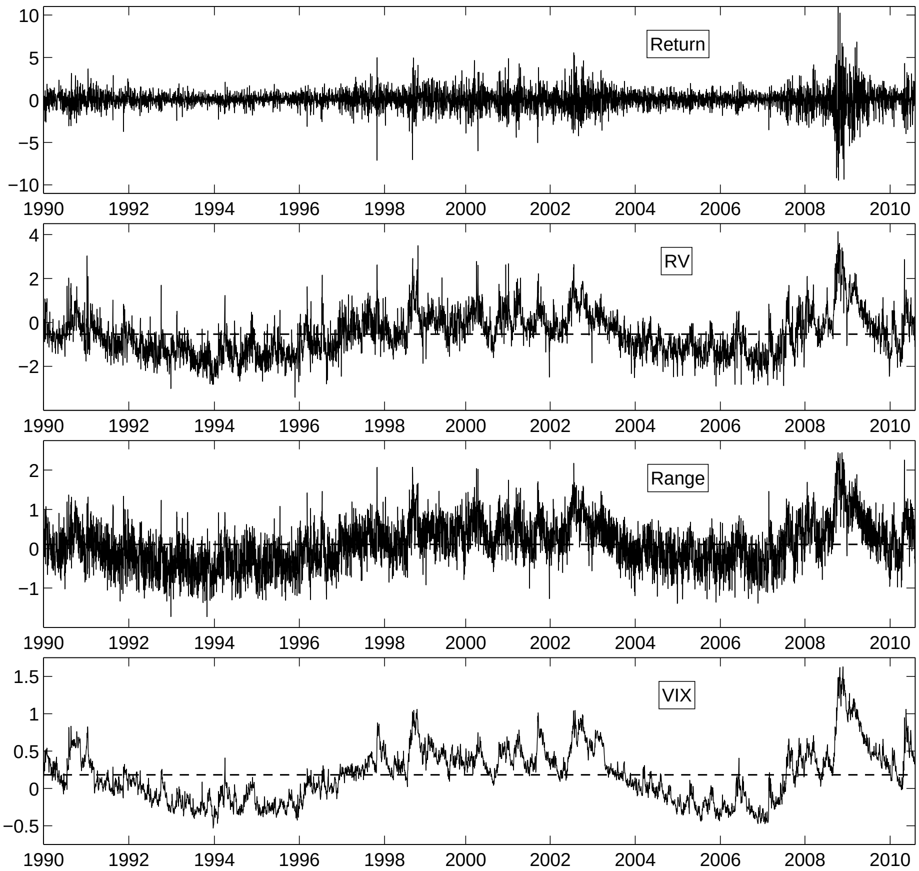

Figure 1: S&P 500 Returns and log of S&P 500 Futures RV, S&P 500 Range, and VIX. Dashed horizontal line in lower three panels is unconditional mean.

$\sigma_{\mu_0} = 1$. For the regime switching MS3SVsk$t$ variant, a mildly informative prior reflects the specific state identification constraint $\mu_1 > \mu_2 > \mu_3$,

$$\boldsymbol{\mu}_0 = (1.5, 0, -1)', \quad \boldsymbol{\Sigma}_{\mu_0} = \boldsymbol{I}_3.$$

Regarding skew parameter γ more centered priors are chosen,

$$\gamma \backsim \mathcal{N}(0, 1), \quad \boldsymbol{\gamma} \backsim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I}_3), \tag{22}$$

the reason being better predictive ability of the models, see Sec. 6.7 for a prior sensitivity analysis. Further, a moderate belief of state persistence is imposed on the transition probability matrix via the priors

$$\boldsymbol{P}_{1\cdot} \backsim \mathcal{D}(30, 1.5, 1.5), \quad \boldsymbol{P}_{2\cdot} \backsim \mathcal{D}(1.5, 30, 1.5), \quad \boldsymbol{P}_{3\cdot} \backsim \mathcal{D}(1.5, 1.5, 30),$$

implying a probability of about 90.9% to remain in the current state and of 4.5% to switch to another state.

For both models, the first $5{,}000$ draws are discarded as burn-in, collecting the following $20{,}000$ draws for parameter inference. In the multi-move sampler, average

Table 4: Estimation Results of the SVsk$t$ Model

	Mean	Stdev	95% c.b.	IF	CD
μ	-0.3959	0.1198	[-0.6341, -0.1624]	2.5	0.86
ϕ	0.9872	0.0021	[0.9827, 0.9910]	11.9	0.67
σ	0.1553	0.0111	[0.1347, 0.1785]	163.6	0.44
ρ	-0.7771	0.0306	[-0.8333, -0.7136]	101.3	0.97
γ	-0.6207	0.1752	[-1.0145, -0.3286]	161.0	0.50
ν	20.734	3.3102	[14.980, 27.734]	280.2	0.61
skewness	-0.2349	0.0488	[-0.3309, -0.1404]	68.8	0.75
kurtosis	0.5088	0.1000	[0.3368, 0.7302]	126.4	0.83

Posterior mean, standard deviation, 95% credible bounds, inefficiency factor, and Geweke's convergence diagnostic (p-value). Lower panel: Implied higher moments of ϵ_t.

Table 5: Estimation Results of the MS3SVsk$t$ Model

	Mean	Stdev	95% c.b.	IF	CD
μ_1	1.5935	0.2731	[1.0189, 2.1045]	57.7	0.91
μ_2	0.0940	0.0776	[-0.0632, 0.2371]	22.4	0.88
μ_3	-1.0971	0.0815	[-1.2628, -0.9411]	19.5	1.00
ϕ	0.9532	0.0080	[0.9357, 0.9673]	62.4	0.85
σ	0.1903	0.0160	[0.1625, 0.2224]	295.7	0.82
ρ	-0.8005	0.0312	[-0.8609, -0.7369]	133.9	0.81
γ_{12}	-0.4745	0.2092	[-0.9307, -0.1207]	100.9	0.79
γ_3	-0.7706	0.2827	[-1.3934, -0.2900]	160.8	0.98
ν	22.621	4.0860	[16.284, 31.638]	322.2	0.62
skewness$_{12}$	-0.1597	0.0597	[-0.2777, -0.0467]	52.4	0.97
skewness$_3$	-0.2603	0.0782	[-0.4144, -0.1101]	107.3	0.75
kurtosis$_{12}$	0.4051	0.0881	[0.2622, 0.5972]	149.3	0.71
kurtosis$_3$	0.5085	0.1286	[0.2983, 0.7943]	146.0	0.65

Posterior mean, standard deviation, 95% credible bounds, inefficiency factor, and Geweke's convergence diagnostic (p-value). Lower panel: Implied higher moments of ϵ_t. Transition probabilities in App. D, Tab. 19.

block size as a tuning parameter is set to 25 and number of iterations to achieve convergence to 5. Posterior means, standard deviations, 95% credible bounds, inefficiency factors (IF), and p-values of Geweke's (1992) convergence diagnostic (CD) (see App. C for the deployed formulae) are reported.

Analyzing results of the SVsk$t$ model in Tab. 4 first, we observe common high persistence of specifically $\phi = 0.987$ and a strong leverage effect $\rho = -0.777$. Skew parameter γ is significantly negative, resulting in a skewness of -0.235 and excess kurtosis of 0.510 for observation error ϵ_t. The null of the CD statistic ('chain has converged') can not be rejected for all parameters at conventional significance levels, as it can not for all parameters of the remaining models in this work.

The SVsk$t$ is extended to the MS3SVsk$t$ model, containing three regimes, with state 1 the rare event regime. However, stabilizing the latter was rather difficult (in the sense of containing a stable portion of observations during the McMC run, gauged by visual inspection), consequently leaving volatility persistence ϕ, volatility of volatility σ, and leverage ρ constant across states. Motivated by similar concerns, only one skew parameter γ_{12} is estimated for states 1 and 2 with higher

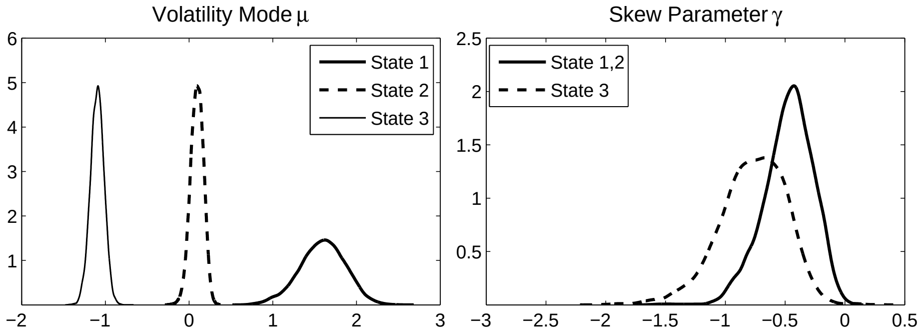

Figure 2: MS3SVsk$t$ - Selected Posterior Probability Densities

volatility. Moreover, relatively few observations in the rare event state (≈ 216 obs., see Eq. (23)) would make it difficult to obtain a more precise estimate of higher moment parameter γ_1. Note further that a Gaussian MS3SV version of the model could not be estimated without imposing a rather strong informative prior on the volatility mode of the rare event state, preventing a state collapse in that way.[9] Estimates are not reported for this variant.

Inspecting results in Tab. 5, volatility modes of the regimes are significantly different, see also Fig. 2 for a graphical visualization of selected posterior probability densities. Volatility persistence $\phi = 0.953$ is significantly lower than in the SVsk$t$ model. Interestingly, skew parameter $\gamma_3 = -0.771$ of the low volatility state is considerably more negative than $\gamma_{12} = -0.475$ of the high volatility states. Although not significantly different from a classical viewpoint, modes of the respective posterior probability densities appear rather well separated. One may thus conjecture that more data supports different skewness across regimes.

Persistence is high in all regimes:

$$\begin{array}{cc} \text{Vol} & \begin{array}{ccc} \text{high} & & \text{low} \end{array} \\ & \begin{array}{ccc} & s_t & \\ 1 & 2 & 3 \end{array} \\ s_{t-1} \begin{array}{c} 1 \\ 2 \\ 3 \end{array} & \begin{pmatrix} \mathbf{0.9766} & 0.0155 & 0.0079 \\ 0.0014 & \mathbf{0.9970} & 0.0016 \\ 0.0009 & 0.0020 & \mathbf{0.9971} \end{pmatrix} \\ & \text{avg. observations} \\ & \begin{array}{ccc} 215.9 & 2{,}633.1 & 2{,}315.1 \\ (4.2\%) & (51.0\%) & (44.8\%) \end{array} \end{array} \tag{23}$$

An observed high persistence of 97.7% for the rare event state is rather uncommon but readily explained by Fig. 3, which shows the states' posterior probability for the complete sample period. Division into the different states appears almost deterministic and most notably, the meltdown of the recent financial crisis 2008/09 is modeled by its own prolonged rare event state. This emphasizes the distinct nature of that event in recent history.

To put the analysis in a macroeconomic context, NBER US business cycle turning points are additionally graphed. Related dates are reported in Tab. 6. Several regime

[9] Consider in this context e.g. Bulla (2011), who observes that a fat tailed distribution like the Student-t can lead to an increased persistence of states.

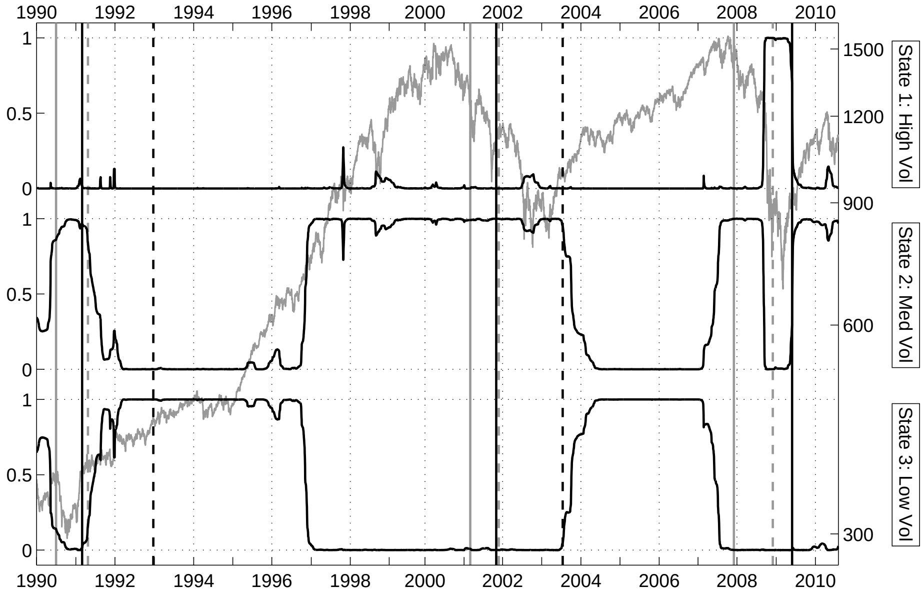

Figure 3: MS3SVsk$t$ - Posterior Probability of States (1990/01/02-2010/08/06). Background: S&P 500 (log-scale, right). Vertical lines relate to NBER US business cycles (solid: Turning points; dashed: Announcement dates; see Tab. 6 for details)

shifts match turning points (first trading day of month is taken) rather close. Most notably the trough in June 2009, where the extremely high volatility period ends, but also the peak in July 1990 and subsequent trough in March 1991, which encapsulate a high volatility period. The December 2007 peak is also reached quite soon after a shift into high volatility has been observed. Only the short recession from March to November 2001 remains undetected by the model, but it is consistently located in a high volatility period. These overlaps are remarkably, as the current analysis is purely technical but the NBER bases its decision on multiple fundamental criteria. Moreover, announcement dates of the July 1990 peak and November 2001 trough, but also of the December 2007 peak coincide with or are rather close to volatility regime shifts in the S&P 500. Especially the coincidence of the trough announcement in July 2003 with the shift to a low volatility period after more than 6 years is striking. One may suggest a significant influence of the NBER announcement on market sentiment in this case.

Table 6: NBER US Business Cycle Expansions and Contractions

Turning Point Date	Peak or Trough	Duration (months)	Announcement Date
June 2009	Trough	18	September 20, 2010
December 2007	Peak	73	December 1, 2008
November 2001	Trough	8	July 17, 2003
March 2001	Peak	120	November 26, 2001
March 1991	Trough	8	December 22, 1992
July 1990	Peak	92	April 25, 1991

US business cycle turning point and announcement dates. Contractions (recessions) start at the peak of a cycle and end at the trough. Duration is length in month from previous turning point. Source: National Bureau of Economic Research web page www.nber.org.

Table 7: Estimation Results of the SVsk$t$-RV Model

	Mean	Stdev	95% c.b.	IF	CD
μ	-0.2698	0.0928	[-0.4514, -0.0864]	1.5	0.91
ϕ	0.9694	0.0034	[0.9625, 0.9757]	4.5	0.94
σ	0.2207	0.0096	[0.2025, 0.2402]	54.0	1.00
ρ	-0.4661	0.0292	[-0.5228, -0.4080]	22.8	0.96
γ	-0.2623	0.1857	[-0.6465, 0.0820]	53.3	0.88
ν	28.771	5.0516	[20.132, 40.138]	215.1	0.96
skewness	-0.0663	0.0442	[-0.1514, 0.0207]	45.9	0.90
kurtosis	0.2667	0.0554	[0.1749, 0.3913]	180.7	0.97
ζ	-0.2159	0.0233	[-0.2616, -0.1710]	50.4	0.86
ξ	0.9040	0.0213	[0.8634, 0.9478]	41.5	0.99
σ_u	0.4083	0.0059	[0.3969, 0.4200]	7.7	0.99
SNR	0.5407	0.0276	[0.4889, 0.5973]	47.4	0.99

Posterior mean, standard deviation, 95% credible bounds, inefficiency factor, and Geweke's convergence diagnostic (p-value). Second panel: Implied higher moments of ϵ_t. Third panel: Parameters auxiliary measurement eq. Lower panel: Implied signal-to-noise ratio.

6.3 The MS4SVsk$t$-RV Model

This section covers the SVsk$t$-RV model with realized volatility as auxiliary measurement variable and a regime switching four state MS4SVsk$t$-RV variant. Priors are as in Sec. 5 and Eq. (22), except for the following. Regarding the volatility mode of the SVsk$t$-RV model, $\mu_0 = -0.5$, $\sigma_{\mu_0} = 1$. For the regime switching MS4SVsk$t$-RV extension, a mildly informative prior reflects the specific state identification constraint $\mu_1 > \mu_2 > \mu_3 > \mu_4$,

$$\boldsymbol{\mu}_0 = (1, 0, -0.5, -1.5)', \quad \boldsymbol{\Sigma}_{\mu_0} = \boldsymbol{I}_4.$$

Burn-in is $5{,}000$ draws for the SVsk$t$-RV and $10{,}000$ for the MS4SVsk$t$-RV model. The next $20{,}000$ draws are recorded for parameter inference. Average block size and number of iterations to achieve convergence in the multi-move sampler are as in Sec. 6.2.

Consider first the SVsk$t$-RV model, Tab. 7. Relative to the SVsk$t$ model with no auxiliary equation, we observe lower persistence $\phi = 0.970$, higher volatility of volatility $\sigma = 0.221$, and lower leverage $\rho = -0.466$, all differences significant. Moreover, error term ϵ_t is much less negatively skewed. Although the 95% credible bounds of skew parameter $\gamma = -0.262$ contain zero, most of the posterior probability mass is located in the negative domain, consider Fig. 4. Bias correction term ζ is negative, indicating a dominance of non-trading hours over market microstructure effects. Scaling parameter ξ lies in the vicinity of unity, as would be expected theoretically. The important positive constant $\sigma/\sigma_u = 54.1\%$ relates to the engineering concept of a signal-to-noise ratio (SNR), measuring the sustained system variance relative to the ephemeral observation variance (see e.g. West and Harrison, 1997).

Turning to the four state MS4SVsk$t$-RV variant, some experimentation was necessary to find an appropriate parameter configuration, as initial runs with the full regime switching parameter set resulted in a quite unstable state partition (especially regarding the state with lowest volatility, gauged by visual inspection).

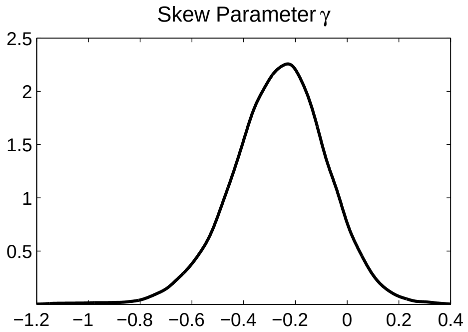

Figure 4: SVsk$t$-RV - Selected Posterior Probability Densities

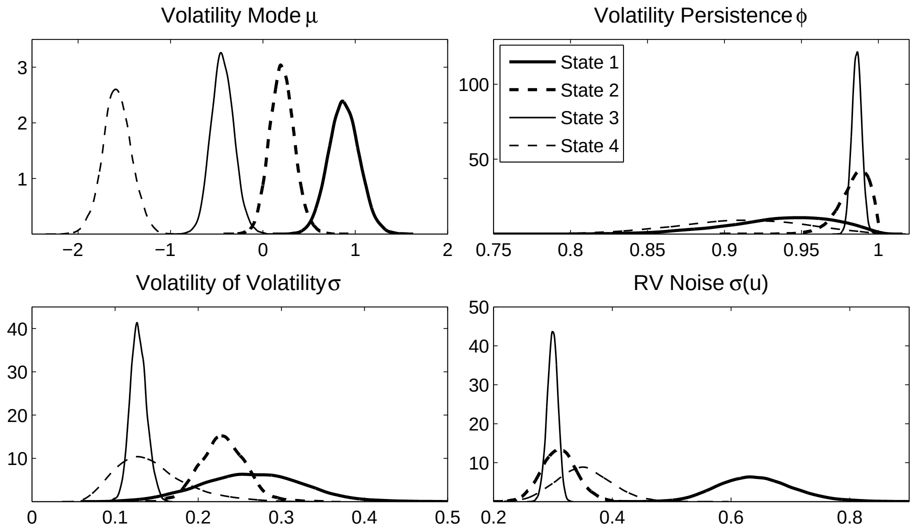

Figure 5: MS4SVsk$t$-RV - Selected Posterior Probability Densities

The first measure was then to make leverage ρ regime independent, as posterior distributions of this parameter overlapped in large part.[10] Moreover, the skew centered around zero for some states, which appeared to destabilize the system when switching continuously between positive and negative values during the McMC run. Consequently, several variants were investigated in the search for parameter configurations yielding stable equilibria, i.e. merging the skew parameter of different states or setting it to zero. Further, setting volatility persistence and volatility of volatility regime invariant. Then the DIC (see Sec. 6.6) was deployed as decision criterion to come up with the proposed parameter configuration.

Analyzing estimation results of Tab. 8, the identified four states show significantly different volatility modes, see also Fig. 5. Persistence levels appear distinct, with high and low volatility states 1/4 (which make up only for 5.6% of the data, see Eq. (24)) featuring lower persistence. Roughly speaking, volatility of volatility is separated

[10]Step 11 of the McMC algorithm has to be modified, as the combination $\{\boldsymbol{\sigma}, \rho\}$ can not be sampled using the proposed Cholesky decomposition. Instead, a MH algorithm similar to Step 6 is applied, sampling $\{\boldsymbol{\sigma}, \rho\}$ jointly.

