Leverage, asymmetry and heavy tails in the high-dimensional factor stochastic volatility model*

Mengheng Li†

University of Technology Sydney, UTS Business School, Australia

Marcel Scharth‡

University of Sydney Business School, Australia

Working paper; this version: August 24, 2018

Abstract

We develop a flexible modeling and estimation framework for a high-dimensional factor stochastic volatility (SV) model. Our specification allows for leverage effects, asymmetry and heavy tails across all systematic and idiosyncratic components of the model. This framework accounts for well-documented features of univariate financial time series, while introducing a flexible dependence structure that incorporates tail dependence and asymmetries such as stronger correlations following downturns. We develop an efficient Markov chain Monte Carlo (MCMC) algorithm for posterior simulation based on the particle Gibbs, ancestor sampling, and particle efficient importance sampling methods. We build computationally efficient model selection into our estimation framework to obtain parsimonious specifications in practice. We validate the performance of our proposed estimation method via extensive simulation studies for univariate and multivariate simulated datasets. An empirical study shows that the model outperforms other multivariate models in terms of value-at-risk evaluation and portfolio selection performance for a sample of US and Australian stocks.

Keywords: *Generalised hyperbolic skew Student's t-distribution; Metropolis-Hastings algorithm; Importance sampling; Particle filter; Particle Gibbs; State space model; Time-varying covariance matrix; Factor model*

JEL Classification: C11; C32; C53; C55; G32

*We would like to thank George Tauchen, Richard Gerlach, Robert Kohn, Gary Koop, Siem Jan Koopman, Frank Kleibergen, Catherine Forbes, Anne Opschoor, and seminar and workshop participants at the University of Sydney Business School, VU University Amsterdam, Tinbergen Institute, the 10th International Conference on Computational and Financial Econometrics (Seville, 2016), the 10th Society of Financial Econometrics Annual Conference (New York, 2017), the 1st International Conference on Econometrics and Statistics (Hong Kong, 2017), the 8th European Seminar on Bayesian Econometrics (Maastricht, 2017) for useful comments and helpful suggestions on previous versions of this paper. Any remaining errors are the responsibility of the authors.

†Email: mengheng.li@uts.edu.au; PO BOX 123, Broadway NSW 2007, Australia

‡Email: marcel.scharth@sydney.edu.au

1 Introduction

The vast literature on univariate and multivariate financial times series provides compelling evidence for the presence of time-varying volatilities, time-varying correlations, leverage effects, conditionally heavy tails, and often skewness as common features of stock and index return series. Stochastic volatility (SV) is large class of models that incorporate these stylized facts into a range of univariate and multivariate specifications. Among many others, Shephard and Pitt (1997) and Durbin and Koopman (1997) develop estimation procedures for SV models with Student's t errors. Koopman and Hol Uspensky (2002) and Yu (2005) discuss leverage effects in SV models. Asai et al. (2006) and Chib et al. (2009) review several approaches for multivariate stochastic volatility (MSV).

This paper addresses the challenge of extending flexible stochastic volatility models to the multivariate setting. We follow Pitt and Shephard (1999a) and Chib et al. (2006) and consider a factor stochastic volatility (FSV) framework. FSV models represent each individual asset return as a linear combination of factor (shared) innovations and an asset-specific innovation, where each model component follows a univariate SV process. To achieve flexibility in this multivariate setting, we introduce a model that allows every systematic and idiosyncratic component to follow the univariate SV model proposed by Nakajima and Omori (2012), which incorporates leverage effects, skewness, heavy-tails via the (generalized hyperbolic) skew Student's t distribution of Aas and Haff (2006).

The motivation for Nakajima and Omori (2012) model as the basis of the factor SV specification is twofold. First, it ensures that each marginal series is consistent with a flexible specification found to be empirically accurate in the univariate SV context. Second, the Nakajima and Omori (2012) specification for the factor components leads to higher flexibility in the dependence structure, allowing us to incorporate important multivariate features into the model. Ang and Chen (2002) and Patton (2004) find that correlations between US stocks are much higher during downturns, and for downside moves. By incorporating factor leverage effects, and allowing for asymmetric factor innovations, our model can account for these asymmetries. Beine et al. (2010) and Oh and Patton (2017a), among others, emphasize the important of tail co-movements, moving the use of a heavy-tailed distribution for the factor innovations in our model.

We follow a Bayesian approach and develop an efficient Markov Chain Monte Carlo (MCMC) algorithm for the univariate and factor versions of the model based on the particle Gibbs method

of Andrieu et al. (2010). Our sampling scheme for latent SV processes builds on the efficient importance sampling (EIS) algorithm of Richard and Zhang (2007) and the particle Gibbs with ancestral sampling algorithm of Lindsten et al. (2014). The EIS method constructs a globally optimal approximation to the distribution of the latent volatility process conditional on the data which is typically orders of magnitude more accurate than methods based on local approximation, such as those used in multi-move sampling. In this paper, we use the particle Gibbs sampler of (Lindsten et al., 2014)as an efficient way to implement the EIS proposal within our MCMC scheme. Grothe et al. (2017) follow a similar approach for general state space models, including a canonical SV model.

Other common methods for posterior simulation in SV models include the pseudo-marginal Metropolis-Hastings algorithm of (Andrieu and Roberts, 2009; Andrieu et al., 2010), Gibbs sampling with data augmentation (Kim et al., 1998), and Metropolis-within-Gibbs (Gilks et al., 1995, Geweke and Tanizaki, 2001, Koop et al., 2007, Watanabe and Omori, 2004) sampling. In a classical setting, several authors have considered simulated maximum likelihood (Durbin and Koopman, 1997; Liesenfeld and Richard, 2006). All of these approaches may infeasible or inefficient for complex SV specifications, or in high-dimensional settings. Our simulation and empirical results of Sections 4 and 5 show that our particle Gibbs approach is instrumental for efficient estimation in practice, whether in the univariate or multivariate setting.

Scalability in the number of assets is a primary concern in multivariate financial time series modeling. Quantitative investment funds typically have tens and even hundreds of positions in their portfolio, requiring the development of high-dimensional multivariate models for risk and investment management (Dempster et al. 2