Table 8: Estimation Results of the MS4SVsk$t$-RV Model

	Mean	Stdev	95% c.b.	IF	CD
μ_1	0.8775	0.1681	[0.5515, 1.2142]	24.6	0.99
μ_2	0.2082	0.1351	[-0.0464, 0.4866]	11.6	0.80
μ_3	-0.4356	0.1250	[-0.6711, -0.1805]	8.3	0.98
μ_4	-1.5771	0.1549	[-1.8855, -1.2675]	21.9	0.68
ϕ_1	0.9353	0.0361	[0.8542, 0.9911]	15.5	0.90
ϕ_2	0.9839	0.0102	[0.9597, 0.9982]	13.2	0.78
ϕ_3	0.9860	0.0032	[0.9797, 0.9924]	12.0	0.96
ϕ_4	0.9078	0.0422	[0.8174, 0.9819]	50.6	0.88
σ_1	0.2659	0.0606	[0.1500, 0.3858]	69.4	0.94
σ_2	0.2300	0.0271	[0.1791, 0.2858]	90.6	0.43
σ_3	0.1281	0.0100	[0.1093, 0.1488]	132.9	0.92
σ_4	0.1433	0.0425	[0.0783, 0.2461]	53.8	0.73
ρ	-0.5149	0.0323	[-0.5760, -0.4507]	37.9	0.96
γ_{23}	0.4498	0.2353	[0.0145, 0.9535]	76.7	0.94
ν	34.160	5.1186	[25.535, 45.217]	151.8	0.93
$\mathrm{skewness}_{23}$	0.0920	0.0453	[0.0031, 0.1845]	70.7	0.98
$\mathrm{kurtosis}_{14}$	0.2047	0.0347	[0.1455, 0.2786]	142.4	0.95
$\mathrm{kurtosis}_{23}$	0.2277	0.0406	[0.1582, 0.3174]	115.6	0.98
ζ	-0.1940	0.0254	[-0.2427, -0.1442]	85.2	0.99
ξ	0.8931	0.0188	[0.8561, 0.9306]	53.3	0.99
$\sigma_{u,1}$	0.6432	0.0640	[0.5282, 0.7803]	11.4	0.98
$\sigma_{u,2}$	0.3107	0.0296	[0.2543, 0.3712]	58.5	0.48
$\sigma_{u,3}$	0.2992	0.0091	[0.2807, 0.3162]	37.8	0.85
$\sigma_{u,4}$	0.3521	0.0461	[0.2660, 0.4461]	15.9	0.69
SNR_1	0.4184	0.1080	[0.2217, 0.6420]	47.6	0.95
SNR_2	0.7493	0.1286	[0.5274, 1.0298]	84.9	0.32
SNR_3	0.4285	0.0363	[0.3598, 0.5038]	108.8	1.00
SNR_4	0.4143	0.1382	[0.2138, 0.7552]	44.6	0.93

Posterior mean, standard deviation, 95% credible bounds, inefficiency factor, and Geweke's convergence diagnostic (p-value). Second panel: Implied higher moments of ϵ_t. Third panel: Parameters auxiliary measurement eq. Lower panel: Implied signal-to-noise ratios. Transition probabilities in App. D, Tab. 22.

between high 1/2 and low 3/4 volatility states. Observing a larger transition error for the high volatility states is rather intuitive. Observation error σ_u of realized volatility is distinctively high for state 1. The signal-to-noise ratios indicate that RV as an observable proxy for latent volatility is most informative in state 2, which is the second most frequent state with 17.5% of the observations. Moreover, compared to the SNR of the SVsk$t$-RV model, states 1/3/4 show a lower and state 2 a higher SNR. Interestingly, skew of medium volatility states 2/3 is now significantly positive as compared to the SVsk$t$-RV model.

It is important to note that skewness and excess kurtosis of the observation error are lowest for the RV class compared to all other models estimated in this work. This implies that more dynamics are explained by the $\mathrm{AR}(1)$ volatility process, which appears to be a desirable feature especially regarding forecasting ability.

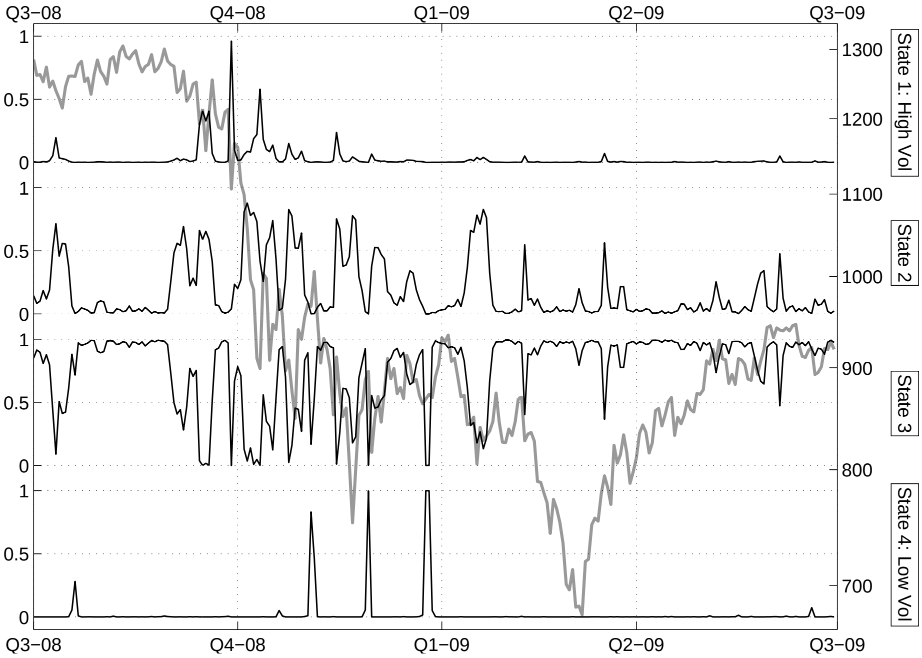

Figure 6: MS4SVsk$t$-RV - Posterior Probability of States (08/07/01-09/06/30). Background: S&P 500 (log-scale, right).

Persistence of regimes is considerably less than in the MS3SVsk$t$ Model:

$$
\begin{array}{cc}
\text{Vol} & \begin{array}{cccc} \text{high} & & & \text{low} \end{array} \\
 & s_t \\
 & \begin{array}{cccc} 1 & 2 & 3 & 4 \end{array} \\
s_{t-1} \begin{array}{c} 1 \\ 2 \\ 3 \\ 4 \end{array} & \begin{pmatrix} \mathbf{0.3512} & 0.3118 & 0.3218 & 0.0153 \\ 0.0318 & \mathbf{0.5479} & 0.4135 & 0.0068 \\ 0.0198 & 0.0845 & \mathbf{0.8747} & 0.0210 \\ 0.0448 & 0.1424 & 0.5833 & \mathbf{0.2294} \end{pmatrix} \\
 & \text{avg. observations} \\
 & \begin{array}{cccc} 170.6 & 901.2 & 3{,}976.0 & 116.2 \\ (3.3\%) & (17.5\%) & (77.0\%) & (2.2\%) \end{array}
\end{array}
\qquad (24)
$$

Posterior probability of the states are visualized for the meltdown period during the financial crisis 2008/09, see Fig. 6. First observe a distinct accumulation of probability mass in high volatility state 1 at the top of the cascade. Note also that this is the state where the information signal contained in the RV volatility proxy is buried under the largest amount of noise. As the downturn gains momentum, state 2 starts to occur more frequently. Actually, state 2 appears to dominate large part of the first half of the downturn. Signal intensity is almost twice that of the other states, mirroring a heavy information flow during the meltdown. Interestingly, in the approximately last third of the downturn, including the time when the market touches bottom, most frequent state 3 has clearly taken over again. Finally, there are a few relatively inactive periods during the second half of the bear market, located on what one may recognize as an intermediate plateau.

Table 9: Estimation Results of the SVsk$t$-Range Model

	Mean	Stdev	95% c.b.	IF	CD
μ	-0.6279	0.1102	[-0.8525, -0.4187]	13.1	0.97
ϕ	0.9577	0.0063	[0.9441, 0.9690]	38.8	0.95
σ	0.3049	0.0238	[0.2620, 0.3545]	264.2	0.95
ρ	-0.6795	0.0424	[-0.7677, -0.6036]	185.3	0.99
γ	-1.9463	0.4226	[-2.8464, -1.2076]	463.7	0.94
ν	32.135	4.7584	[23.632, 42.010]	333.9	0.91
skewness	-0.4142	0.0547	[-0.5263, -0.3112]	139.8	0.95
kurtosis	0.5915	0.1125	[0.3938, 0.8384]	131.3	0.90
ζ	0.3943	0.0315	[0.3422, 0.4653]	176.5	1.00
ξ	0.4504	0.0109	[0.4297, 0.4723]	22.2	0.96
σ_u	0.3641	0.0051	[0.3538, 0.3740]	11.7	0.97
SNR	0.8383	0.0734	[0.7081, 0.9938]	213.3	0.95

Posterior mean, standard deviation, 95% credible bounds, inefficiency factor, and Geweke's convergence diagnostic (p-value). Second panel: Implied higher moments of ϵ_t. Third panel: Parameters auxiliary measurement eq. Lower panel: Implied signal-to-noise ratio.

6.4 The MS4SVsk$t$-Range Model

Attractiveness of the range stems from its broad availability and a probably condensed information set contained in the market's intraday top and bottom. Accordingly, this section sheds light on the information content of the range as a proxy for latent volatility, analyzing the SVsk$t$-Range model and a four state MS4SVsk$t$-Range variant. Priors are as in Sec. 5 and Eq. (22), except for the following. Regarding the volatility mode of the SVsk$t$-Range model, $\mu_0 = -0.5$, $\sigma_{\mu_0} = 1$. For the regime switching MS4SVsk$t$-Range extension, a mildly informative prior is chosen that reflects the specific state identification constraint $\mu_1 > \mu_2 > \mu_3 > \mu_4$,

$$\boldsymbol{\mu}_0 = (0.5, 0, -1, -2)', \quad \boldsymbol{\Sigma}_{\mu_0} = \boldsymbol{I}_4.$$

Burn-in is $5{,}000$ draws for the SVsk$t$-Range and $10{,}000$ for the MS4SVsk$t$-Range model. The next $20{,}000$ draws are recorded for parameter inference, except for the SVsk$t$-Range model, where $40{,}000$ draws are collected. The reason for the latter measure is that especially sample paths of higher moment parameters γ, ν show strong autocorrelations for this model. The conjecture that the regime invariant version is difficult to fit to the data is further supported by failure of the SV-Range model with Gaussian error to extract a persistent volatility process ($\phi = 0.555$). Results for the latter are not reported. Average block size and number of iterations to achieve convergence in the multi-move sampler are as in Sec. 6.2.

Estimation results for the SVsk$t$-Range model are reported in Tab. 9. We observe a lower volatility mode μ compared to the SVsk$t$-RV model, but the difference is barely significant. Moreover, volatility of volatility σ (or more technically, signal strength) is significantly higher. The leverage effect can be considered still strong, $\rho = -0.680$, and is significantly higher in absolute magnitude than in the SVsk$t$-RV model. Observation error ϵ_t shows a pronounced negative skewness of -0.414, almost double in size compared to the SVsk$t$ model. Analyzing the parameters of the auxiliary measurement equation, we observe that the SNR is with 83.8% significantly higher than in the SVsk$t$-RV model, mainly due to the more intense

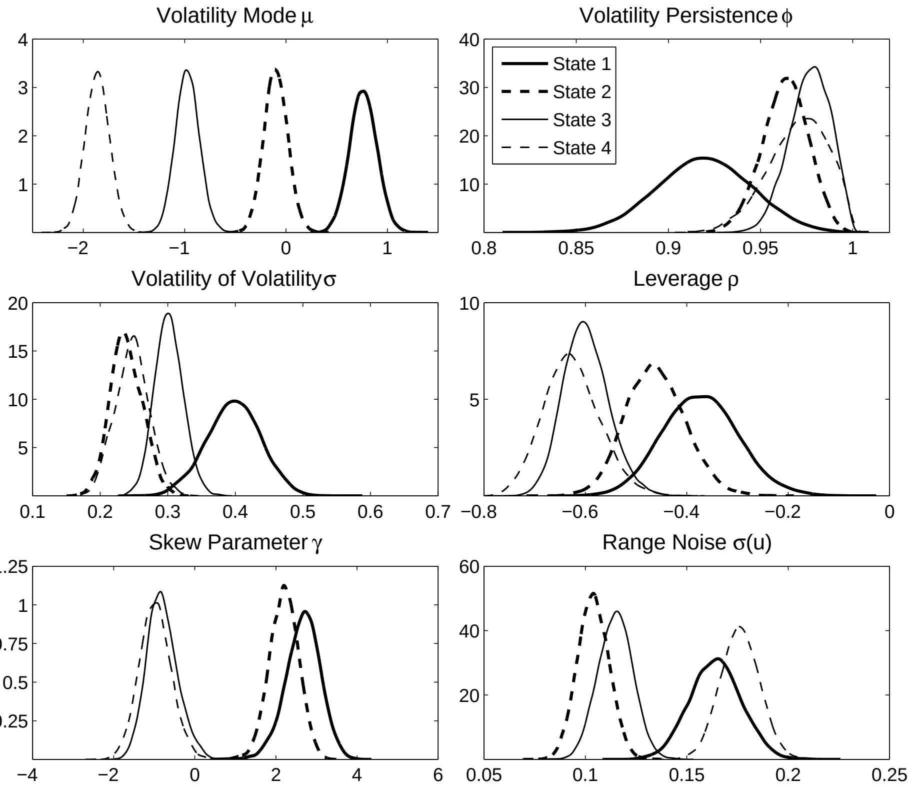

Figure 7: MS4SVsk$t$-Range - Selected Posterior Probability Densities

signal. Interestingly, bias correction term ζ is now positive. However, an explanation in terms of dominating non-trading hours or market microstructure effects as in the SVsk$t$-RV model appears inadequate.

As already mentioned above, it has been relatively difficult to extract a persistent latent volatility process for the SVsk$t$-Range model. However, the MS4sk$t$-Range model with four regimes fits the data rather naturally. Equilibrium is stable upfront without a need to make certain parameters regime invariant. Results are reported in Tab. 10, and selected posterior probability densities of the parameters are shown in Fig. 7. Extracted volatility modes are significantly different. Moreover, some interesting patterns are visible inspecting posterior density plots of the remaining parameters. Persistence ϕ and volatility of volatility σ appear to be distinctively lower respective higher for high volatility state 1. Leverage ρ is higher the lower volatility is (taking mode μ as proxy for volatility), which is in accordance with the findings of the MS3SVsk$t$ model, and appears to be grouped into states 1/2 and 3/4. This separation is even more obvious for skew parameter γ, indicating a strong positive skew for high volatility regimes 1/2 and a negative one for low volatility regimes 3/4. Regime switching range noise σ_u shows distinctive behavior as well, discriminating between more observation noise in the high/low volatility regimes and a less contaminated signal for the states featuring moderate volatility. Importantly, when we compare the signal-to-noise ratios with that of the SVsk$t$-Range model, we recognize that they are significantly higher. This stems in large part from lower

Table 10: Estimation Results of the MS4SVsk$t$-Range Model

	Mean	Stdev	95% c.b.	IF	CD
μ_1	0.7627	0.1331	[0.5038, 1.0242]	12.0	0.93
μ_2	-0.1007	0.1166	[-0.3284, 0.1310]	6.7	0.99
μ_3	-0.9723	0.1174	[-1.1979, -0.7371]	6.5	0.92
μ_4	-1.8596	0.1198	[-2.0911, -1.6200]	7.6	0.80
ϕ_1	0.9195	0.0255	[0.8688, 0.9688]	19.2	0.99
ϕ_2	0.9633	0.0124	[0.9380, 0.9865]	34.1	0.92
ϕ_3	0.9767	0.0110	[0.9537, 0.9956]	20.6	0.94
ϕ_4	0.9701	0.0159	[0.9351, 0.9959]	33.3	0.91
σ_1	0.3975	0.0400	[0.3167, 0.4747]	51.2	0.87
σ_2	0.2391	0.0240	[0.1938, 0.2873]	112.0	0.83
σ_3	0.3017	0.0213	[0.2615, 0.3451]	71.2	0.98
σ_4	0.2473	0.0254	[0.1984, 0.2978]	69.7	0.96
ρ_1	-0.3725	0.0747	[-0.5143, -0.2201]	17.8	0.97
ρ_2	-0.4616	0.0590	[-0.5722, -0.3419]	34.7	0.90
ρ_3	-0.5986	0.0450	[-0.6811, -0.5041]	34.7	0.95
ρ_4	-0.6268	0.0555	[-0.7294, -0.5111]	27.4	0.92
γ_1	2.6680	0.4337	[1.7559, 3.4741]	37.8	0.97
γ_2	2.1983	0.3588	[1.4758, 2.8821]	61.3	0.92
γ_3	-0.8000	0.3805	[-1.4968, 0.0204]	46.9	0.94
γ_4	-0.9752	0.3948	[-1.7371, -0.1678]	28.4	0.87
ν	50.564	6.3158	[39.069, 64.236]	121.1	0.55
skewness_1	0.3378	0.0598	[0.2196, 0.4561]	52.6	0.72
skewness_2	0.2835	0.0505	[0.1854, 0.3819]	74.0	0.66
skewness_3	-0.1070	0.0532	[-0.2075, 0.0027]	56.6	0.81
skewness_4	-0.1300	0.0553	[-0.2397, -0.0212]	37.7	0.74
kurtosis_1	0.3706	0.0941	[0.2133, 0.5831]	63.7	0.68
kurtosis_2	0.3018	0.0715	[0.1835, 0.4583]	84.8	0.64
kurtosis_3	0.1614	0.0356	[0.1083, 0.2451]	86.2	0.67
kurtosis_4	0.1732	0.0423	[0.1119, 0.2751]	63.4	0.66
ζ	0.4866	0.0183	[0.4524, 0.5246]	247.1	0.73
ξ	0.4274	0.0083	[0.4120, 0.4447]	143.1	1.00
$\sigma_{u,1}$	0.1643	0.0128	[0.1396, 0.1895]	28.0	0.87
$\sigma_{u,2}$	0.1039	0.0078	[0.0889, 0.1192]	49.2	0.95
$\sigma_{u,3}$	0.1148	0.0086	[0.0981, 0.1317]	63.4	0.86
$\sigma_{u,4}$	0.1762	0.0097	[0.1574, 0.1953]	25.7	0.71
SNR_1	2.4424	0.3678	[1.7772, 3.2280]	47.2	0.84
SNR_2	2.3185	0.3239	[1.7280, 2.9978]	85.1	0.83
SNR_3	2.6464	0.3120	[2.1166, 3.3328]	75.5	0.91
SNR_4	1.4087	0.1743	[1.0851, 1.7780]	57.2	0.88

Posterior mean, standard deviation, 95% credible bounds, inefficiency factor, and Geweke's convergence diagnostic (p-value). Second panel: Implied higher moments of ϵ_t. Third panel: Parameters auxiliary measurement eq. Lower panel: Implied signal-to-noise ratios. Transition probabilities in App. D, Tab. 26.

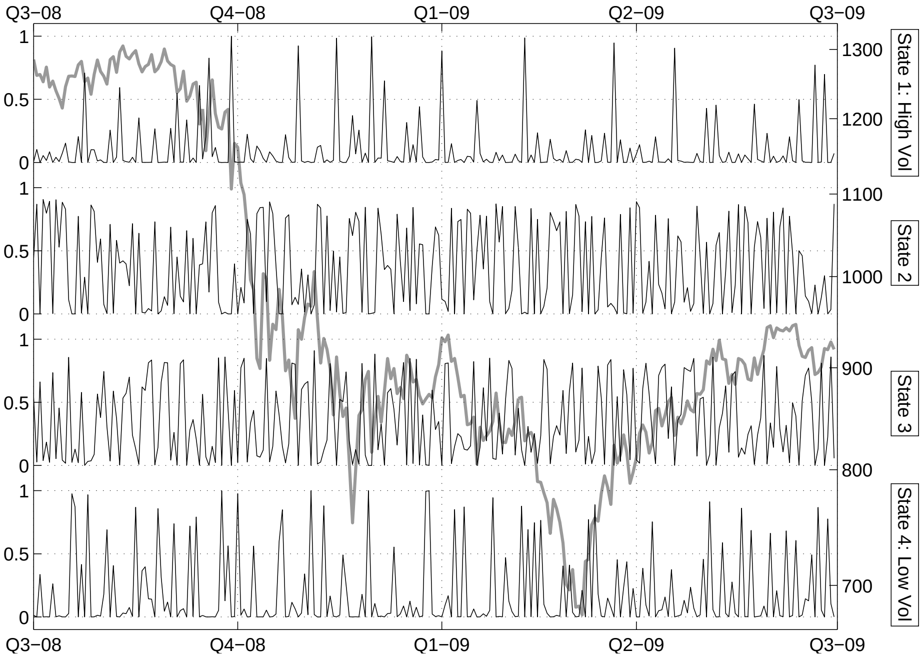

Figure 8: MS4SVsk$t$-Range - Post. Probability of States (08/07/01-09/06/30). Background: S&P 500 (log-scale).

observation errors σ_u, indicating that the model extracts far more information out of the range measure than its regime invariant counterpart.

Comparing the time series of the logarithmized range and RV in Fig. 1, a more diffuse nature of the former may let suggest heavier noise contamination. This can be made more precise now. Taking the estimation results of the MS4SVsk$t$-RV and MS4SVsk$t$-Range models as reference, one can say that RV features a less erratic information flow, meaning the received signal is subject to smaller shocks. Contamination by ephemeral observation error, on the contrary, can be relatively large. For the range the argument goes the other way around, explaining the higher signal-to-noise ratios and counteracting hereby its perception as a noisy measure. A more intense signal, however, is all else equal not advantageous for prediction, as will be made concrete in Sec. 6.6 and 6.8.

Relative to the MS4SVsk$t$-RV model, the states show an even more transient behavior, and observations are more equally divided between the regimes:

$$
\begin{array}{ccc}
\text{Vol} & & \begin{array}{cccc} \text{high} & & & \text{low} \end{array} \\
& & s_t \\
& & \begin{array}{cccc} 1 & 2 & 3 & 4 \end{array} \\
s_{t-1} & \begin{array}{c} 1 \\ 2 \\ 3 \\ 4 \end{array} & \begin{pmatrix} \mathbf{0.0952} & 0.2412 & 0.3985 & 0.2651 \\ 0.0835 & \mathbf{0.2908} & 0.3937 & 0.2320 \\ 0.1485 & 0.3621 & \mathbf{0.3003} & 0.1892 \\ 0.1552 & 0.3671 & 0.3527 & \mathbf{0.1250} \end{pmatrix} \\
& & \text{avg. observations} \\
& & \begin{array}{cccc} 628.2 & 1{,}680.0 & 1{,}826.4 & 1{,}029.4 \\ (12.2\%) & (32.5\%) & (35.4\%) & (19.9\%) \end{array}
\end{array}
$$

Posterior probability of the states are visualized for the meltdown period during the financial crisis 2008/09, see Fig. 8. Compared to the posterior state probability of the MS4SVsk$t$-RV model for the same period, Fig. 6, we observe a more erratic regime process. However, observations appear rather well divided between the states. Specifically, note some distinctive occurrences of the positively skewed high volatility state.

6.5 The MS2SVsk$t$-VIX Model

This section presents the SVsk$t$-VIX model and a two state MS2SVsk$t$-VIX extension, using the VIX as additional measurement series. In contrast to RV and the range, which measure ex-post volatility, the VIX as volatility proxy measures future expectations. It is also a much smoother measure, clearly visible from Fig. 1. However, the latter trait makes it more difficult to extract regimes and the approach is adopted in the following way. Specifically, the volatility mode has been shown inappropriate for regime identification using pilot runs and is left constant across states. Instead, states are identified by signal intensity σ. Further, auxiliary series bias ζ and scale ξ are made regime dependent. There is a convergence issue regarding these two parameters, probably a result of the aforementioned smooth characteristic of the VIX. Making them regime dependent has been found to partly alleviate this issue. Moreover, experimentation with more than two states has shown notable potential. However, convergence issues prevail, equilibria are very difficult to extract and often unstable. Consequently, a robust two state version is presented.

Discussing the SVsk$t$-VIX model first, priors are as in Sec. 5 and Eq. (22), except for the following,

$$\mu \backsim \mathcal{N}(-0.5, 1),$$
$$(\zeta, \xi)' \backsim \mathcal{N}((0, 0.5)', \mathrm{diag}(10^2, 1)), \quad \sigma_u^2 \backsim \mathcal{IG}(5/2, 0.001/2).$$

The prior on the VIX measurement noise variance implies a σ_u with mean 0.017 and standard deviation 0.007 (moments obtained by MC simulation), reflecting the smooth nature of the series. The first $10{,}000$ draws are discarded as burn-in, collecting the following $100{,}000$ draws for parameter inference. The larger number of draws reflects higher inefficiencies incurred when sampling bias correction ζ and scale ξ. In the multi-move sampler, average block size is set to 50 and number of iterations to achieve convergence to 3. The higher block size and lower number of iterations are motivated by computational cost savings. Practically, inference on latent volatility is very efficient due to only minor noise in the VIX series, keeping acceptance rates high in the block sampler also for larger block sizes.

Tab. 11 reports results. Latent process parameters μ, ϕ, σ are not significantly different from the SVsk$t$ model, whereupon volatility persistence is even slightly higher and volatility of volatility lower. This is not surprising, as the VIX is a very clean proxy for volatility. However, leverage and skew differ significantly from the former. Respective posterior densities are visualized in Fig. 9. Evidence for leverage is very weak and is not significant at the 95% level. This fits well with results of JM, who report estimates close to zero. The former authors give a possible explanation, arguing that the VIX estimator is based on future expectations of volatility and may thus incorporate the information content of the leverage parameter. Skew parameter γ is also not significantly different from zero, but most of its probability mass lies in the positive domain. This is in contrast to all

Table 11: Estimation Results of the SVsk$t$-VIX Model

	Mean	Stdev	95% c.b.	IF	CD
μ	-0.3348	0.1724	[-0.6751, 0.0036]	1.1	0.87
ϕ	0.9884	0.0021	[0.9842, 0.9925]	1.8	0.97
σ	0.1390	0.0051	[0.1291, 0.1491]	143.1	0.70
ρ	-0.0155	0.0212	[-0.0582, 0.0247]	17.1	0.89
γ	0.1359	0.0975	[-0.0455, 0.3383]	33.5	0.93
ν	16.816	2.7303	[12.269, 22.856]	378.0	0.71
skewness	0.0675	0.0462	[-0.0246, 0.1572]	25.5	0.82
kurtosis	0.5064	0.1071	[0.3311, 0.7475]	298.1	0.66
ζ	0.3131	0.0101	[0.2967, 0.3345]	1,854.1	0.43
ξ	0.3871	0.0102	[0.3692, 0.4080]	2,383.3	0.50
σ_u	0.0181	0.0020	[0.0138, 0.0219]	279.0	0.79
SNR	7.7989	1.1403	[6.0130, 10.459]	306.4	0.73

Posterior mean, standard deviation, 95% credible bounds, inefficiency factor, and Geweke's convergence diagnostic (p-value). Second panel: Implied higher moments of ϵ_t. Third panel: Parameters auxiliary measurement eq. Lower panel: Implied signal-to-noise ratio.

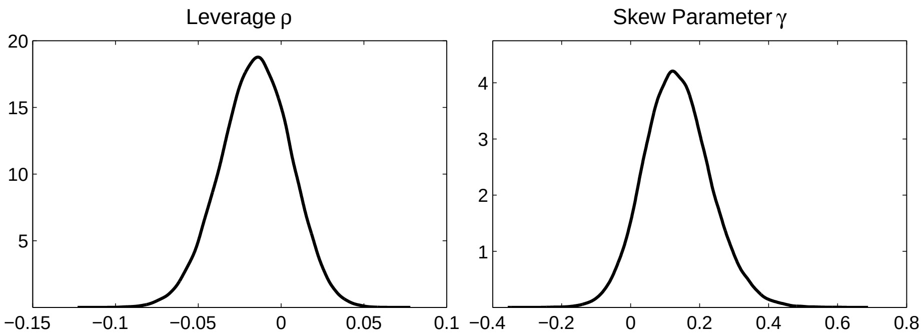

Figure 9: SVsk$t$-VIX - Selected Posterior Probability Densities

other regime invariant models analyzed, which feature negative skewness. However, kurtosis equals that of the SVsk$t$ model. Analyzing auxiliary measurement equation related parameters, bias correction ζ is positive as in the range class of models. Mentioned already, the inefficiency factors for ζ and ξ are very high, but the range around which these parameters fluctuate is rather small, as can be inferred from corresponding standard deviations and 95% credible bounds. Finally, the SNR is very high, a result of low VIX measurement noise $\sigma_u = 0.018$.

Turning to the proposed two state MS2SVsk$t$-VIX variant, priors are as in Sec. 5 and Eq. (22), except for the following,

$$\mu \backsim \mathcal{N}(-1,1),$$
$$\lambda_1 \backsim \mathcal{IG}(5/2, 0.2/2), \quad \lambda_2 \backsim \mathcal{IG}(5/2, 0.05/2),$$
$$\boldsymbol{\zeta} \backsim \mathcal{N}((0.5, 0.5)', \boldsymbol{I}_2), \quad \boldsymbol{\xi} \backsim \mathcal{N}((0.4, 0.4)', \boldsymbol{I}_2),$$
$$\sigma^2_{u,i} \backsim \mathcal{IG}(5/2, 0.001/2), \quad i = 1, 2.$$

The priors on λ_1, λ_2 reflect state identification constraint $\sigma_1 > \sigma_2$, implying means of 0.24, 0.12 and standard deviations of 0.10, 0.05 for σ_1, σ_2, respectively, in the case

Table 12: Estimation Results of the MS2SVsk$t$-VIX Model

	Mean	Stdev	95% c.b.	IF	CD
μ	-1.1072	0.1658	[-1.4544, -0.7978]	4.9	0.90
ϕ	0.9898	0.0019	[0.9860, 0.9936]	6.6	0.97
σ_1	0.2129	0.0084	[0.1971, 0.2301]	41.5	0.79
σ_2	0.0798	0.0032	[0.0736, 0.0862]	125.2	0.78
ρ_1	-0.0522	0.0412	[-0.1333, 0.0281]	5.5	0.98
ρ_2	0.0124	0.0247	[-0.0365, 0.0604]	8.0	0.95
γ_1	-7.3662	0.4298	[-8.2482, -6.5468]	136.2	0.88
γ_2	0.4561	0.2300	[0.0134, 0.9186]	71.7	0.86
ν	34.4165	3.8903	[27.3706, 42.6165]	211.3	0.97
skewness_1	-0.9235	0.0729	[-1.0791, -0.7923]	181.9	1.00
skewness_2	0.0931	0.0462	[0.0028, 0.1841]	73.4	0.88
kurtosis_1	1.8361	0.3042	[1.3293, 2.5239]	188.5	0.99
kurtosis_2	0.2241	0.0346	[0.1664, 0.3011]	151.8	0.95
ζ_1	0.5296	0.0168	[0.4981, 0.5692]	2,822.7	0.70
ζ_2	0.5271	0.0169	[0.4957, 0.5670]	11,638.5	0.44
ξ_1	0.4203	0.0103	[0.4015, 0.4408]	1,529.9	0.83
ξ_2	0.4237	0.0103	[0.4051, 0.4441]	9,189.9	0.69
$\sigma_{u,1}$	0.0316	0.0025	[0.0265, 0.0365]	55.5	0.99
$\sigma_{u,2}$	0.0090	0.0012	[0.0069, 0.0114]	261.0	0.95
SNR_1	6.7822	0.6493	[5.6830, 8.2300]	50.8	0.89
SNR_2	9.0434	1.3416	[6.7309, 11.9585]	262.2	0.87

Posterior mean, standard deviation, 95% credible bounds, inefficiency factor, and Geweke's convergence diagnostic (p-value). Second panel: Implied higher moments of ϵ_t. Third panel: Parameters auxiliary measurement eq. Lower panel: Implied signal-to-noise ratios. Transition probabilities in App. D, Tab. 31.

of zero correlation (moments obtained by MC simulation). Inefficiencies encountered when drawing inference on auxiliary measurement equation parameters ζ, ξ increase in the regime switching case even more. To account for this, $300{,}000$ draws are collected, after discarding the first $30{,}000$ as burn-in. In the multi-move sampler, average block size is set to 100 and number of iterations to achieve convergence to 3.

Tab. 12 reports results and selected posterior probability densities are shown in Fig. 10. Two regimes are clearly identified by signal intensity σ. Leverage is barely present in both states, albeit for intense signal state 1 there is a considerable portion of probability mass located in the negative domain. Volatility mode μ is significantly lower than in the SVsk$t$-VIX model, as more volatility dynamics are shifted into the tails. Strong tail characteristics are especially present in state 1, with measurement error ϵ_t featuring a negative skewness of -0.924 and excess kurtosis of 1.836. About one fourth of the observations fall into this extremely negative skewed state, see Eq. (25). Inefficiency factors for ζ, ξ are high, but corresponding standard deviations are low and 95% credible bounds narrow, indicating only minor fluctuations of these parameters. Note that ζ, ξ are almost equal across states; however, as already mentioned, merging the respective states would result in even

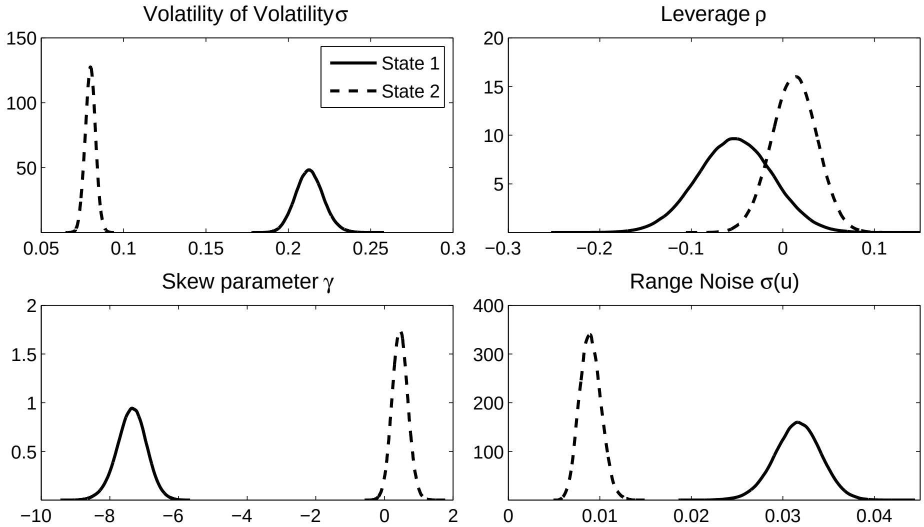

Figure 10: MS2SVsk$t$-VIX - Selected Posterior Probability Densities

higher inefficiency factors.[11] Observation noise σ_u is more than three times higher for high signal intensity state 1. Moreover, the SNR is larger in state 2, albeit not significantly, despite state 1 exhibiting a more intense signal.

Transition probabilities and observations per state are summarized:

$$\begin{array}{ccc} \text{Vol} & \text{high} & \text{low} \\ & \multicolumn{2}{c}{s_t} \\ & 1 & 2 \\ s_{t-1} \begin{array}{c} 1 \\ 2 \end{array} & \multicolumn{2}{c}{\begin{pmatrix} \mathbf{0.2701} & 0.7299 \\ 0.2471 & \mathbf{0.7529} \end{pmatrix}} \\ & \multicolumn{2}{c}{\text{avg. observations}} \\ & 1{,}304.1 & 3{,}859.9 \\ & (25.3\%) & (74.7\%) \end{array} \tag{25}$$

State 1 is transitory, with the probability to remain in this state only 27.0%. Persistence of state 2 is with 75.3% rather moderate too small. Interestingly, the model variants with Gaussian or Student-t error exhibit higher persistent states, around 88% for state 1 and 97% for state 2 (estimates are reported in App. D). From Fig. 11, which plots state posterior probability for the meltdown period during the financial crisis 2008/09, an increased occurrence of state 1 during the downturn around Q4-08 is clearly visible. This fosters the perception of this state modeling negative unexpected news flow.

6.6 Model Selection

Models are compared using the deviance information criterion (DIC) (Spiegelhalter, Best, Carlin, and van der Linde, 2002). In addition, the deviance of the in-sample predictive log-likelihood is reported to give an indication of out-of-sample forecasting

[11] Moreover, for model variants with Student-t and Gaussian error, scale parameters ξ_1 and ξ_2 have found to be significantly different, see App. D.

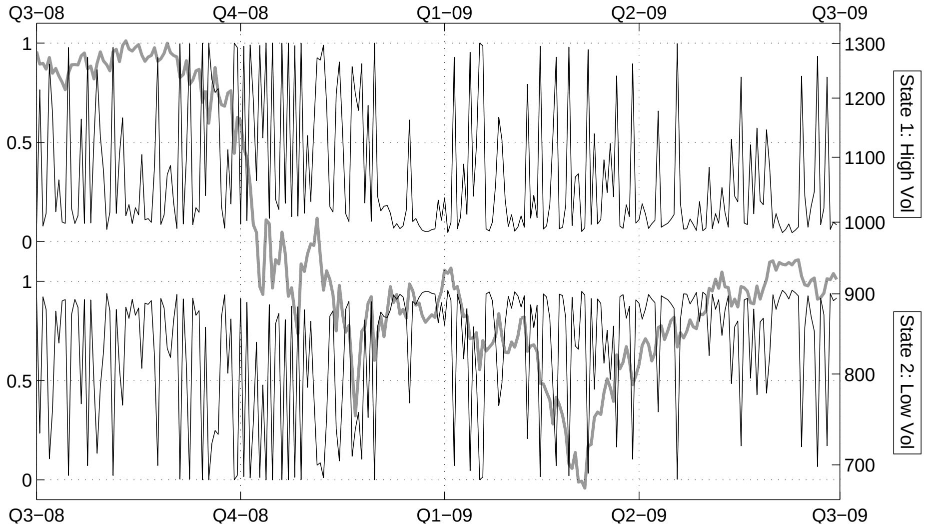

Figure 11: MS2SVsk$t$-VIX - Posterior Probability of States (08/07/01-09/06/30). Background: S&P 500 (log-scale, right).

performance. As outlined in Sec. 4, in-sample fit and predictive ability are evaluated for the return distribution only, which is the object of interest in this work. The DIC is defined by

$$DIC = \bar{D} + p_d,$$

where

$$p_d = \bar{D} - D(\bar{\boldsymbol{\theta}}), \quad \bar{D} = \mathrm{E}_{\theta|y}[D(\boldsymbol{\theta})], \quad D(\boldsymbol{\theta}) = -2 \log f(\boldsymbol{y}_{1,1:n}|\boldsymbol{\theta}),$$

with $D(\boldsymbol{\theta})$ the deviance and $\bar{\boldsymbol{\theta}}$ the posterior mean. It is a Bayesian measure of model fit, attaining smaller values for better models, with p_d a penalty term for model complexity. Intuitively, higher parameter uncertainty lets $D(\boldsymbol{\theta})$ fluctuate more widely around $D(\bar{\boldsymbol{\theta}})$, yielding larger p_D values. Numerical standard errors of the estimates are obtained the following way. Posterior expectation of the deviance $\bar{D}$, readily available as a byproduct of the McMC run, is calculated by dividing the sample in 10 equally sized batches. Regarding deviance of the posterior mean $D(\bar{\boldsymbol{\theta}})$, the particle filter outlined in Sec. 4 is repeatedly applied 10 times to obtain estimates of log-likelihood ordinate $\log f_{\text{post}}(\boldsymbol{y}_{1,1:n}|\bar{\boldsymbol{\theta}}) = \sum_{t=1}^{n} \log \hat{f}(y_{1,t}|\boldsymbol{y}_{1:t}, \bar{\boldsymbol{\theta}})$. The deviance of the predictive log-likelihood is calculated in a similar way from ordinate $\log f_{\text{prior}}(\boldsymbol{y}_{1,1:n}|\bar{\boldsymbol{\theta}}) = \sum_{t=1}^{n} \log \hat{f}(y_{1,t}|\boldsymbol{y}_{1:t-1}, \bar{\boldsymbol{\theta}})$. The above approach yields 100 different DIC values, from which statistics are calculated.

Tab. 13 reports DIC values, parameter penalty terms and the deviances of the predictive likelihoods. The regime switching range class of models, but also the SVsk$t$-Range and the MS2SVsk$t$-VIX model show distinctively low DIC values, indicating a superior in-sample fit. However, their predictive performance (in-sample) is among the worst. Specifically, the MS4SVsk$t$-Range model has the best DIC but the worst predictive likelihood. This discrepancy can in large part be attributed to the regime switching dynamics and tail behavior. First, especially in case of the range class of models, rather transient regime dynamics increase the difference between mixture forecast and a posterior realized state. Second, the above models

Table 13: Model Selection

	DIC	(s.e.)	Rank	p_D	(s.e.)	D_{pred}	(s.e.)	Rank
SV	13,761.7	(8.3)	22	311.2	(4.3)	13,799.7	(0.6)	14
SVt	13,541.2	(33.5)	18	363.5	(16.8)	13,770.3	(1.3)	12
SVsk$t$	13,434.8	(32.9)	17	486.9	(16.8)	13,747.5	(2.1)	8
MS3SVt	13,547.9	(35.2)	19	400.9	(17.7)	13,731.9	(1.7)	7
MS3SVsk$t$	13,413.2	(55.9)	15	479.5	(28.4)	13,712.1	(2.9)	5
SV-RV	13,337.4	(3.5)	12	125.8	(2.0)	13,749.9	(0.5)	9
SVt-RV	13,268.7	(12.4)	10	211.0	(6.3)	13,716.1	(0.6)	6
SVsk$t$-RV	13,235.4	(16.1)	9	217.4	(8.2)	13,697.8	(0.9)	4
MS4SV-RV	13,169.2	(4.2)	8	110.9	(2.1)	13,679.8	(0.4)	2
MS4SVt-RV	13,128.6	(11.8)	7	185.0	(5.9)	13,674.1	(0.5)	1
MS4SVsk$t$-RV	13,081.9	(29.6)	6	226.1	(14.8)	13,696.5	(0.6)	3
SVt-Range	13,285.2	(26.9)	11	480.9	(13.4)	13,893.3	(0.6)	19
SVsk$t$-Range	12,545.1	(277.4)	5	994.9	(138.7)	13,872.7	(1.8)	18
MS4SV-Range	10,899.7	(5.2)	2	96.2	(2.7)	14,032.6	(1.9)	20
MS4SVt-Range	10,942.5	(5.3)	3	124.3	(2.7)	14,035.5	(1.9)	21
MS4SVsk$t$-Range	10,708.6	(110.3)	1	531.6	(55.2)	14,073.8	(2.3)	22
SV-VIX	13,619.1	(3.1)	21	1.4	(2.9)	13,849.8	(0.8)	17
SVt-VIX	13,384.8	(42.6)	14	160.2	(22.3)	13,751.6	(1.1)	10
SVsk$t$-VIX	13,374.3	(41.7)	13	148.9	(22.8)	13,776.9	(0.9)	13
MS2SV-VIX	13,596.5	(6.0)	20	-2.5	(4.6)	13,831.2	(2.8)	16
MS2SVt-VIX	13,427.4	(39.9)	16	193.8	(20.9)	13,755.0	(3.5)	11
MS2SVsk$t$-VIX	11,917.3	(219.5)	4	1408.8	(110.7)	13,809.9	(1.0)	15

Deviance information criterion, complexity penalty term and deviance of the predictive log-likelihood. Numerical standard error in parentheses.

all feature a pronounced tail behavior. A considerable part of the volatility dynamics has been shifted into the tails. As these are unconditional within regimes, forecasting performance likely deteriorates, but a posterior fit increases. On the other side, the basic SV model has the worst DIC but performs considerably better according to the predictive criterion. The more sophisticated Markov switching and fat tailed versions in this class improve on predictive ability and in-sample fit, with the MS3SVsk$t$ the best according to both criteria. The RV class of models shows best predictive performance, most notably the MS4SVt-RV variant, and has a good to medium in-sample fit. As for the models with no auxiliary measurement equation, modeling tail behavior and regime dynamics overall improves both in-sample fit and predictive ability. However, introducing skewness in addition to regime switching and fat tails does not yield an improvement in predictive performance. Performance of the VIX class of models is mixed under both criteria. Only the MS2SVsk$t$-VIX shows a superior in-sample fit, but its predictive likelihood is rather low. The fat tailed but symmetric SVt-VIX and MS2SVt-VIX models show best predictive performance in this category.

Especially a Student-t distributed error term has a positive impact on forecasting ability. Except in the range class, these model variants perform uniformly better than their Gaussian counterparts. However, whereas the additional effect of asymmetric tails on in-sample fit is overall positive, impact on forecasting ability is mixed. For the models using no auxiliary series, introducing skewness yields an improvement. When RV or the range as additional volatility proxy is used, this is so only for the

Table 14: Prior Sensitivity of Skew Parameter γ - Model Selection Criteria

MS3SVsk$t$								
γ Prior	DIC	(s.e.)	p_D	(s.e.)	$\bar{D}$	(s.e.)	D_{pred}	(s.e.)
#1	13,413.2	(55.9)	479.5	(28.4)	12,933.8	(27.8)	13,712.1	(2.9)
#2	13,421.7	(68.0)	495.7	(34.2)	12,926.0	(33.9)	13,711.4	(3.0)

MS4SVsk$t$-Range								
γ Prior	DIC	(s.e.)	p_D	(s.e.)	$\bar{D}$	(s.e.)	D_{pred}	(s.e.)
#1	10,708.6	(110.3)	531.6	(55.2)	10,177.0	(55.2)	14,073.8	(2.3)
#2	10,561.8	(180.8)	711.6	(90.4)	9,850.3	(90.4)	14,091.4	(1.9)

Deviance information criterion, complexity penalty term, posterior mean of the deviance, and deviance of the predictive log-likelihood. Numerical standard error in parentheses.

regime invariant model variants. Modeling asymmetry in the error term for the VIX class yields an improvement relative to the Gaussian version, but the variants with Student-t error dominate significantly.

Considering complexity penalty term p_D, three points are to mention. First, a positive effect on parameter uncertainty by incorporating an additional proxy for latent volatility is clearly visible from in general higher p_D values for the models using no auxiliary equation, comparing e.g. the SV and SV-RV model. Second, modeling skewness in the error distribution substantially increases model complexity, which is especially striking for the SVsk$t$-Range and MS2SVsk$t$-VIX model. Third, using the smooth VIX series as additional time series may substantially improve parameter inference, which is most pronounced in the Gaussian case, where the p_D term even turns slightly negative for the MS2SV-VIX model. Further specification tests are provided in App. E for the interested reader.

6.7 Prior Sensitivity Analysis

Sensitivity of skewness and kurtosis regarding prior selection for degrees of freedom parameter ν has been investigated in Sec. 5. Now, prior selection for skew parameter γ and resulting consequences for in-sample fit and predictive ability are analyzed. As already mentioned, a diffuse prior for γ can be motivated by the desire 'to let the data speak for itself'. On the contrary and as in the current implementation, having an inherently time varying character of skewness in mind, one may argue for a prior more centered around zero to constrain fluctuations and potentially increasing predictive ability as a result. To investigate this point in more detail, following Prior #2 is additionally proposed (as employed in the simulation study of Sec. 5),

$$\text{Prior } \#1: \gamma \backsim \mathcal{N}(0,1), \qquad \text{Prior } \#2: \gamma \backsim \mathcal{N}(0,10^{12}).$$

Model selection criteria along the lines of Sec. 6.6 are provided for regime switching models MS3SVsk$t$ and MS4SVsk$t$-Range, which feature pronounced skewness.[12]

Tab. 14 reports results. Not very surprising, diffuse Prior #2 provides a better in-sample fit as indicated by posterior likelihood measure $\bar{D}$. However, the difference is only significant for the MS4SVsk$t$-Range model, featuring a more pronounced

[12]The MS2SVsk$t$-VIX model has not been chosen, as the already high inefficiency factors for bias and scale parameters ζ, ξ increased further using diffuse Prior #2.

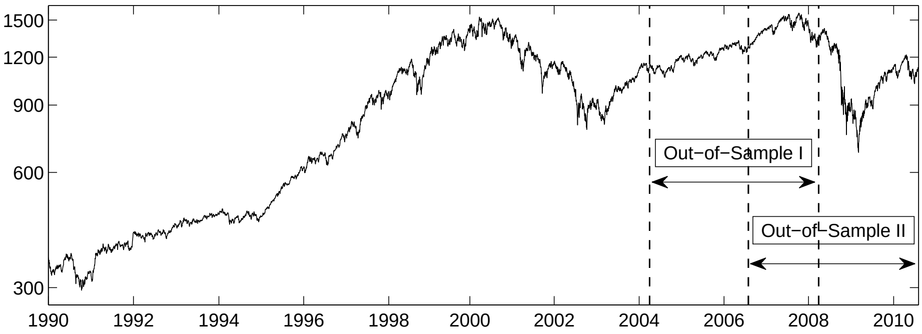

Figure 12: Partitioning of the Data In-/Out-of-Sample

skewed error term. A positive effect of less diffuse Prior $\#1$ on parameter fluctuation is indicated by lower parameter complexity penalty p_D, the effect again significant only for the MS4SVsk$t$-Range model. The lower complexity penalty leads the DIC criterion to favor the MS3SVsk$t$ model under Prior $\#1$, albeit not significantly, despite a worse in-sample fit. In contrast, lower complexity penalty can not outweigh worse in-sample fit in case of the MS4SVsk$t$-Range model, and more diffuse Prior $\#2$ is favored significantly according to the DIC. Inspecting (in-sample) predictive deviance criterion D_{pred}, more centered Prior $\#1$ yields significantly better results for the MS4SVsk$t$-Range model. In case of the less heavy skewed MS3SVsk$t$ model differences are not significant. In accordance with these positive indications regarding predictive ability, common Prior $\#1$ is also employed for all models in the out-of-sample forecasting exercise that follows.

6.8 Out-of-Sample Predictive Density Tests

This section investigates out-of-sample performance. Models are estimated on a subsample period, and forecasts are generated for two distinct out-of-sample periods spanning approximately 4 years each, see Tab. 15 for descriptive statistics and Fig. 12 for a graphical illustration. The in-sample period contains the dotcom peak and subsequent trough, 1990/01/02 - 2004/03/31, spanning $3{,}567$ observations ($14^1/_4$ years). Out-of-sample periods are chosen as to reflect a quiet period on the one side, 2004/04/01 - 2008/03/31 ($1{,}004$ obs., 4 yrs) and the volatile recent financial crisis on the other, 2006/08/01 - 2010/08/06 ($1{,}011$ obs., ≈ 4 yrs). Distinctive characteristics of the latter period are reflected in much higher volatility and kurtosis. Note that skew is negative for all samples and about the same magnitude for both out-of-sample periods. Moreover, the Ljung-Box statistic indicates far lower autocorrelation in squared returns for the more quiet period.

When generating out-of-sample forecasts, the particle filter of Sec. 4 is started 50 observations prior to the start of the respective periods to avoid any deficiencies due to unconditional initialization. Further, the out-of-sample data is demeaned over the complete period.[13] Some model variants have been fitted to the data with slight modifications or have been dropped, for the following reasons. In the MS3SVsk$t$ class with no auxiliary series, difficulties were encountered when extracting a stable third rare event regime, and thus a more precise prior $\sigma_{\mu_{0,1}} = 0.1$ on the volatility mode of state 1 has been imposed, all else equal. Consequently, the two state versions

[13] This actually induces a forward-looking bias. However, mean dynamics are generally assumed negligible when working with daily volatility.

Table 15: Descriptive Statistics In-/Out-of-Sample

	In-sample	Out-of-sample Period I	Out-of-sample Period II
Period	90/01/02 - 04/03/31	04/04/01 - 08/03/31	06/08/01 - 10/08/06
Obs.	3,567	1,004	1,011
Mean	0.0305	0.0158	-0.0125
Stdev	1.0509	0.8302	1.7109
Skew	-0.1001	-0.2293	-0.2112
Kurtosis	6.5780	5.4358	10.2863
Min	-7.1127	-3.5343	-9.4695
Max	5.573	4.1535	10.9572
$LB^2(15)$	1162.4	377.7	1210.4

The $LB^2(15)$ test is on squared return $y_{1,t}^2$, critical value is 30.6 at the 1% sig. level.

MS2SV, MS2SVt, and MS2SVsk$t$ have been estimated in addition, employing the usual mildly informative priors. Moreover, the SVsk$t$-RV variant is dropped, as the probability mass of skew parameter γ centers around zero and can not be considered significant. Further, no persistent volatility process could be extracted for the one regime SVsk$t$-Range class, irrespective of modeling the tails. These variants are also dropped. Finally, to make the selection of models more comprehensible from a forecaster's point of view, model variants using RV or the range with less than four regimes have been estimated. Thereby, respective parameter configurations were retained, and a symmetric heavy tailed error was deployed. However, these model variants have to be considered inferior relative to their four state versions according to the DIC criterion, both regarding the complete likelihood and the return distribution only (in-sample).

Tab. 16 reports log predictive density ratios relative to the basic SV model for the two out-of-sample periods for estimated models, with higher values indicating better fit. Analyzing calm forecasting period I first, performance is rather well divided between model classes. The models using no auxiliary time series perform uniformly better, with the basic SV model the least strong and the MS2SVsk$t$ variant on top. Clearly, modeling regimes and tail behavior pays off, with the asymmetric versions the best. However, a third state does not appear to deliver significant improvements, as the respective two state variants all show higher LPDRs. Recall that the third state has been enforced by a very informative prior. The RV class of models follows, with the MS4SVt-RV variant located most closely behind the baseline SV model. Interestingly, both the most basic SV-RV and sophisticated MS4SVsk$t$ model are the worst performer in this class. Moreover, it appears crucial to model symmetric fat tails especially for the one regime variant. Note a strong negative effect of modeled skewness on forecasting performance for the regime switching case, qualitatively comparable to the respective in-sample results of Sec. 6.6. The range variants are located on the lowest ranks, reinforcing the in-sample results for the range as a bad predictor. LPDRs of these models are all distinctively low. The VIX class also shows an unsatisfying forecasting ability for calm period I. Only the SVt-VIX model relatively stands out, again congruent with the respective in-sample findings.

Regarding volatile forecasting period II, performance can not be anymore differentiated between model classes as clearly as for period I. High performers are found in all categories, except again in the range class. There is a considerable improvement compared to period I for the Gaussian and Student-t version of the latter, but performance remains unsatisfying. The overall best performer is the

Table 16: Forecasting - Predictive Density Ratios

	Period I (calm)			Period II (volatile)		
	LPDR	(s.e.)	Rank	LPDR	(s.e.)	Rank
SV	0	(0.2)	9	0	(0.4)	10
SVt	3.7	(0.2)	8	2.9	(0.6)	8
SVsk$t$	7.8	(0.2)	4	8.9	(0.5)	1
MS2SV	5.8	(0.2)	6	-6.5	(0.3)	16
MS2SVt	9.1	(0.1)	2	-3.1	(0.4)	12
MS2SVsk$t$	11.9	(0.3)	1	3.2	(0.3)	7
MS3SV	4.7	(0.1)	7	-6.8	(0.3)	17
MS3SVt	6.7	(0.1)	5	-3.7	(0.2)	13
MS3SVsk$t$	8.1	(0.1)	3	-1.9	(0.3)	11
SV-RV	-11.9	(0.4)	15	-7.3	(0.3)	19
SVt-RV	-5.6	(0.6)	13	1.7	(0.3)	9
MS4SV-RV	-2.1	(0.1)	11	3.8	(0.1)	6
MS4SVt-RV	-0.8	(0.2)	10	6.9	(0.3)	2
MS4SVsk$t$-RV	-10.8	(0.1)	14	-7.2	(0.2)	18
MS4SV-Range	-29.0	(0.3)	21	-5.1	(0.3)	15
MS4SVt-Range	-29.2	(0.2)	22	-4.6	(0.4)	14
MS4SVsk$t$-Range	-31.2	(0.3)	23	-12.4	(0.3)	22
SV-VIX	-17.8	(0.5)	18	-14.0	(0.2)	23
SVt-VIX	-5.5	(0.4)	12	5.9	(0.5)	3
SVsk$t$-VIX	-20.2	(0.4)	19	-8.6	(0.6)	20
MS2SV-VIX	-20.3	(0.5)	20	-9.6	(0.7)	21
MS2SVt-VIX	-11.9	(1.1)	16	5.7	(0.6)	4
MS2SVsk$t$-VIX	-15.6	(0.2)	17	5.0	(0.3)	5

Cumulative log predictive density ratio. Numerical standard error in parentheses.

relatively simple SVsk$t$ model. This is perhaps surprising, but can be related to its relatively parsimonious specification especially in conjunction with no re-estimation. It is followed by the MS4SVt-RV model on rank 2. Notably, ranks 3 to 5 are now occupied by models using the VIX, with the SVt-VIX variant the best. This is in stark contrast to the overall inferior predictive performance of this model class found for the calmer period. JM investigate forecasting ability of SV models using RV and the VIX as additional measurement variables during the crash period 2008/09, using a sequential Monte Carlo sampler with time varying parameters. They come to qualitatively comparable results, supporting seizable information content w.r.t. volatility forecasting contained in the VIX and RV.[14] Having four competitors on the first five ranks that use an auxiliary variable and/or regime switching also puts the success of the top performing SVsk$t$ model into a broader perspective. Further, note quite heavy differences in the rankings especially within the RV and VIX classes. Similar patterns have already been observed for calmer period I, but less pronounced. In case of the VIX, modeling the tails appears essential. Concerning skewness, the big difference in performance between the SVsk$t$-VIX and the regime switching MS2SVsk$t$-VIX model indicates a crucial necessity to model the strong negatively skewed second state, mirroring bad unexpected news.

[14] JM further estimate a model using the VIX and RV simultaneously as two auxiliary equations. Considering the findings of the current paper, this clearly is an attractive option.

Regarding both out-of-sample periods and analogous to the in-sample results of Sec. 6.6, a Student-t error in general increases predictive ability. Interestingly, a skewed error term improves upon its symmetric counterpart only for the models featuring no auxiliary series in this out-of-sample study. However, the MS2SVsk$t$-VIX model capturing extreme skewness is competitive in this respect. Clearly, introducing switching regimes can increase predictive ability, but more frequent and timely re-estimation is necessary to get a clearer picture. However, the superior performance of the models using no auxiliary times series during less volatile markets is rather obvious. Moreover, results suggest that gains w.r.t. to volatility forecasting using RV or the VIX as additional time series can be expected in more turbulent market periods.

7 Conclusion

A very general SV model specification featuring leverage, heavy asymmetric tails, switching regimes and an auxiliary time series to improve inference on latent volatility is provided. The model encompasses a wide variety of models encountered in the literature. A McMC sampler and associated particle filter are proposed and outlined in detail.

Regime switching applications of model variants using RV, the range and the VIX as additional volatility proxy are presented, identifying up to four regimes from the data. Depending on the model, significantly skewed and heavy tailed errors are found. Regime switching dynamics are analyzed around the crash period 2008/09 in more detail.

In-sample model selection and out-of-sample predictive density tests are conducted. The range has been found to provide the best in-sample fit, but its predictive ability is weak. On the other side, RV and the VIX do provide seizable content with respect to volatility forecasting. This has also been found by JM in a comparable context. However, current findings suggest that the potential additional utility of RV and especially the VIX appears to pay off mostly in more turbulent market conditions. In stable environments, SV specifications using no auxiliary time series may be preferred. An overall assessment of (asymmetric) tail modeling and regime switching regarding predictive ability yields the following insights. Whereas heavy tails in general attribute positively, an asymmetric error term improves forecasting ability only for models using no auxiliary volatility proxy. An exception appears to be the skewed two state model variant using the VIX, which shows competitive performance for the volatile period. Thereby, the extremely negative skewed second state models unexpected bad news flow. Finally, the findings support the concept of regime switching as a useful tool to improve model fit and predictive performance.

Extensions are imaginable. Using various volatility measures simultaneously to draw inference on latent volatility is certainly useful. However, how to implement this in a regime switching scenario is not a straightforward question to answer. Going a step further in this direction, and considering the higher utility of RV and the VIX during turbulent market phases, making the inclusion of these measures into the model regime dependent could be another option. Then, a more elaborated estimator for realized volatility, as e.g. two scale RV (JM), can be used. Finally, one may drop the Gaussian error assumption of the auxiliary series and introduce foremost skewness to even better separate persistent information flow from spurious noise in the data.

A Multi-Move Sampler

Nakajima and Omori (2012) extend the multi-move sampler (or block sampler) with leverage described by Omori and Watanabe (2008) to handle the Generalized Hyperbolic skew-t distribution. Takahashi et al. (2009) introduce realized volatility as additional measurement equation. By defining $\alpha_t = h_t - \mu_{s_t}$, $t = 1, \ldots, n$, and $\varkappa_{s_t} = \exp(\mu_{s_t}/2)$, the following alternative representation of Eq. (7)-(9) as a state space model with respect to $\boldsymbol{\alpha}_{1:n}$ is obtained,

$$
\begin{aligned}
y_{1,t} &= (\gamma_{s_t}\bar{z}_t + \sqrt{z_t}\epsilon_t)\exp(\alpha_t/2)\varkappa_{s_t}, && && (26)\\
y_{2,t} &= \zeta + \xi(\mu_{s_t} + \alpha_t) + u_t, && t = 1, \ldots, n,\\
\alpha_{t+1} &= \phi_{s_{t+1}}\alpha_t + \eta_{t+1}, && t = 1, \ldots, n-1. && (27)
\end{aligned}
$$

To achieve more efficient sampling, $\boldsymbol{\alpha}_{1:n}$ is divided into $K+1$ blocks at random, say $(\boldsymbol{\alpha}_{k_{i-1}+1}, \ldots, \boldsymbol{\alpha}_{k_i})'$ for $i = 1, \ldots, K+1$ with $k_0 = 0$ and $k_{K+1} = n$, where K is a tuning parameter. I use stochastic knots given by $k_i = \mathrm{int}[n(i + \mathcal{U}_i)/(K+2)]$, for $i = 1, \ldots, K$ and $k_i - k_{i-1} \geq 2$, with $\mathcal{U}_i$ a random sample from the uniform distribution $\mathcal{U}(0,1)$ (Shephard and Pitt, 1997).

Suppose that $k_{i-1} = r$ and $k_i = r + d$ for the i^{th} block. The sampler then exploits the independence of disturbances η_t, block sampling $\underline{\boldsymbol{\eta}} \equiv (\eta_{r+1}, \ldots, \eta_{r+d})'$ instead of $\underline{\boldsymbol{\alpha}} \equiv (\alpha_{r+1}, \ldots, \alpha_{r+d})'$ from the respective full joint conditional posterior density ($r \geq 0$, $d \geq 2$, $r + d \leq n$),

$$
\begin{aligned}
&\pi(\underline{\boldsymbol{\eta}}|\alpha_{r+d+1}, s_{r+d+1}, \boldsymbol{\Theta}) \propto && (28)\\
&\quad\Bigg(\prod_{t=r+1}^{r+d} f(\boldsymbol{y}_t|\alpha_t, \alpha_{t+1})\Bigg)\Bigg(\prod_{t=r+1}^{r+d} f(\boldsymbol{\eta}_t)\Bigg) f(\alpha_{r+d}), \qquad r = 0,\\
&\pi(\underline{\boldsymbol{\eta}}|\alpha_r, \alpha_{r+d+1}, s_r, s_{r+d+1}, y_{1,r}, z_r, \boldsymbol{\Theta}) \propto\\
&\quad f(y_{1,r}|\alpha_r, \alpha_{r+1})\Bigg(\prod_{t=r+1}^{r+d} f(\boldsymbol{y}_t|\alpha_t, \alpha_{t+1})\Bigg)\Bigg(\prod_{t=r+1}^{r+d} f(\boldsymbol{\eta}_t)\Bigg) f(\alpha_{r+d}), \qquad r + d < n,\\
&\pi(\underline{\boldsymbol{\eta}}|\alpha_r, s_r, y_{1,r}, z_r, \boldsymbol{\Theta}) \propto && (29)\\
&\quad f(y_{1,r}|\alpha_r, \alpha_{r+1})\Bigg(\prod_{t=r+1}^{r+d-1} f(\boldsymbol{y}_t|\alpha_t, \alpha_{t+1})\Bigg) f(\boldsymbol{y}_n|\alpha_n)\Bigg(\prod_{t=r+1}^{r+d} f(\boldsymbol{\eta}_t)\Bigg), \qquad r + d = n,
\end{aligned}
$$

where $\boldsymbol{y}_t = (y_{1,t}, y_{2,t})'$ and $\boldsymbol{\Theta} \equiv \{\boldsymbol{y}_{r+1:r+d}, \boldsymbol{s}_{r+1:r+d}, \boldsymbol{z}_{r+1:r+d}, \boldsymbol{\theta}\}$. The logarithm of $f(y_{1,r}|\alpha_r, \alpha_{r+1})$, $f(\boldsymbol{y}_t|\alpha_t, \alpha_{t+1})$, $f(\boldsymbol{y}_n|\boldsymbol{\alpha}_n)$ in Eq. (28)-(29) is (excluding constant terms)

$$
l_t = -\frac{\alpha_t}{2} - \frac{(y_{1,t} - \tilde{\mu}_t)^2}{2\tilde{\sigma}_t^2} - \frac{(\tilde{y}_{2,t} - \xi\alpha_t)^2}{2\sigma_u^2}\mathrm{I}_{t>r},
$$

with

$$
\begin{aligned}
\tilde{\mu}_t &= \begin{cases} [\gamma_{s_t}\bar{z}_t + \rho_{s_{t+1}}\sqrt{z_t}(\alpha_{t+1} - \phi_{s_{t+1}}\alpha_t)/\sigma_{s_{t+1}}]\exp(\alpha_t/2)\varkappa_{s_t} & t < n\\ \gamma_{s_n}\bar{z}_n \exp(\alpha_n/2)\varkappa_{s_n} & t = n, \end{cases}\\
\tilde{\sigma}_t^2 &= \begin{cases} (1 - \rho_{s_{t+1}}^2) z_t \exp(\alpha_t)\varkappa_{s_t}^2 & t < n\\ z_n \exp(\alpha_n)\varkappa_{s_n}^2 & t = n, \end{cases}\\
\tilde{y}_{2,t} &= y_{2,t} - \zeta - \xi\mu_{s_t}.
\end{aligned}
$$

Then the logarithm of Eq. (28)-(29) can be written as $L - \sum_{t=r+1}^{r+d} \eta_t^2/(2\sigma_{s_t}^2)$, where (excluding constant terms)

$$
\begin{aligned}
L &= \log f(\alpha_{r+d})\mathrm{I}_{r+d<n} + \begin{cases} \sum_{r+1}^{r+d} l_t & r=0 \\ \sum_{r}^{r+d} l_t & r>0, \end{cases} \\
\log f(\alpha_{r+d}) &= -\frac{(\alpha_{r+d+1} - \phi_{s_{r+d+1}}\alpha_{r+d})^2}{2\sigma_{s_{r+d+1}}^2}.
\end{aligned}
$$

To construct a proposal density based on a normal approximation of the posterior density of $\underline{\boldsymbol{\eta}}$, define further

$$
\begin{aligned}
\boldsymbol{\delta} &= (\delta_{r+1}, \dots, \delta_{r+d})', \quad \delta_t = \frac{\partial L}{\partial \alpha_t}, \\
\boldsymbol{Q} &= -\mathrm{E}\left(\frac{\partial^2 L}{\partial \underline{\boldsymbol{\alpha}} \partial \underline{\boldsymbol{\alpha}}'}\right) = \begin{pmatrix} A_{r+1} & B_{r+2} & 0 & \cdots & 0 \\ B_{r+2} & A_{r+2} & B_{r+3} & \cdots & 0 \\ 0 & B_{r+3} & A_{r+3} & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & B_{r+d} \\ 0 & \cdots & 0 & B_{r+d} & A_{r+d} \end{pmatrix}, \\
A_t &= -\mathrm{E}\left(\frac{\partial^2 L}{\partial \alpha_t^2}\right), \quad t = r+1, \dots, r+d, \\
B_t &= -\mathrm{E}\left(\frac{\partial^2 L}{\partial \alpha_t \partial \alpha_{t-1}}\right), \quad t = r+2, \dots, r+d, \quad B_{r+1} = 0.
\end{aligned}
$$

The first derivatives are

$$
\begin{aligned}
\delta_t = &-\frac{1}{2} + \frac{(y_{1,t} - \tilde{\mu}_t)^2}{2\tilde{\sigma}_t^2} + \frac{y_{1,t} - \tilde{\mu}_t}{\tilde{\sigma}_t^2} \cdot \frac{\partial \tilde{\mu}_t}{\partial \alpha_t} + \frac{y_{t-1} - \tilde{\mu}_{t-1}}{\tilde{\sigma}_{t-1}^2} \cdot \frac{\partial \tilde{\mu}_{t-1}}{\partial \alpha_t} + \\
&\frac{\xi(\tilde{y}_{2,t} - \xi\alpha_t)}{\sigma_u^2} + j(\alpha_t),
\end{aligned}
$$

where

$$
\begin{aligned}
\frac{\partial \tilde{\mu}_t}{\partial \alpha_t} &= \begin{cases} \left[\frac{\gamma_{s_t}\bar{z}_t}{2} + \rho_{s_{t+1}}\sqrt{z_t}\left(-\phi_{s_{t+1}} + \frac{\alpha_{t+1} - \phi_{s_{t+1}}\alpha_t}{2}\right)/\sigma_{s_{t+1}}\right] \times \exp(\alpha_t/2)\varkappa_{s_t} & t<n \\ 0 & t=n, \end{cases} \\
\frac{\partial \tilde{\mu}_{t-1}}{\partial \alpha_t} &= \begin{cases} 0 & t=1 \\ \rho_{s_t}\sqrt{z_{t-1}}\exp(\alpha_{t-1}/2)\varkappa_{s_{t-1}}/\sigma_{s_t} & t>1, \end{cases} \\
j(\alpha_t) &= \begin{cases} \frac{\phi_{s_{t+1}}(\alpha_{t+1} - \phi_{s_{t+1}}\alpha_t)}{\sigma_{s_{t+1}}^2} & t = r+d<n \\ 0 & \text{else.} \end{cases}
\end{aligned}
$$

For the second derivatives, take expectations with respect to $\boldsymbol{y}_t$ multiplied by -1 and obtain

$$
\begin{aligned}
A_t &= \frac{1}{2} + \tilde{\sigma}_t^{-2}\left(\frac{\partial \tilde{\mu}_t}{\partial \alpha_t}\right)^2 + \tilde{\sigma}_{t-1}^{-2}\left(\frac{\partial \tilde{\mu}_{t-1}}{\partial \alpha_t}\right)^2 + \frac{\xi^2}{\sigma_u^2} + j'(\alpha_t), \\
B_t &= \tilde{\sigma}_{t-1}^{-2} \cdot \frac{\partial \tilde{\mu}_{t-1}}{\partial \alpha_{t-1}} \cdot \frac{\partial \tilde{\mu}_{t-1}}{\partial \alpha_t},
\end{aligned}
$$

where

$$j'(\alpha_t) = \begin{cases} \phi^2_{s_{r+d}}/\sigma^2_{s_{r+d}} & t = r+d < n \\ 0 & \text{else.} \end{cases}$$

Applying a second order Taylor expansion to the log of the posterior density around mode $\underline{\hat{\boldsymbol{\eta}}}$ yields an approximating normal density as follows (excluding constant terms):

$$\begin{aligned}
&\log\, \pi(\underline{\boldsymbol{\eta}}|\cdot) \\
&\approx \hat{L} + \left.\frac{\partial L}{\partial \underline{\boldsymbol{\eta}}'}\right|_{\underline{\boldsymbol{\eta}}=\underline{\hat{\boldsymbol{\eta}}}} (\underline{\boldsymbol{\eta}} - \underline{\hat{\boldsymbol{\eta}}}) + \frac{1}{2}(\underline{\boldsymbol{\eta}} - \underline{\hat{\boldsymbol{\eta}}})' \left.\mathrm{E}\left(\frac{\partial^2 L}{\partial \underline{\boldsymbol{\eta}} \partial \underline{\boldsymbol{\eta}}'}\right)\right|_{\underline{\boldsymbol{\eta}}=\underline{\hat{\boldsymbol{\eta}}}} (\underline{\boldsymbol{\eta}} - \underline{\hat{\boldsymbol{\eta}}}) - \frac{1}{2}\sum_{t=r+1}^{r+d} \frac{\eta_t^2}{\sigma^2_{s_t}} \\
&= \hat{L} + \hat{\boldsymbol{\delta}}'(\underline{\boldsymbol{\alpha}} - \underline{\hat{\boldsymbol{\alpha}}}) - \frac{1}{2}(\underline{\boldsymbol{\alpha}} - \underline{\hat{\boldsymbol{\alpha}}})'\hat{\boldsymbol{Q}}(\underline{\boldsymbol{\alpha}} - \underline{\hat{\boldsymbol{\alpha}}}) - \frac{1}{2}\sum_{t=r+1}^{r+d} \frac{\eta_t^2}{\sigma^2_{s_t}} \\
&\equiv \log\, q(\underline{\boldsymbol{\eta}}|\cdot),
\end{aligned}$$

where $\hat{L}$, $\hat{\boldsymbol{\delta}}$ and $\hat{\boldsymbol{Q}}$ are the values of L, $\boldsymbol{\delta}$ and $\boldsymbol{Q}$ at $\underline{\boldsymbol{\alpha}} = \underline{\hat{\boldsymbol{\alpha}}}$ (or at $\underline{\boldsymbol{\eta}} = \underline{\hat{\boldsymbol{\eta}}}$, equivalently). It can be shown that proposal density $q(\underline{\boldsymbol{\eta}}|\cdot)$ is the posterior of $\underline{\boldsymbol{\eta}}$ obtained from an ordinary linear state space model given by Eq. (30)-(31) below (see Omori and Watanabe, 2007, in a related context). The mode $\underline{\hat{\boldsymbol{\eta}}}$ can be obtained by repeating the following algorithm until it converges (usually 2 to 5 iterations).

Algorithm 1 (Disturbance smoother):

1. Initialize $\underline{\hat{\boldsymbol{\eta}}}$, and compute $\underline{\hat{\boldsymbol{\alpha}}}$ at $\underline{\boldsymbol{\eta}} = \underline{\hat{\boldsymbol{\eta}}}$ using state equation (27) recursively.
2. Evaluate $\hat{\delta}_t$, $\hat{A}_t$, and $\hat{B}_t$ at $\underline{\boldsymbol{\alpha}} = \underline{\hat{\boldsymbol{\alpha}}}$, $t = r+1, \ldots, r+d$.
3. Let $\hat{D}_{r+1} = \hat{A}_{r+1}$ and $\hat{b}_{r+1} = \hat{\delta}_{r+1}$. Compute the following variables recursively for $t = r+2, \ldots, r+d$:
$$\begin{aligned} \hat{D}_t &= \hat{A}_t - \hat{D}_{t-1}^{-1}\hat{B}_t^2, \quad \hat{K}_t = \sqrt{\hat{D}_t}, \\ \hat{b}_t &= \hat{\delta}_t - \hat{B}_t \hat{D}_{t-1}^{-1}\hat{b}_{t-1}. \end{aligned}$$
4. Define auxiliary variable $\hat{y}_t = \hat{\varphi}_t + \hat{D}_t^{-1}\hat{b}_t$, where $\hat{\varphi}_t = \hat{\alpha}_t + \hat{D}_t^{-1}\hat{B}_{t+1}\hat{\alpha}_{t+1}$, $t = r+1, \ldots, r+d$, and set $\hat{B}_{d+r+1} = 0$, $\hat{\alpha}_{r+d+1} = \alpha_{r+d+1}$.
5. Consider the linear Gaussian state space model formulated by
$$\begin{aligned} \hat{y}_t &= Z_t\alpha_t + G_t \mathrm{e}_t, && t = r+1, \ldots, r+d, && (30) \\ \alpha_{t+1} &= \phi_{s_{t+1}}\alpha_t + H_{t+1}\mathrm{u}_{t+1}, && t = r, \ldots, r+d-1, \quad t > 0, \\ \alpha_1 &= \mathcal{N}(0, \sigma_{s_1}/\sqrt{1-\phi_{s_1}}), && t = 0, \\ \begin{pmatrix} \mathrm{e}_t \\ \mathrm{u}_{t+1} \end{pmatrix} &\backsim \mathcal{N}(\mathbf{0}, \boldsymbol{I}_2), && && (31) \end{aligned}$$
where
$$Z_t = 1 + \phi_{s_{t+1}}\hat{D}_t^{-1}\hat{B}_{t+1}, \quad G_t = (\hat{K}_t^{-1}, \hat{D}_t^{-1}\hat{B}_{t+1}\sigma_{s_{t+1}}), \quad H_{t+1} = (0, \sigma_{s_{t+1}}).$$

Apply the Kalman filter (e.g. Durbin and Koopman, 2008) and disturbance smoother (Koopman, 1993) to this state space model to obtain the posterior mode $\hat{\underline{\boldsymbol{\eta}}}$ or $\hat{\underline{\boldsymbol{\alpha}}}$, equivalently.

6. Go to 2.

In the McMC sampling procedure, the current sample of $\underline{\boldsymbol{\eta}}$ may be taken as initial value of $\hat{\underline{\boldsymbol{\eta}}}$ in Step 1. Further note that the above steps are equivalent to the method of scoring used to maximize the conditional posterior density. After convergence of Algorithm 1, $\underline{\boldsymbol{\eta}}$ is sampled from the conditional posterior density by conducting an AR (Accept - Reject) - MH algorithm (Tierney, 1994) using the simulation smoother (de Jong and Shephard, 1995, or Durbin and Koopman, 2002).

Algorithm 2 (AR-MH step and simulation smoother):

1. Let $\underline{\boldsymbol{\eta}}_0$ denote the current value. Find mode $\hat{\underline{\boldsymbol{\eta}}}$ using Algorithm 1.
2. Proceed with Steps 2-4 in Algorithm 1 to obtain the approximated linear Gaussian state space model of Eq. (30)-(31).
3. Propose a candidate $\underline{\boldsymbol{\eta}}^*$ by sampling from $\tilde{q}(\underline{\boldsymbol{\eta}}^*) \propto \min\{\pi(\underline{\boldsymbol{\eta}}^*|\cdot), cq(\underline{\boldsymbol{\eta}}^*|\cdot)\}$ using the AR algorithm as follows:
   (a) Generate $\underline{\boldsymbol{\eta}}^*$ using the simulation smoother for the approximated state space model of Eq. (30)-(31).
   (b) Accept $\underline{\boldsymbol{\eta}}^*$ with probability
   $$\frac{\min\{\pi(\underline{\boldsymbol{\eta}}^*|\cdot), cq(\underline{\boldsymbol{\eta}}^*|\cdot)\}}{cq(\underline{\boldsymbol{\eta}}^*|\cdot)},$$
   where c is a scaling constant. If it is rejected, go to (a).
4. Conduct the MH algorithm using candidate $\underline{\boldsymbol{\eta}}^*$, with acceptance probability
$$\min\left\{\frac{\pi(\underline{\boldsymbol{\eta}}^*|\cdot) \times \min\{\pi(\underline{\boldsymbol{\eta}}_0|\cdot), cq(\underline{\boldsymbol{\eta}}_0|\cdot)\}}{\pi(\underline{\boldsymbol{\eta}}_0|\cdot) \times \min\{\pi(\underline{\boldsymbol{\eta}}^*|\cdot), cq(\underline{\boldsymbol{\eta}}^*|\cdot)\}}, 1\right\}.$$

B The $\mathcal{GH}$ Skew Student-t Distribution

This appendix includes some important properties of the generalized hyperbolic skew Student-t distribution ($\mathcal{GH}$ skew-t). For a more complete treatment, see Aas and Haff (2006). Consider further Scott, Würtz, Dong, and Tran (2011), Hu (2009), and Prause (1999). The $\mathcal{GH}$ skew-t distribution is a limiting case of the $\mathcal{GH}$ distribution, which was introduced in Barndorff-Nielsen (1977). Starting with the latter, a possible parameterization is as follows,

$$f_{\mathcal{GH}}(x) = \frac{(\beta^2-\gamma^2)^{\lambda/2} K_{\lambda-1/2}\left(\beta\sqrt{\delta^2+(x-\mu)^2}\right)\exp(\gamma(x-\mu))}{\sqrt{2\pi}\beta^{\lambda-1/2}\delta^{\lambda}K_{\lambda}\left(\delta\sqrt{\beta^2-\gamma^2}\right)(\delta+(x-\mu)^2)^{1/2-\lambda}}, \tag{32}$$

where K_j is the modified Bessel function of the third kind of order j (Abramowitz and Stegun, 1972). Parameters must fulfill the conditions

$$\begin{aligned} \delta \geq 0,\ |\gamma| < \beta \quad & \text{if } \lambda > 0, \\ \delta > 0,\ |\gamma| < \beta \quad & \text{if } \lambda = 0, \\ \delta > 0,\ |\gamma| \leq \beta \quad & \text{if } \lambda < 0. \end{aligned}$$

The tails of the generalized hyperbolic ($\mathcal{GH}$) distribution behave as

$$f_{\mathcal{GH}}(x) \backsim \text{const}|x|^{\lambda-1}\exp(-\beta|x|+\gamma x) \quad \text{as } x \to \pm\infty\ \forall\ \lambda. \tag{33}$$

It follows that as long as $|\gamma| \neq \beta$, both tails are semi-heavy.

The $\mathcal{GH}$ distribution can be represented as a normal mean-variance mixture with generalized inverse Gaussian ($\mathcal{GIG}$) distributed mixture variable (Barndorff-Nielsen and Blæsid, 1981),

$$X = \mu + \gamma Z + \sqrt{Z}Y, \quad Y \backsim \mathcal{N}(0,1), \quad Z \backsim \mathcal{GIG}(\lambda, \delta, \psi), \tag{34}$$

and $\psi = \sqrt{\beta^2-\gamma^2}$. It follows from Eq. (34) that $X|Z = z \backsim \mathcal{N}(\mu+\gamma z, z)$. However, parameters are difficult to estimate due to flatness of the likelihood function (see e.g. Aas and Haff, 2006). A parsimonious representation that is more amenable to estimation can be obtained by setting $\lambda = -\nu/2$ $(\nu > 0)$, $\delta = \sqrt{\nu}$ and $\beta \to |\gamma|$ in Eq. (32). Then $Z \backsim \mathcal{GIG}(-\nu/2, \sqrt{\nu}, 0)$ or equivalently $Z \backsim \mathcal{IG}(\nu/2, \nu/2)$. This limiting case of the $\mathcal{GH}$ distribution is referred to as the $\mathcal{GH}$ skew-t distribution (e.g. Barndorff-Nielsen and Shephard, 2001b), with skew parameter γ, degrees of freedom parameter ν and mixture representation

$$X = \mu + \gamma Z + \sqrt{Z}Y, \quad Y \backsim \mathcal{N}(0,1), \quad Z \backsim \mathcal{IG}(\nu/2, \nu/2).$$

Its closed form density is given by

$$f_{\mathcal{GH}\text{skew}t}(x) = \frac{2^{\frac{1-\nu}{2}}\nu^{\frac{\nu}{2}}|\gamma|^{\frac{\nu+1}{2}} K_{\frac{\nu+1}{2}}\left(\sqrt{\gamma^2(\nu+(x-\mu)^2)}\right)\exp(\gamma(x-\mu))}{\Gamma\left(\frac{\nu}{2}\right)\sqrt{\pi}\left(\sqrt{\nu+(x-\mu)^2}\right)^{\frac{\nu+1}{2}}}, \quad \gamma \neq 0,$$

and

$$f_{\mathcal{GH}t}(x) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\pi\nu}\Gamma\left(\frac{\nu}{2}\right)}\left[1+\frac{(x-\mu)^2}{\nu}\right]^{-(\nu+1)/2}, \qquad \gamma = 0.$$

The latter density is commonly referred to as the non-central Student-t distribution with ν degrees of freedom, expectation μ, and variance $\nu/(\nu-2)$.

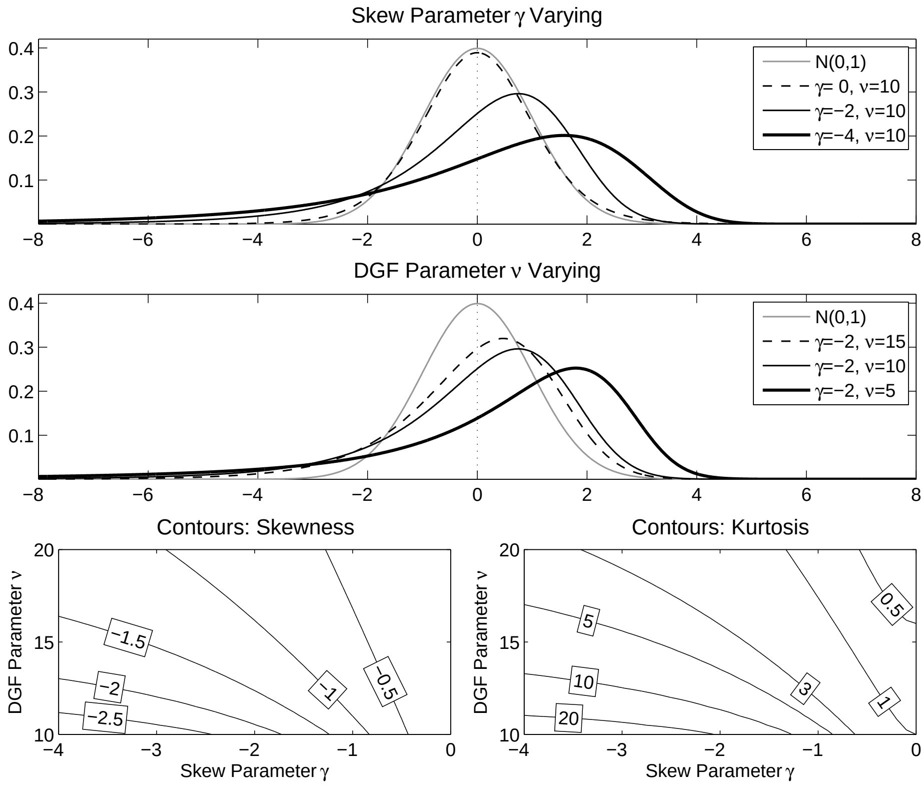

Figure 13: The $\mathcal{GH}$ skew-t distribution: Tail Behavior and Identification

The first four moments of a $\mathcal{GH}$ skew-t distributed random variate X are (Aas and Haff, 2006)

$$
\begin{aligned}
\mathrm{E}(X) &= \mu + \frac{\gamma\nu}{\nu-2}, \\
\mathrm{var}(X) &= \frac{2\gamma^2\nu^2}{(\nu-2)^2(\nu-4)} + \frac{\nu}{\nu-2}, \\
\mathrm{s}(X) &= \frac{2\gamma[\nu(\nu-4)]^{1/2}}{[2\gamma^2\nu + (\nu-2)(\nu-4)]^{3/2}}\left[3(\nu-2) + \frac{8\gamma^2\nu}{\nu-6}\right], \\
\mathrm{k(X)} &= \frac{6}{[2\gamma^2\nu + (\nu-2)(\nu-4)]^2} \times \\
&\quad \left[(\nu-2)^2(\nu-4) + \frac{16\gamma^2\nu(\nu-2)(\nu-4)}{\nu-6} + \frac{8\gamma^4\nu^2(5\nu-22)}{(\nu-6)(\nu-8)}\right].
\end{aligned}
$$

We observe that for the mean and variance to exist, $\nu > 2$ and $\nu > 4$, respectively. Skewness and excess kurtosis are defined only if $\nu > 6$ and $\nu > 8$, respectively.

From Eq. (33) it follows that the tails of the $\mathcal{GH}$ skew-t distribution are given by

$$f_{\mathcal{GH}\mathrm{skew}t}(x) \backsim \mathrm{const}|x|^{-\nu/2-1}\exp(-|\gamma x| + \gamma x) \quad \text{as } x \to \pm\infty.$$

Thus we have a heavy tail decaying as

$$f_{\mathcal{GH}\text{skew}t}(x) \backsim \text{const}|x|^{-\nu/2-1} \quad \text{if } \begin{cases} \gamma < 0 \text{ and } x \to -\infty \\ \gamma > 0 \text{ and } x \to +\infty, \end{cases}$$

and a light tail decaying as

$$f_{\mathcal{GH}\text{skew}t}(x) \backsim \text{const}|x|^{-\nu/2-1} \exp(-2|\gamma x|) \quad \text{if } \begin{cases} \gamma < 0 \text{ and } x \to +\infty \\ \gamma > 0 \text{ and } x \to -\infty. \end{cases}$$

The heavy tail shows polynomial and the light tail exponential behavior. Thus the $\mathcal{GH}$ skew-t distribution is able to model substantially skewed and heavy tailed data, as found in e.g. financial markets. This feature distinguishes it well from alternative skew Student-t distributions in the literature (for a more detailed discussion, see Aas and Haff, 2006). The tails of the $\mathcal{GH}$ skew-t distribution are characterized uniquely by parameters γ and ν, which determine jointly the degree of skewness and heavy tailedness. Fig. 13 demonstrates how both parameters shift probability mass into the tails of the distribution (upper two plots). Observe that a specific skewness and kurtosis of the distribution can be obtained by different combinations of higher moment parameters γ and ν (lower two plots).

Finally, note that the heavy tail of the $\mathcal{GH}$ skew-t distribution is heavier than the tails of the symmetric Student-t distribution, which decay as

$$f_{\mathcal{GH}t}(x) \backsim \text{const}|x|^{-\nu-1} \quad \text{as } x \to \pm\infty.$$

C Inefficiency Factor / Convergence Diagnostic

The inefficiency factor is the estimated ratio of the numerical variance of the posterior sample mean to the variance of the hypothetical sample mean from uncorrelated draws. With an inefficiency factor of m, we need to draw McMC samples m times as many as uncorrelated samples. The following formula is used to calculate the inefficiency factor,

$$IF = 1 + 2\frac{b_w}{b_w - 1}\sum_{i=1}^{b_w} K_{\text{QS}}\left(\frac{i}{b_w}\right)\hat{\rho}_i, \tag{35}$$

where $\hat{\rho}_i$ is the sample autocorrelation at lag i, K_{QS} the Quadratic Spectral (QS) kernel and b_w the bandwidth parameter (the correction term $b_w/(b_w - 1)$ dropped if $b_w = 1$). Andrews (1991) finds the QS kernel to be superior in terms of an asymptotic truncated mean squared error criterion, relative to other kernels. The QS kernel is defined as

$$K_{\text{QS}}(x) = \frac{25}{12\pi^2x^2}\left(\frac{\sin(6\pi x/5)}{6\pi x/5} - \cos(6\pi x/5)\right).$$

The bandwidth b_w is determined automatically as a function of the data, following Andrews (1991). Specifically,

$$\hat{b}_w = 1.3221(\hat{\alpha}(2)N)^{1/5},$$

where N is the number of collected draws and

$$\hat{\alpha}(2) = \frac{4\hat{\rho}_a^2\hat{\sigma}_a^4}{(1-\hat{\rho}_a)^8} \bigg/ \left[\frac{\hat{\sigma}_a^4}{(1-\hat{\rho}_a)^4}\right],$$

with $\{\hat{\rho}_a,\, \hat{\sigma}_a\}$ the autoregressive and variance innovation parameter, respectively, obtained by running a first-order autoregressive linear regression on the demeaned draws of the Markov chain. As Andrews (1991) notes, there can be sensitivity of the estimator to the choice of the bandwidth parameter. In the current context, this would suggest obtaining multiple bandwidth values by varying $\hat{\rho}_a$ by one or two standard deviations, $\hat{\rho}_a \pm 1/\sqrt{N}$ or $\hat{\rho}_a \pm 2/\sqrt{N}$, respectively.

Geweke (1992) proposes the widely used convergence diagnostic (CD) to test if a Markov chain actually has converged. The basic idea is to compare values sampled early in the sequence (subsample A) with those late in the sequence (subsample B). Denote by $\theta^{(i)}$ the i^{th} draw of parameter θ in the recorded N draws, and calculate $\bar{\theta}_A = \frac{1}{n_A}\sum_{i=1}^{n_A}\theta^{(i)}$, $\bar{\theta}_B = \frac{1}{n_B}\sum_{i=N-n_B+1}^{N}\theta^{(i)}$. Further, let σ_A and σ_B denote the standard errors of $\bar{\theta}_A$ and $\bar{\theta}_B$, respectively. The convergence diagnostic is then defined as

$$CD = \frac{\bar{\theta}_A - \bar{\theta}_B}{\sqrt{\sigma_A^2 + \sigma_B^2}} \xrightarrow{d} \mathcal{N}(0,1),$$

under the assumption that sequence $\theta^{(i)}$ is stationary. For window lengths n_A and n_B Geweke (1992) proposes $0.1N$ and $0.5N$, respectively, which have become standard in the literature and are also employed in this work. With a corresponding estimate of the inefficiency factor in Eq. (35) at hand, variances σ_A^2, σ_B^2 are calculated as $\sigma_A^2 = s_A^2(b_{w,A}+1)^{-1}IF_A$, $\sigma_B^2 = s_B^2(b_{w,B}+1)^{-1}IF_B$, where s_A^2 and s_B^2 are the variances of subsamples A and B, respectively.

D Parameter Estimates

This appendix contains estimation statistics of the transition probabilities of the presented models. The results for estimated nested models (i.e. models with Gaussian or Student-t error) are also reported. All estimates are based on the full sample 1990/01/02 - 2010/08/06. If not stated otherwise, priors and McMC parameters are those in the main text.

SVsk$t$ Model Variants

Table 17: Estimation Results of the SVt Model

	Mean	Stdev	95% c.b.	IF	CD
μ	-0.2822	0.1173	[-0.5101, -0.0470]	1.6	0.96
ϕ	0.9871	0.0022	[0.9825, 0.9910]	12.0	0.97
σ	0.1484	0.0106	[0.1292, 0.1696]	186.4	0.99
ρ	-0.7294	0.0356	[-0.7933, -0.6580]	94.8	0.97
ν	17.901	3.2304	[12.430, 24.700]	254.7	0.71
kurtosis	0.4562	0.1112	[0.2898, 0.7116]	265.4	0.74

Posterior mean, standard deviation, 95% credible bounds, inefficiency factor, and Geweke's convergence diagnostic (p-value). Lower panel: Implied higher moments of ϵ_t.

Table 18: Estimation Results of the SV Model

	Mean	Stdev	95% c.b.	IF	CD
μ	-0.2026	0.1143	[-0.4283, 0.0242]	1.2	0.98
ϕ	0.9850	0.0025	[0.9798, 0.9896]	12.5	0.87
σ	0.1590	0.0113	[0.1384, 0.1833]	148.3	0.82
ρ	-0.6845	0.0360	[-0.7499, -0.6116]	76.6	0.70

Posterior mean, standard deviation, 95% credible bounds, inefficiency factor, and Geweke's convergence diagnostic (p-value).

MS3SVsk$t$ Model Variants

Table 19: Estimation Results of the MS3SVsk$t$ Model: Transition Probabilities

	Mean	Stdev	95% c.b.	IF	CD
p_{11}	0.9766	0.0112	[0.9502, 0.9933]	3.0	0.97
p_{12}	0.0155	0.0093	[0.0028, 0.0378]	3.1	0.91
p_{13}	0.0079	0.0065	[0.0006, 0.0250]	2.9	0.92
p_{21}	0.0014	0.0010	[0.0002, 0.0038]	6.3	0.98
p_{22}	0.9970	0.0013	[0.9939, 0.9989]	4.6	0.97
p_{23}	0.0016	0.0009	[0.0004, 0.0037]	2.7	0.98
p_{31}	0.0009	0.0007	[0.0001, 0.0027]	3.9	0.80
p_{32}	0.0020	0.0011	[0.0005, 0.0046]	2.6	0.98
p_{33}	0.9971	0.0012	[0.9943, 0.9990]	2.4	0.87

Posterior mean, standard deviation, 95% credible bounds, inefficiency factor, and Geweke's convergence diagnostic (p-value).

Table 20: Estimation Results of the MS3SVt Model

	Mean	Stdev	95% c.b.	IF	CD
μ_1	1.5364	0.2725	[0.9521, 2.0312]	67.9	0.59
μ_2	0.1013	0.0776	[-0.0602, 0.2449]	21.1	0.66
μ_3	-1.0333	0.0812	[-1.2010, -0.8777]	16.0	0.78
ϕ	0.9563	0.0078	[0.9407, 0.9717]	59.0	0.73
σ	0.1768	0.0129	[0.1520, 0.2037]	210.0	0.84
ρ	-0.7756	0.0324	[-0.8364, -0.7097]	103.9	0.88
ν	18.480	3.1091	[13.122, 25.668]	234.1	0.80
kurtosis	0.4338	0.0945	[0.2769, 0.6576]	238.0	0.85
p_{11}	0.9745	0.0135	[0.9428, 0.9928]	7.1	0.70
p_{12}	0.0168	0.0106	[0.0030, 0.0425]	5.6	0.68
p_{13}	0.0087	0.0077	[0.0006, 0.0283]	5.5	0.88
p_{21}	0.0015	0.0011	[0.0002, 0.0042]	10.5	0.72
p_{22}	0.9970	0.0013	[0.9938, 0.9989]	6.2	0.80
p_{23}	0.0015	0.0009	[0.0003, 0.0036]	3.1	0.95
p_{31}	0.0010	0.0009	[0.0001, 0.0033]	5.1	0.63
p_{32}	0.0020	0.0011	[0.0004, 0.0046]	3.5	0.77
p_{33}	0.9970	0.0013	[0.9939, 0.9990]	3.4	0.57

Posterior mean, standard deviation, 95% credible bounds, inefficiency factor, and Geweke's convergence diagnostic (p-value). Middle panel: Implied higher moments of ϵ_t. Bottom panel: Transition probabilities.

SVsk$t$-RV Model Variants

Table 21: Results of the SVt-RV (upper block)/SV-RV Model (lower block)

	Mean	Stdev	95% c.b.	IF	CD
μ	-0.2589	0.0931	[-0.4378, -0.0749]	1.5	1.00
ϕ	0.9690	0.0035	[0.9620, 0.9755]	5.2	0.86
σ	0.2218	0.0096	[0.2039, 0.2414]	54.0	0.76
ρ	-0.4552	0.0291	[-0.5112, -0.3970]	23.1	0.84
ν	28.581	4.5704	[20.862, 38.382]	202.6	0.92
kurtosis	0.2526	0.0472	[0.1745, 0.3557]	202.5	0.92
ζ	-0.2224	0.0229	[-0.2679, -0.1783]	42.9	0.98
ξ	0.9047	0.0204	[0.8665, 0.9456]	35.0	0.82
σ_u	0.4072	0.0059	[0.3955, 0.4188]	8.5	0.87
SNR	0.5449	0.0281	[0.4933, 0.6029]	50.1	0.78
μ	-0.1971	0.0918	[-0.3760, -0.0143]	1.3	0.98
ϕ	0.9683	0.0035	[0.9612, 0.9749]	4.9	0.99
σ	0.2252	0.0095	[0.2066, 0.2440]	51.0	0.97
ρ	-0.4422	0.0283	[-0.4982, -0.3874]	19.1	0.94
ζ	-0.2814	0.0191	[-0.3195, -0.2448]	24.4	0.95
ξ	0.9029	0.0197	[0.8657, 0.9427]	33.0	1.00
σ_u	0.4056	0.0058	[0.3942, 0.4169]	7.6	0.97
SNR	0.5556	0.0280	[0.5023, 0.6118]	45.2	0.97

Posterior mean, standard deviation, 95% credible bounds, inefficiency factor, and convergence diagnostic (p-value). Second panel (SVt-RV only): Implied higher moments of ϵ_t. Lower two panels: Parameters auxiliary measurement eq. and implied signal-to-noise ratio.

MS4SVsk$t$-RV Model Variants

Table 22: Results of the MS4SVsk$t$-RV Model: Transition Probabilities

	Mean	Stdev	95% c.b.	IF	CD
p_{11}	0.3512	0.1030	[0.1443, 0.5492]	35.6	0.94
p_{12}	0.3118	0.1232	[0.0920, 0.5637]	42.0	0.84
p_{13}	0.3218	0.0986	[0.1304, 0.5139]	28.4	0.72
p_{14}	0.0153	0.0148	[0.0004, 0.0545]	5.5	0.85
p_{21}	0.0318	0.0210	[0.0039, 0.0846]	41.6	0.88
p_{22}	0.5479	0.0713	[0.4104, 0.6922]	82.6	0.54
p_{23}	0.4135	0.0676	[0.2757, 0.5415]	77.7	0.52
p_{24}	0.0068	0.0059	[0.0002, 0.0221]	15.0	0.80
p_{31}	0.0198	0.0054	[0.0106, 0.0318]	27.9	0.98
p_{32}	0.0845	0.0202	[0.0475, 0.1260]	87.7	0.95
p_{33}	0.8747	0.0211	[0.8287, 0.9117]	82.4	0.86
p_{34}	0.0210	0.0050	[0.0127, 0.0321]	28.3	0.61
p_{41}	0.0448	0.0368	[0.0018, 0.1391]	12.3	0.91
p_{42}	0.1424	0.0761	[0.0165, 0.3067]	20.9	0.64
p_{43}	0.5833	0.0924	[0.3963, 0.7555]	18.2	0.62
p_{44}	0.2294	0.0642	[0.1130, 0.3636]	12.2	0.88

Posterior mean, standard deviation, 95% credible bounds, inefficiency factor, and Geweke's convergence diagnostic (p-value).

Table 23: Estimation Results of the MS4SVt-RV Model

	Mean	Stdev	95% c.b.	IF	CD
μ_1	0.8792	0.1833	[0.5301, 1.2430]	29.4	0.63
μ_2	0.2381	0.1414	[-0.0347, 0.5214]	12.8	0.77
μ_3	-0.4185	0.1273	[-0.6619, -0.1612]	8.2	0.80
μ_4	-1.5878	0.1556	[-1.8926, -1.2798]	20.5	0.73
ϕ_1	0.9346	0.0373	[0.8502, 0.9911]	16.1	0.92
ϕ_2	0.9836	0.0104	[0.9590, 0.9982]	14.5	1.00
ϕ_3	0.9863	0.0032	[0.9800, 0.9927]	13.1	0.94
ϕ_4	0.9072	0.0433	[0.8134, 0.9812]	44.0	0.81
σ_1	0.2730	0.0622	[0.1534, 0.4002]	66.6	0.74
σ_2	0.2267	0.0312	[0.1644, 0.2903]	121.1	0.83
σ_3	0.1286	0.0113	[0.1064, 0.1516]	195.4	0.52
σ_4	0.1450	0.0428	[0.0788, 0.2425]	67.1	0.67
ρ	-0.5167	0.0321	[-0.5789, -0.4554]	37.9	0.98
ν	33.329	5.0549	[24.455, 44.635]	159.3	0.99
kurtosis	0.2107	0.0365	[0.1476, 0.2933]	157.4	0.93
ζ	-0.2124	0.0215	[-0.2541, -0.1689]	55.6	0.89
ξ	0.8913	0.0198	[0.8539, 0.9311]	61.5	0.90
$\sigma_{u,1}$	0.6528	0.0654	[0.5318, 0.7911]	12.2	0.84
$\sigma_{u,2}$	0.3206	0.0308	[0.2612, 0.3829]	57.5	0.78
$\sigma_{u,3}$	0.3034	0.0092	[0.2829, 0.3199]	40.2	0.99
$\sigma_{u,4}$	0.3511	0.0473	[0.2620, 0.4478]	15.4	0.86
SNR_1	0.4236	0.1113	[0.2308, 0.6611]	47.7	0.86
SNR_2	0.7155	0.1340	[0.4939, 1.0219]	99.1	0.99
SNR_3	0.4243	0.0406	[0.3475, 0.5081]	146.7	0.61
SNR_4	0.4216	0.1413	[0.2131, 0.7540]	48.5	0.81
p_{11}	0.3476	0.1183	[0.1234, 0.5874]	51.5	0.98
p_{12}	0.3225	0.1317	[0.0895, 0.6000]	51.0	0.82
p_{13}	0.3145	0.1044	[0.1182, 0.5207]	34.3	0.82
p_{14}	0.0153	0.0153	[0.0004, 0.0566]	6.0	0.86
p_{21}	0.0338	0.0238	[0.0032, 0.0904]	55.1	0.63
p_{22}	0.5641	0.0743	[0.4194, 0.7031]	80.6	0.70
p_{23}	0.3956	0.0720	[0.2596, 0.5378]	86.7	0.86
p_{24}	0.0065	0.0059	[0.0002, 0.0219]	14.6	0.85
p_{31}	0.0195	0.0056	[0.0097, 0.0318]	34.8	0.65
p_{32}	0.0762	0.0215	[0.0410, 0.1273]	116.2	0.73
p_{33}	0.8847	0.0226	[0.8264, 0.9198]	104.9	0.85
p_{34}	0.0196	0.0047	[0.0116, 0.0298]	27.6	0.91
p_{41}	0.0450	0.0372	[0.0017, 0.1392]	10.0	0.96
p_{42}	0.1317	0.0770	[0.0106, 0.3025]	21.6	0.98
p_{43}	0.5935	0.0933	[0.4042, 0.7648]	18.6	0.96
p_{44}	0.2298	0.0646	[0.1136, 0.3648]	11.9	0.97

Posterior mean, standard deviation, 95% credible bounds, inefficiency factor, and Geweke's convergence diagnostic (p-value). Second panel: Implied higher moments of ϵ_t. Third panel: Parameters auxiliary measurement eq. Fourth panel: Implied signal-to-noise ratios. Bottom panel: Transition probabilities.

Table 24: Estimation Results of the MS4SV-RV Model

	Mean	Stdev	95% c.b.	IF	CD
μ_1	0.8959	0.1654	[0.5829, 1.2283]	23.9	0.97
μ_2	0.2823	0.1393	[0.0148, 0.5652]	13.1	0.96
μ_3	-0.3698	0.1259	[-0.6117, -0.1110]	8.7	0.93
μ_4	-1.5391	0.1526	[-1.8402, -1.2393]	20.6	0.94
ϕ_1	0.9336	0.0378	[0.8478, 0.9910]	18.1	0.92
ϕ_2	0.9841	0.0102	[0.9596, 0.9981]	12.7	0.97
ϕ_3	0.9864	0.0032	[0.9800, 0.9926]	12.8	0.93
ϕ_4	0.9034	0.0426	[0.8171, 0.9804]	45.3	0.89
σ_1	0.2684	0.0611	[0.1526, 0.3947]	66.7	0.52
σ_2	0.2279	0.0304	[0.1657, 0.2907]	117.9	0.69
σ_3	0.1265	0.0103	[0.1074, 0.1473]	180.2	0.98
σ_4	0.1458	0.0423	[0.0805, 0.2435]	59.0	0.94
ρ	-0.5146	0.0324	[-0.5769, -0.4504]	39.5	0.96
ζ	-0.2568	0.0197	[-0.2954, -0.2181]	41.9	0.88
ξ	0.8952	0.0192	[0.8595, 0.9347]	52.1	0.70
$\sigma_{u,1}$	0.6539	0.0643	[0.5410, 0.7935]	11.8	0.97
$\sigma_{u,2}$	0.3183	0.0301	[0.2573, 0.3752]	56.8	0.86
$\sigma_{u,3}$	0.3040	0.0083	[0.2879, 0.3199]	29.1	0.99
$\sigma_{u,4}$	0.3494	0.0458	[0.2657, 0.4445]	14.1	0.99
SNR_1	0.4151	0.1057	[0.2234, 0.6376]	48.1	0.64
SNR_2	0.7255	0.1368	[0.4839, 1.0255]	104.4	0.73
SNR_3	0.4165	0.0367	[0.3494, 0.4904]	131.9	0.98
SNR_4	0.4246	0.1378	[0.2209, 0.7539]	43.7	0.95
p_{11}	0.3307	0.1106	[0.1147, 0.5403]	41.7	0.96
p_{12}	0.3350	0.1298	[0.1019, 0.6014]	44.7	0.96
p_{13}	0.3197	0.0982	[0.1311, 0.5127]	29.0	0.99
p_{14}	0.0147	0.0144	[0.0004, 0.0528]	6.2	0.98
p_{21}	0.0378	0.0277	[0.0037, 0.1118]	57.8	0.90
p_{22}	0.5640	0.0762	[0.4078, 0.7071]	86.8	0.82
p_{23}	0.3920	0.0701	[0.2556, 0.5259]	80.0	0.83
p_{24}	0.0063	0.0054	[0.0002, 0.0202]	12.3	0.99
p_{31}	0.0207	0.0057	[0.0108, 0.0333]	31.6	0.97
p_{32}	0.0744	0.0176	[0.0422, 0.1115]	75.4	0.86
p_{33}	0.8854	0.0181	[0.8463, 0.9171]	66.9	0.88
p_{34}	0.0194	0.0043	[0.0120, 0.0288]	22.2	0.95
p_{41}	0.0483	0.0393	[0.0017, 0.1473]	12.1	0.87
p_{42}	0.1316	0.0765	[0.0129, 0.3037]	21.9	0.87
p_{43}	0.5892	0.0921	[0.4022, 0.7591]	17.8	0.87
p_{44}	0.2309	0.0653	[0.1122, 0.3684]	12.7	0.91

Posterior mean, standard deviation, 95% credible bounds, inefficiency factor, and Geweke's convergence diagnostic (p-value). Second panel: Parameters auxiliary measurement eq. Third panel: Implied signal-to-noise ratios. Bottom panel: Transition probabilities.

SVsk$t$-Range Model Variants

Table 25: Estimation Results of the SVt-Range Model

	Mean	Stdev	95% c.b.	IF	CD
μ	-0.3344	0.0871	[-0.5031, -0.1618]	1.4	0.96
ϕ	0.9601	0.0060	[0.9475, 0.9705]	25.6	0.95
σ	0.2838	0.0222	[0.2451, 0.3291]	164.4	0.97
ρ	-0.5378	0.0347	[-0.6054, -0.4702]	57.1	0.88
ν	30.973	5.0957	[21.310, 41.563]	218.0	0.81
kurtosis	0.2309	0.0466	[0.1597, 0.3464]	247.2	0.84
ζ	0.2865	0.0120	[0.2631, 0.3102]	15.6	0.97
ξ	0.4522	0.0112	[0.4310, 0.4745]	20.8	0.95
σ_u	0.3673	0.0051	[0.3574, 0.3771]	8.9	0.92
SNR	0.7732	0.0675	[0.6573, 0.9127]	148.5	0.97

Posterior mean, standard deviation, 95% credible bounds, inefficiency factor, and Geweke's convergence diagnostic (p-value). Second panel: Implied higher moments of ϵ_t. Third panel: Parameters auxiliary measurement eq. Lower panel: Implied signal-to-noise ratio.

MS4SVsk$t$-Range Model Variants

Table 26: Results of the MS4SVsk$t$-Range Model: Transition Probabilities

	Mean	Stdev	95% c.b.	IF	CD
p_{11}	0.0952	0.0218	[0.0552, 0.1413]	16.7	0.91
p_{12}	0.2412	0.0374	[0.1672, 0.3138]	26.4	1.00
p_{13}	0.3985	0.0490	[0.3016, 0.4919]	35.4	0.83
p_{14}	0.2651	0.0456	[0.1833, 0.3598]	43.4	0.86
p_{21}	0.0835	0.0195	[0.0495, 0.1252]	54.7	0.90
p_{22}	0.2908	0.0246	[0.2407, 0.3372]	28.8	0.97
p_{23}	0.3937	0.0320	[0.3296, 0.4538]	44.1	0.76
p_{24}	0.2320	0.0371	[0.1665, 0.3132]	87.5	0.88
p_{31}	0.1485	0.0228	[0.1068, 0.1944]	47.6	0.82
p_{32}	0.3621	0.0294	[0.3016, 0.4180]	41.1	0.94
p_{33}	0.3003	0.0294	[0.2396, 0.3571]	42.9	0.79
p_{34}	0.1892	0.0309	[0.1359, 0.2568]	76.5	0.87
p_{41}	0.1552	0.0264	[0.1073, 0.2103]	32.8	0.84
p_{42}	0.3671	0.0335	[0.3011, 0.4336]	27.3	0.87
p_{43}	0.3527	0.0321	[0.2906, 0.4154]	24.6	0.92
p_{44}	0.1250	0.0306	[0.0693, 0.1878]	55.2	0.92

Posterior mean, standard deviation, 95% credible bounds, inefficiency factor, and Geweke's convergence diagnostic (p-value).

Table 27: Estimation Results of the MS4SVt-Range Model

	Mean	Stdev	95% c.b.	IF	CD
μ_1	0.9816	0.1258	[0.7464, 1.2367]	7.8	0.85
μ_2	0.0287	0.1156	[-0.1921, 0.2640]	5.1	0.86
μ_3	-0.9221	0.1177	[-1.1469, -0.6887]	5.3	0.85
μ_4	-1.8424	0.1225	[-2.0758, -1.6004]	7.2	0.78
ϕ_1	0.9252	0.0260	[0.8731, 0.9742]	21.6	0.92
ϕ_2	0.9658	0.0117	[0.9417, 0.9875]	31.8	0.71
ϕ_3	0.9758	0.0114	[0.9516, 0.9958]	23.4	0.75
ϕ_4	0.9677	0.0179	[0.9284, 0.9958]	38.2	0.86
σ_1	0.4034	0.0403	[0.3247, 0.4849]	49.0	0.93
σ_2	0.2554	0.0235	[0.2093, 0.3020]	91.0	0.65
σ_3	0.3054	0.0204	[0.2647, 0.3444]	62.4	0.71
σ_4	0.2359	0.0290	[0.1795, 0.2932]	97.0	0.80
ρ_1	-0.3468	0.0764	[-0.4959, -0.1940]	18.5	0.92
ρ_2	-0.4463	0.0559	[-0.5527, -0.3350]	33.2	0.78
ρ_3	-0.6049	0.0415	[-0.6820, -0.5200]	33.9	0.96
ρ_4	-0.6339	0.0634	[-0.7476, -0.5008]	33.3	0.93
ν	53.860	6.1432	[42.640, 66.448]	72.0	0.98
kurtosis	0.1222	0.0152	[0.0961, 0.1553]	73.9	0.97
ζ	0.4220	0.0102	[0.4023, 0.4423]	62.3	0.93
ξ	0.4104	0.0065	[0.3974, 0.4231]	81.7	0.93
$\sigma_{u,1}$	0.1622	0.0123	[0.1376, 0.1862]	25.0	1.00
$\sigma_{u,2}$	0.1062	0.0075	[0.0920, 0.1212]	49.3	0.91
$\sigma_{u,3}$	0.1165	0.0086	[0.0992, 0.1331]	59.3	0.89
$\sigma_{u,4}$	0.1768	0.0101	[0.1574, 0.1968]	24.5	0.89
SNR_1	2.5081	0.3673	[1.8491, 3.3179]	44.9	0.95
SNR_2	2.4221	0.3168	[1.8485, 3.0787]	70.0	0.76
SNR_3	2.6402	0.2974	[2.1060, 3.2688]	66.6	0.90
SNR_4	1.3397	0.1909	[0.9843, 1.7299]	73.4	0.88
p_{11}	0.0912	0.0199	[0.0544, 0.1321]	13.1	0.91
p_{12}	0.2759	0.0369	[0.2033, 0.3477]	22.7	0.91
p_{13}	0.3999	0.0482	[0.3070, 0.4954]	33.7	0.90
p_{14}	0.2331	0.0419	[0.1569, 0.3192]	38.8	0.79
p_{21}	0.0866	0.0168	[0.0553, 0.1210]	41.7	0.77
p_{22}	0.3079	0.0234	[0.2619, 0.3530]	27.3	0.97
p_{23}	0.3983	0.0301	[0.3373, 0.4563]	42.2	0.87
p_{24}	0.2072	0.0316	[0.1510, 0.2711]	74.0	0.79
p_{31}	0.1454	0.0219	[0.1010, 0.1883]	45.5	0.72
p_{32}	0.3890	0.0278	[0.3335, 0.4412]	35.3	0.85
p_{33}	0.3015	0.0284	[0.2401, 0.3532]	40.4	0.75
p_{34}	0.1642	0.0280	[0.1156, 0.2218]	72.9	0.70
p_{41}	0.1495	0.0254	[0.1003, 0.2005]	27.1	0.80
p_{42}	0.3924	0.0322	[0.3297, 0.4573]	19.7	0.91
p_{43}	0.3487	0.0332	[0.2844, 0.4149]	23.4	0.89
p_{44}	0.1094	0.0266	[0.0604, 0.1636]	38.4	0.84

Posterior mean, standard deviation, 95% credible bounds, inefficiency factor, and Geweke's convergence diagnostic (p-value). Second panel: Implied higher moments of ϵ_t. Third panel: Parameters auxiliary measurement eq. Fourth panel: Implied signal-to-noise ratios. Bottom panel: Transition probabilities.

Table 28: Estimation Results of the MS4SV-Range Model

	Mean	Stdev	95% c.b.	IF	CD
μ_1	0.9967	0.1199	[0.7662, 1.2368]	6.2	0.84
μ_2	0.0458	0.1108	[-0.1664, 0.2677]	3.9	0.89
μ_3	-0.8959	0.1170	[-1.1202, -0.6601]	5.4	0.97
μ_4	-1.8118	0.1249	[-2.0585, -1.5671]	8.4	0.87
ϕ_1	0.9239	0.0260	[0.8718, 0.9731]	20.7	0.85
ϕ_2	0.9646	0.0122	[0.9402, 0.9873]	34.7	0.92
ϕ_3	0.9763	0.0111	[0.9527, 0.9956]	22.2	0.93
ϕ_4	0.9680	0.0174	[0.9300, 0.9958]	37.0	0.85
σ_1	0.4028	0.0389	[0.3287, 0.4787]	46.1	0.97
σ_2	0.2548	0.0235	[0.2084, 0.3005]	93.0	0.76
σ_3	0.3074	0.0213	[0.2663, 0.3502]	61.4	0.99
σ_4	0.2402	0.0281	[0.1850, 0.2944]	80.7	0.90
ρ_1	-0.3444	0.0730	[-0.4879, -0.1988]	16.0	0.89
ρ_2	-0.4469	0.0542	[-0.5502, -0.3391]	29.4	0.86
ρ_3	-0.5957	0.0427	[-0.6735, -0.5074]	33.5	0.88
ρ_4	-0.6230	0.0635	[-0.7339, -0.4832]	34.7	0.95
ζ	0.4165	0.0093	[0.3980, 0.4353]	49.0	0.90
ξ	0.4113	0.0062	[0.3990, 0.4231]	63.0	0.97
$\sigma_{u,1}$	0.1607	0.0120	[0.1373, 0.1847]	24.4	0.94
$\sigma_{u,2}$	0.1054	0.0073	[0.0917, 0.1201]	55.6	0.88
$\sigma_{u,3}$	0.1156	0.0088	[0.0989, 0.1334]	64.3	0.89
$\sigma_{u,4}$	0.1779	0.0098	[0.1584, 0.1969]	23.1	0.90
SNR_1	2.5268	0.3542	[1.8875, 3.2643]	40.1	0.98
SNR_2	2.4336	0.3270	[1.8367, 3.1218]	91.1	0.90
SNR_3	2.6792	0.3115	[2.1262, 3.3505]	62.4	0.92
SNR_4	1.3557	0.1827	[1.0078, 1.7351]	63.0	0.87
p_{11}	0.0902	0.0197	[0.0538, 0.1304]	12.8	0.84
p_{12}	0.2747	0.0359	[0.2028, 0.3453]	20.8	0.97
p_{13}	0.3949	0.0492	[0.2980, 0.4904]	34.1	0.77
p_{14}	0.2402	0.0467	[0.1539, 0.3343]	48.7	0.81
p_{21}	0.0841	0.0162	[0.0545, 0.1188]	38.5	0.81
p_{22}	0.3049	0.0234	[0.2587, 0.3497]	26.9	0.82
p_{23}	0.3962	0.0314	[0.3354, 0.4581]	45.9	0.95
p_{24}	0.2148	0.0373	[0.1452, 0.2910]	99.9	0.94
p_{31}	0.1437	0.0202	[0.1058, 0.1850]	35.6	0.82
p_{32}	0.3863	0.0283	[0.3298, 0.4408]	38.5	0.85
p_{33}	0.3004	0.0261	[0.2465, 0.3494]	31.5	0.82
p_{34}	0.1696	0.0308	[0.1135, 0.2319]	85.1	0.86
p_{41}	0.1475	0.0230	[0.1047, 0.1945]	21.1	0.80
p_{42}	0.3902	0.0324	[0.3297, 0.4562]	21.1	0.85
p_{43}	0.3475	0.0348	[0.2793, 0.4148]	28.0	0.86
p_{44}	0.1148	0.0288	[0.0606, 0.1718]	46.1	0.85

Posterior mean, standard deviation, 95% credible bounds, inefficiency factor, and Geweke's convergence diagnostic (p-value). Second panel: Parameters auxiliary measurement eq. Third panel: Implied signal-to-noise ratios. Bottom panel: Transition probabilities.

SVsk$t$-VIX Model Variants

Table 29: Estimation Results of the SVt-VIX Model

	Mean	Stdev	95% c.b.	IF	CD
μ	-0.3421	0.1741	[-0.6859, 0.0019]	1.2	0.77
ϕ	0.9885	0.0021	[0.9843, 0.9926]	1.8	0.93
σ	0.1383	0.0045	[0.1298, 0.1472]	108.8	0.88
ρ	-0.0199	0.0210	[-0.0622, 0.0204]	15.8	0.92
ν	15.512	2.7446	[11.263, 21.724]	492.5	0.53
kurtosis	0.5503	0.1284	[0.3385, 0.8261]	453.0	0.52
ζ	0.3170	0.0122	[0.2946, 0.3386]	3,214.6	0.17
ξ	0.3868	0.0082	[0.3719, 0.4029]	1,513.4	0.53
σ_u	0.0187	0.0019	[0.0147, 0.0223]	220.4	0.86
SNR	7.5075	0.9942	[5.9333, 9.844]	241.6	0.94

Posterior mean, standard deviation, 95% credible bounds, inefficiency factor, and Geweke's convergence diagnostic (p-value). Second panel: Implied higher moments of ϵ_t. Third panel: Parameters auxiliary measurement eq. Lower panel: Implied signal-to-noise ratio.

Table 30: Estimation Results of the SV-VIX Model

	Mean	Stdev	95% c.b.	IF	CD
μ	-0.2032	0.1729	[-0.5472, 0.1370]	1.1	0.98
ϕ	0.9885	0.0021	[0.9843, 0.9926]	1.8	0.93
σ	0.1392	0.0048	[0.1299, 0.1486]	126.0	0.91
ρ	-0.0121	0.0213	[-0.0547, 0.0285]	17.6	0.86
ζ	0.2604	0.0068	[0.2461, 0.2733]	1,071.4	0.70
ξ	0.3871	0.0085	[0.3726, 0.4046]	1,955.3	0.93
σ_u	0.0180	0.0022	[0.0133, 0.0219]	312.7	0.70
SNR	7.8803	1.2670	[6.0212, 10.969]	357.8	0.71

Posterior mean, standard deviation, 95% credible bounds, inefficiency factor, and Geweke's convergence diagnostic (p-value). Middle panel: Parameters auxiliary measurement eq. Lower panel: Implied signal-to-noise ratio.

MS2SVsk$t$-VIX Model Variants

Table 31: Results of the MS2SVsk$t$-VIX Model: Transition Probabilities

	Mean	Stdev	95% c.b.	IF	CD
p_{11}	0.2701	0.0356	[0.2027, 0.3421]	39.0	0.96
p_{12}	0.7299	0.0356	[0.6579, 0.7973]	39.0	0.96
p_{21}	0.2471	0.0235	[0.2031, 0.2944]	100.9	0.87
p_{22}	0.7529	0.0235	[0.7056, 0.7969]	100.9	0.87

Posterior mean, standard deviation, 95% credible bounds, inefficiency factor, and Geweke's convergence diagnostic (p-value).

Table 32: The MS2SVt-VIX (upper block)/MS2SV-VIX (lower block) Model

	Mean	Stdev	95% c.b.	IF	CD
μ	-0.8697	0.1444	[-1.1763, -0.6088]	4.1	0.94
ϕ	0.9869	0.0020	[0.9831, 0.9908]	5.5	0.78
σ_1	0.2532	0.0160	[0.2210, 0.2838]	103.1	0.91
σ_2	0.1006	0.0036	[0.0939, 0.1080]	140.4	0.76
ρ_1	-0.0055	0.0450	[-0.1040, 0.0725]	28.8	0.93
ρ_2	0.0238	0.0214	[-0.0191, 0.0650]	9.6	0.93
ν	15.8231	2.6897	[11.4959, 21.8092]	541.7	0.73
kurtosis	0.5336	0.1203	[0.3369, 0.8004]	511.8	0.69
ζ_1	0.3191	0.0138	[0.2948, 0.3486]	1,562.7	0.56
ζ_2	0.3239	0.0128	[0.2976, 0.3492]	11,124.0	0.55
ξ_1	0.3465	0.0108	[0.3272, 0.3713]	1,211.3	0.71
ξ_2	0.4075	0.0107	[0.3848, 0.4266]	6,868.8	0.67
$\sigma_{u,1}$	0.0250	0.0077	[0.0113, 0.0399]	587.3	0.90
$\sigma_{u,2}$	0.0101	0.0014	[0.0074, 0.0130]	306.9	0.83
SNR_1	11.4948	4.8244	[5.7237, 23.8847]	562.7	0.95
SNR_2	10.2058	1.6612	[7.4699, 13.9673]	305.2	0.80
p_{11}	0.8832	0.0187	[0.8430, 0.9161]	40.1	0.88
p_{12}	0.1168	0.0187	[0.0839, 0.1570]	40.1	0.88
p_{21}	0.0302	0.0049	[0.0216, 0.0407]	36.2	0.90
p_{22}	0.9698	0.0049	[0.9593, 0.9784]	36.2	0.90
μ	-0.7461	0.1417	[-1.0491, -0.4918]	3.7	0.94
ϕ	0.9867	0.0019	[0.9830, 0.9905]	4.8	0.95
σ_1	0.2584	0.0137	[0.2306, 0.2846]	73.0	0.72
σ_2	0.0996	0.0034	[0.0930, 0.1062]	125.1	0.88
ρ_1	0.0150	0.0375	[-0.0649, 0.0827]	16.2	0.90
ρ_2	0.0224	0.0216	[-0.0207, 0.0637]	10.1	0.85
ζ_1	0.2617	0.0105	[0.2412, 0.2835]	1,033.6	0.89
ζ_2	0.2729	0.0089	[0.2568, 0.2915]	6,228.4	0.39
ξ_1	0.3493	0.0086	[0.3327, 0.3669]	1,028.2	0.65
ξ_2	0.4122	0.0092	[0.3950, 0.4301]	3,703.1	0.89
$\sigma_{u,1}$	0.0214	0.0067	[0.0105, 0.0351]	638.2	0.95
$\sigma_{u,2}$	0.0102	0.0015	[0.0076, 0.0132]	356.6	0.90
SNR_1	13.5754	5.1022	[6.7958, 25.7449]	546.1	0.97
SNR_2	10.0094	1.6225	[7.2926, 13.5356]	331.7	0.94
p_{11}	0.8812	0.0176	[0.8442, 0.9129]	32.6	0.94
p_{12}	0.1188	0.0176	[0.0871, 0.1558]	32.6	0.94
p_{21}	0.0306	0.0048	[0.0222, 0.0409]	34.0	0.72
p_{22}	0.9694	0.0048	[0.9591, 0.9778]	34.0	0.72

Posterior mean, standard deviation, 95% credible bounds, inefficiency factor, and Geweke's convergence diagnostic (p-value). Second panel (MS2SVt only): Implied higher moments of ϵ_t. Middle two panels: Parameters auxiliary measurement eq. and implied signal-to-noise ratios. Bottom panel: Transition probabilities.

Altered priors: MS2SVt-VIX: $\boldsymbol{\zeta} \backsim \mathcal{N}((0.30, 0.30)', \boldsymbol{I}_2), \boldsymbol{\xi} \backsim \mathcal{N}((0.35, 0.40)', \boldsymbol{I}_2)$,
MS2SV-VIX: $\boldsymbol{\zeta} \backsim \mathcal{N}((0.25, 0.25)', \boldsymbol{I}_2), \boldsymbol{\xi} \backsim \mathcal{N}((0.35, 0.40)', \boldsymbol{I}_2)$.

Table 33: Specification Tests I

	Kupier	(s.e.$\times 10$)	Skewness	(s.e.$\times 10^2$)	Kurtosis	(s.e.$\times 10^2$)
SV	3.06	(0.2)	0.089	(0.1)	-0.062	(0.2)
SVt	2.55	(0.3)	0.021	(0.1)	-0.134	(0.2)
SVsk$t$	2.63	(0.3)	0.019	(0.1)	-0.153	(0.2)
MS3SVt	2.38	(0.2)	0.003	(0.9)	-0.133	(3.5)
MS3SVsk$t$	2.43	(0.4)	0.014	(0.5)	-0.105	(1.8)
SV-RV	2.88	(0.2)	0.128	(0.0)	-0.008	(0.1)
SVt-RV	2.61	(0.2)	0.077	(0.6)	-0.066	(2.3)
SVsk$t$-RV	2.62	(0.1)	0.082	(0.5)	-0.048	(2.1)
MS4SV-RV	2.62	(0.2)	0.053	(0.0)	-0.128	(0.1)
MS4SVt-RV	2.39	(0.1)	0.023	(0.1)	-0.126	(0.3)
MS4SVsk$t$-RV	2.46	(0.2)	0.021	(0.0)	-0.130	(0.2)
SVt-Range	3.18	(0.2)	0.074	(0.1)	-0.114	(0.3)
SVsk$t$-Range	3.31	(0.3)	0.036	(0.4)	-0.200	(1.6)
MS4SV-Range	7.04	(0.3)	-0.238	(0.1)	-0.133	(0.5)
MS4SVt-Range	7.04	(0.3)	-0.253	(0.1)	-0.137	(0.2)
MS4SVsk$t$-Range	7.45	(0.5)	-0.290	(0.1)	-0.143	(0.2)
SV-VIX	4.00	(0.2)	0.173	(0.0)	0.005	(0.1)
SVt-VIX	3.02	(0.2)	0.077	(0.1)	-0.113	(0.3)
SVsk$t$-VIX	3.07	(0.1)	0.086	(0.1)	-0.097	(0.7)
MS2SV-VIX	3.92	(0.3)	0.163	(0.1)	-0.016	(0.4)
MS2SVt-VIX	3.07	(0.5)	0.068	(0.1)	-0.135	(0.3)
MS2SVsk$t$-VIX	3.16	(0.2)	-0.135	(0.1)	-0.304	(0.2)

Kupier statistic of $\hat{u}_t$, skewness and kurtosis of transformed forecasts $\hat{n}_t$. Critical value for V is 2.001 at the 1% level (H_0: $\hat{u}_t \backsim \mathcal{U}(0,1)$). Numerical standard error in parentheses.

E Specification Tests

Tab. 33 reports the Kupier statistic of probability integral transforms $\hat{u}_t$ and skewness and kurtosis estimates of transformed forecasts $\hat{n}_t$. Standard errors are provided by repeatedly applying the associated particle filter ten times, see Sec. 4. The Kupier statistic tests for uniformity and is equally sensitive for all values of $\hat{u}_t$. In contrast, the more well-known Kolmogorov-Smirnov statistic is most sensitive around the median. The Kupier statistic V is defined as

$$D^+ = \max_{1\leq t\leq n}[t/n - u_t], \quad D^- = \max_{1\leq t\leq n}[u_t - (t-1)/n], \quad V = D^+ + D^-,$$

under the null of uniformly distributed u_t.

For all models using no auxiliary equation or RV the Kupier statistic is rather close to the critical value of 2.001 for the 1% significance level[15]. Especially the regime switching range variants, but also the SV-VIX and MS2SV-VIX perform worse, mirroring the respective predictive likelihoods in Tab. 13 or equally, the LPDRs in Tab. 34. A graphical analysis of histograms of $\hat{u}_t$ together with approximate confidence intervals in the spirit of Bauwens, Giot, Grammig, and Veredas (2004) visualizes these deficits in predictive ability, see Fig. 14. The remaining skewness in

[15]The asymptotic distribution of the statistic is known, but Stephens (1970) provides a simplified finite-sample statistic with negligible error to its asymptotic distribution.

Table 34: Specification Tests II

	LPDR	(s.e.)	Rank	LB_{15}^2	(s.e.)	LB_{15}^3	(s.e.)	LB_{15}^4	(s.e.)
y_1 (return)				5653.5		877.7		2183.3	
SV	0	(0.3)	14	14.2	(0.1)	40.0	(0.1)	14.7	(0.1)
SVt	14.7	(0.7)	12	15.6	(0.2)	38.6	(0.2)	15.8	(0.2)
SVsk$t$	26.1	(1.0)	8	16.1	(0.1)	39.4	(0.2)	16.1	(0.2)
MS3SVt	33.9	(0.9)	7	15.6	(0.2)	45.7	(0.2)	16.4	(0.2)
MS3SVsk$t$	43.8	(1.5)	5	16.5	(0.3)	45.6	(0.3)	17.3	(0.3)
SV-RV	24.9	(0.3)	9	15.5	(0.0)	50.1	(0.1)	13.6	(0.0)
SVt-RV	41.8	(0.3)	6	15.2	(0.0)	47.6	(0.0)	12.9	(0.0)
SVsk$t$-RV	51.0	(0.5)	4	15.1	(0.1)	46.6	(0.1)	12.8	(0.0)
MS4SV-RV	59.9	(0.2)	2	16.2	(0.1)	47.6	(0.1)	13.1	(0.0)
MS4SVt-RV	62.8	(0.2)	1	16.2	(0.0)	46.1	(0.1)	13.2	(0.1)
MS4SVsk$t$-RV	51.6	(0.3)	3	16.3	(0.0)	46.7	(0.1)	13.2	(0.0)
SVt-Range	-46.8	(0.3)	19	18.3	(0.1)	100.4	(0.1)	17.6	(0.1)
SVsk$t$-Range	-36.5	(0.9)	18	18.2	(0.1)	93.5	(0.3)	17.0	(0.1)
MS4SV-Range	-116.5	(1.0)	20	20.1	(0.3)	31.3	(0.3)	24.2	(0.3)
MS4SVt-Range	-117.9	(1.0)	21	19.9	(0.4)	31.0	(0.3)	24.2	(0.3)
MS4SVsk$t$-Range	-137.1	(1.1)	22	19.8	(0.2)	28.9	(0.2)	24.3	(0.2)
SV-VIX	-25.0	(0.4)	17	30.1	(0.1)	49.3	(0.1)	36.4	(0.2)
SVt-VIX	24.0	(0.5)	10	28.8	(0.1)	46.0	(0.1)	31.2	(0.1)
SVsk$t$-VIX	11.4	(0.5)	13	28.8	(0.1)	46.3	(0.1)	31.7	(0.1)
MS2SV-VIX	-15.8	(1.4)	16	30.4	(0.6)	47.2	(0.5)	35.3	(1.0)
MS2SVt-VIX	22.3	(1.7)	11	30.0	(0.3)	44.6	(0.4)	31.1	(0.3)
MS2SVsk$t$-VIX	-5.1	(0.5)	15	24.2	(0.1)	47.1	(0.1)	22.3	(0.1)

Cumulative log predictive density ratio. Ljung-Box statistic of demeaned $\hat{u}_t^2$, $\hat{u}_t^3$, $\hat{u}_t^4$ and $y_{1,t}^2$, $y_{1,t}^3$, $y_{1,t}^4$. Critical values for the $LB(15)$ test are $30.6/25.0/22.3$ at the 1%/5%/10% significance level. Numerical standard error in parentheses.

the transformed one-step-ahead forecasts appears in general small and is positive in almost all models, indicating that the negative outliers have been appropriately modeled. However, similarly to the Kupier statistic, the regime switching range variants negatively stand out. Interestingly, there is some remaining negative skewness for the MS2SVsk$t$-VIX model, despite its second state modeling extreme negative skewness explicitly. However, as already outlined in the main part, transient regime dynamics may have weaker positive impact on forecasting performance, more so in a time invariant transition probability context. Remaining excess kurtosis in the data is overall negative and small in absolute value. This is rather satisfying under a conservative viewpoint.

Tab. 34 reports log predictive density ratios relative to the basic SV model and Ljung-Box statistics at lag 15 for $\hat{u}_t^2$, $\hat{u}_t^3$, $\hat{u}_t^4$ and y_1. The LPDRs make the better potential predictive ability of models with no additional time series, RV or that of some VIX variants clearly visible. The $LB(15)$ statistics support this picture, but are satisfying for all models in general, especially for dynamics in the second and fourth moment. The relatively high statistics for the third moment indicate some remaining dynamics, but note that the $\mathrm{AR}(1)$ effect, which is common in stock index data, has not been removed. Moreover, Nakajima and Omori (2012) estimate the SVsk$t$ model on different sample periods of the S&P 500 and find changing skewness of the empirical return distribution.

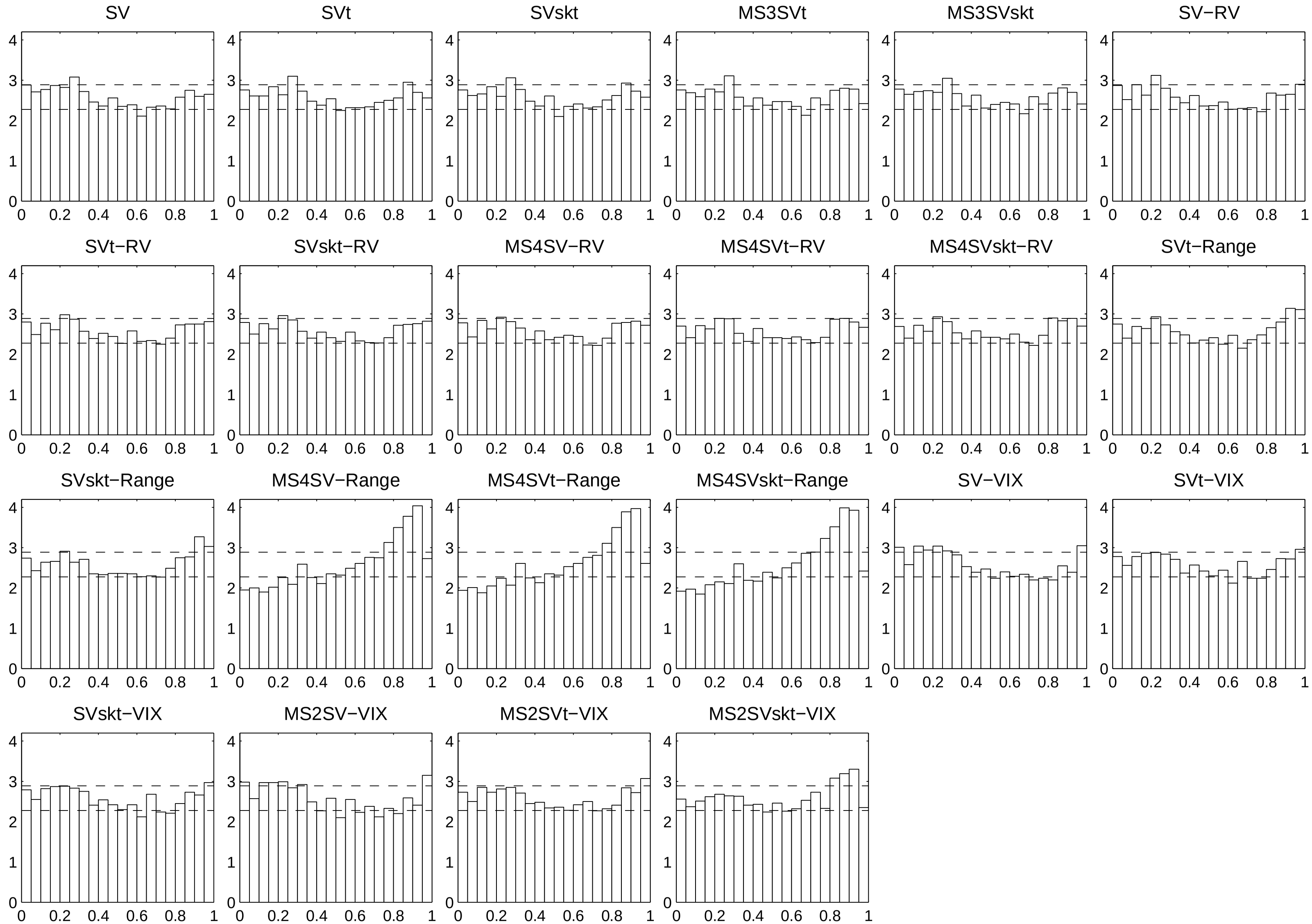

Figure 14: Transformed Predictions $\hat{u}_t$ (in-sample, grand averages). Horizontal dashed lines are 95% confidence intervals based on a normal approximation of the binomially distributed bin elements (H_0: $\hat{u}_t \backsim \mathcal{U}(0,1)$). Each bin contains 5% of the data. Y-axis in hundreds.

References

Aas, K. and I. H. Haff (2006). The Generalized Hyperbolic skew Student's t-distribution. *Journal of Financial Econometrics 4*(2), 275–309.

Abanto-Valle, C. A., D. Bandyopadhyay, V. H. Lachos, and I. Enriquez (2010). Robust Bayesian analysis of heavy-tailed stochastic volatility models using scale mixtures of normal distributions. *Computational Statistics & Data Analysis 54*(12), 2883–2898.

Abanto-Valle, C. A., V. H. Lachos, and D. K. Dey (2012). Stock return volatility, heavy tails, skewness and trading volume: A Bayesian approach. Submitted.

Abanto-Valle, C. A., V. H. Lachos, and D. K. Dey (2013). Bayesian estimation of a Skew-Student-t stochastic volatility model. *Methodology and Computing in Applied Probability*, 1–18.

Abanto-Valle, C. A., H. S. Migon, and H. F. Lopes (2010). Bayesian modeling of financial returns: A relationship between volatility and trading volume. *Applied Stochastic Models in Business and Industry 26*(2), 172–193.

Abramowitz, M. and I. A. Stegun (1972). *Handbook of Mathematical Functions*. New York, USA: Dover Publications.

Alizadeh, S., M. W. Brandt, and F. X. Diebold (2002). Range-based estimation of stochastic volatility models. *The Journal of Finance 57*(3), 1047–1091.

Andersen, T. G. (1996). Return volatility and trading volume: An information flow interpretation of stochastic volatility. *The Journal of Finance 51*(1), 169–204.

Andersen, T. G. and L. Benzoni (2008). Stochastic volatility. Chapter prepared for the Encyclopedia of Complexity and System Science (Springer). Available at SSRN: http://ssrn.com/abstract=1076672 or http://dx.doi.org/10.2139/ssrn.1076672.

Andersen, T. G. and T. Bollerslev (1998). Answering the skeptics: Yes, standard volatility models do provide accurate forecasts. *International Economic Review 39*(4), 885–905.

Andersen, T. G., T. Bollerslev, and F. X. Diebold (2010). Parametric and nonparametric volatility measurement. In Y. Aït-Sahalia and L. P. Hansen (Eds.), *Handbook of Financial Econometrics: Tools and Techniques*, Volume 1 of *Handbooks in Finance*, pp. 67–137. San Diego, USA: North-Holland.

Andrews, D. W. K. (1991). Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica 59*(3), 817–858.

Asai, M. (2008). Autoregressive stochastic volatility models with heavy-tailed distributions: A comparison with multifactor volatility models. *Journal of Empirical Finance 15*(2), 332–341.

Barndorff-Nielsen, O. E. (1977). Exponentially decreasing distributions for the logarithm of particle size. *Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences 353*(1674), 401–419.

Barndorff-Nielsen, O. E. and P. Blæsid (1981). Hyperbolic distributions and ramnifications: Contributions to theory and application. *Statistical Distributions in Scientific Work 4*, 19–44.

Barndorff-Nielsen, O. E., P. R. Hansen, A. Lunde, and N. Shephard (2008). Designing realized kernels to measure the ex post variation of equity prices in the presence of noise. *Econometrica 76*(6), 1481–1536.

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

Barndorff-Nielsen, O. E. and N. Shephard (2001b). Normal modified stable processes. *Theory of Probability and Mathematical Statistics*, 1–19.

Bauwens, L., P. Giot, J. Grammig, and D. Veredas (2004). A comparison of financial duration models via density forecasts. *International Journal of Forecasting 20*(4), 589–609.

Black, F. (1976). Studies in stock price volatility changes. In *Proceedings of the 1976 Meeting of the Business and Economic Statistics Section*, pp. 177–181. American Statistical Association.

Bos, C. S. (2011). Relating stochastic volatility estimation methods. Discussion Paper 11-049/4 Tinbergen Institute and VU University Amsterdam.

Bouchaud, J.-P. and M. Potters (2009). *Theory of Financial Risk and Derivative Pricing*. Cambridge, UK: Cambridge University Press.

Brandt, M. W. and C. S. Jones (2005). Bayesian range-based estimation of stochastic volatility models. *Finance Research Letters 2*(4), 201–209.

Broto, C. and E. Ruiz (2004). Estimation methods for stochastic volatility models: A survey. *Journal of Economic Surveys 18*(5), 613–649.

Bulla, J. (2011). Hidden Markov models with t components. increased persistence and other aspects. *Quantitative Finance 11*(3), 459–475.

Cappuccio, N., D. Lubian, and D. Raggi (2004). McMC Bayesian estimation of a Skew-GED stochastic volatility model. *Studies in Nonlinear Dynamics & Econometrics 8*(2). Article 6.

Carvalho, C. M. and H. F. Lopes (2007). Simulation-based sequential analysis of Markov switching stochastic volatility models. *Computational Statistics & Data Analysis 51*(9), 4526–4542.

Chan, J. C.-C. and I. Jeliazkov (2009). McMC estimation of restricted covariance matrices. *Journal of Computational and Graphical Statistics 18*(2), 457–480.

Chernov, M., A. R. Gallant, E. Ghysels, and G. Tauchen (2003). Alternative models for stock price dynamics. *Journal of Econometrics 116*(1-2), 225–257. Frontiers of financial econometrics and financial engineering.

Chib, S. (1996). Calculating posterior distributions and modal estimates in Markov mixture models. *Journal of Econometrics 75*(1), 79–97.

Chib, S. and E. Greenberg (1995). Understanding the Metropolis-Hastings algorithm. *Journal of the American Statistical Association 49*(4), 327–335.

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

Christie, A. A. (1982). The stochastic behavior of common stock variances: Value, leverage and interest rate effects. *Journal of Financial Economics 10*(4), 407–432.

Clark, P. K. (1973). A subordinated stochastic process model with finite variance for speculative prices. *Econometrica 41*(1), 135–155.

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

Diebold, F. X., T. A. Gunther, and A. S. Tay (1998). Evaluating density forecasts with applications to financial risk management. *International Economic Review 39*(4), 863–883.

Dobrev, D. and P. Szerszen (2010). The information content of high-frequency data for estimating equity return models and forecasting risk. Discussion Paper, Number 1005, Board of Governors of the Federal Reserve System.

Douc, R., O. Cappé, and E. Moulines (2005). Comparison of resampling schemes for particle filtering. In *4th International Symposium on Image and Signal Processing and Analysis (ISPA)*, pp. 64–69.

Doucet, A., N. de Freitas, and N. Gordon (2001). *Sequential Monte Carlo Methods in Practice*. New York, USA: Springer.

Durbin, J. and S. J. Koopman (2002). A simple and efficient simulation smoother for state space time series analysis. *Biometrika 89*(3), 603–615.

Durbin, J. and S. J. Koopman (2008). *Time Series Analysis by State Space Methods*. Oxford, UK: Oxford University Press.

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

Frühwirth-Schnatter, S. (2001). Markov chain Monte Carlo estimation of classical and dynamic switching and mixture models. *Journal of the American Statistical Association 96*(453), 194–209.

Frühwirth-Schnatter, S. (2006). *Finite Mixture and Markov Switching Models*. New York, USA: Springer.

Geweke, J. (1992). Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments. In J. M. Bernardo, J. O. Berger, A. P. Dawid, and A. F. M. Smith (Eds.), *Bayesian Statistics*, Volume 4, pp. 169–188. New York, USA: Oxford University Press.

Hamilton, J. D. (1989). A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica 57*(2), 357–384.

Hansen, P. R. and A. Lunde (2006). Realized variance and market microstructure noise. *Journal of Business & Economic Statistics 24*(2), 127–161.

Harvey, A. C. (1989). *Forecasting, Structural Time Series Models and the Kalman Filter*. Cambridge, UK: Cambridge University Press.

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

Hol, J. D., T. B. Schön, and F. Gustafsson (2006). On resampling algorithms for particle filters. In *Nonlinear Statistical Signal Processing Workshop*, Cambridge, UK. IEEE.

Hu, W. (2009). *Calibration of Multivariate Generalized Hyperbolic Distributions*. Saarbrücken, GER: VDM Publishing.

Jacquier, E. and S. Miller (2012). The information content of realized volatility. Working paper, HEC Montreal.

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

Jacquier, E., N. G. Polson, and P. E. Rossi (2004). Bayesian analysis of stochastic volatility models with fat-tails and correlated errors. *Journal of Econometrics 122*(1), 185–212.

Jensen, M. J. and J. M. Maheu (2010). Bayesian semiparametric stochastic volatility modeling. *Journal of Econometrics 157*(2), 306–316.

Kim, C.-J. and C. R. Nelson (1998). Business cycle turning points, a new coincident index, and tests of duration dependence based on a dynamic factor model with regime switching. *The Review of Economics and Statistics 80*(2), 188–201.

Kim, C.-J. and C. R. Nelson (1999). *State-Space Models with Regime Switching*. Cambridge, USA: MIT Press.

Kim, S., N. Shephard, and S. Chib (1998). Stochastic volatility: Likelihood inference and comparison with ARCH models. *The Review of Economic Studies 65*(3), 361–393.

Koopman, S. J. (1993). Disturbance smoother for state space models. *Biometrika 80*(1), 117–126.

Koopman, S. J., B. Jungbacker, and E. Hol (2005). Forecasting daily variability of the S&P 100 stock index using historical, realised and implied volatility measurements. *Journal of Empirical Finance 12*(3), 445–475.

Koopman, S. J. and M. Scharth (2012). The analysis of stochastic volatility in the presence of daily realized measures. *Journal of Financial Econometrics 11*(1), 76–115.

Liesenfeld, R. and J.-F. Richard (2003). Univariate and multivariate stochastic volatility models: Estimation and diagnostics. *Journal of Empirical Finance 10*(4), 505–531.

Lo, A. W. and M. T. Müller (2010). Warning: Physics envy may be hazardous to your wealth! *Journal of Investment Management (JOIM) 8*(2). Available at SSRN: http://ssrn.com/abstract=1639085.

Mahieu, R. and R. Bauer (1998). A Bayesian analysis of stock return volatility and trading volume. *Applied Financial Economics 8*(6), 671–687.

McAleer, M. and M. C. Medeiros (2008). Realized volatility: A review. *Econometric Reviews 27*(1-3), 10–45.

Molina, G., C.-H. Han, and J.-P. Fouque (2010). McMC estimation of multiscale stochastic volatility models. In C.-F. Lee and J. Lee (Eds.), *Handbook of Quantitative Finance and Risk Management*, Volume 5, pp. 1109–1120. New York, USA: Springer.

Nakajima, J. and Y. Omori (2009). Leverage, heavy-tails and correlated jumps in stochastic volatility models. *Computational Statistics & Data Analysis 53*(6), 2335–2353. The Fourth Special Issue on Computational Econometrics.

Nakajima, J. and Y. Omori (2012). Stochastic volatility model with leverage and asymmetrically heavy-tailed error using GH skew Student's t-distribution. *Computational Statistics & Data Analysis 56*(11), 3690–3704. 1st issue of the Annals of Computational and Financial Econometrics. Sixth Special Issue on Computational Econometrics.

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

Omori, Y. and T. Watanabe (2007). Block sampler and posterior mode estimation for a nonlinear and non-Gaussian state-space model with correlated errors. Working Paper, CARF-F-104, Center for Advanced Research in Finance, Faculty of Economics, The University of Tokyo.

Omori, Y. and T. Watanabe (2008). Block sampler and posterior mode estimation for asymmetric stochastic volatility models. *Computational Statistics & Data Analysis 52*(6), 2892–2910.

Pemstein, D., K. M. Quinn, and A. D. Martin (2011). The Scythe statistical library: An open source C++ library for statistical computation. *Journal of Statistical Software 42*(12), 1–26.

Pitt, M. K. and N. Shephard (1999a). Filtering via simulation: Auxiliary particle filters. *Journal of the American Statistical Association 94*(446), 590–599.

Prause, K. (1999). The Generalized Hyperbolic models: Estimation, financial derivatives and risk measurement. PhD dissertation, University of Freiburg.

Robert, C. P. and G. Casella (2004). *Monte Carlo Statistical Methods*. New York, USA: Springer.

Rosenblatt, M. (1952). Remarks on a multivariate transformation. *The Annals of Mathematical Statistics 23*(3), 470–472.

Scott, D. J., D. Würtz, C. Dong, and T. T. Tran (2011). Moments of the Generalized Hyperbolic distribution. *Computational Statistics 26*(3), 459–476.

Shephard, N. and M. K. Pitt (1997). Likelihood analysis of non-Gaussian measurement time series. *Biometrika 84*(3), 653–667.

Shibata, M. and T. Watanabe (2005). Bayesian analysis of a Markov switching stochastic volatility model. *Journal of the Japan Statistical Society 35*(2), 205–219.

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

Spiegelhalter, D. J., N. G. Best, B. P. Carlin, and A. van der Linde (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 64*(4), 583–639.

Stephens, M. A. (1970). Use of the Kolmogorov-Smirnov, Cramér-von Mises and related statistics without extensive tables. *Journal of the Royal Statistical Society. Series B (Methodological) 32*(1), 115–122.

Takahashi, M., Y. Omori, and T. Watanabe (2009). Estimating stochastic volatility models using daily returns and realized volatility simultaneously. *Computational Statistics & Data Analysis 53*(6), 2404–2426. The Fourth Special Issue on Computational Econometrics.

Takahashi, M., T. Watanabe, and Y. Omori (2014). Volatility and quantile forecasts by realized stochastic volatility models with generalized hyperbolic distribution. Discussion Paper, CIRJE-F-921, Graduate School of Economics, Faculty of Economics, The University of Tokyo.

Tierney, L. (1994). Markov chains for exploring posterior distributions. *The Annals of Statistics 22*(4), 1701–1728.

Watanabe, T. (2000). Bayesian analysis of dynamic bivariate mixture models: Can they explain the behavior of returns and trading volume? *Journal of Business & Economic Statistics 18*(2), 199–210.

Watanabe, T. and Y. Omori (2004). A multi-move sampler for estimating non-Gaussian time series models: Comments on Shephard & Pitt (1997). *Biometrika 91*(1), 246–248.

West, M. and J. Harrison (1997). *Bayesian Forecasting and Dynamic Models* (2nd ed.). New York, USA: Springer.

Whiteley, N. and A. M. Johansen (2011). *Bayesian Time Series Models*. Cambridge, UK: Cambridge University Press.

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

Zhang, L., P. A. Mykland, and Y. Aït-Sahalia (2005). A tale of two time scales: Determining integrated volatility with noisy high-frequency data. *Journal of the American Statistical Association 100*(472), 1394–1411.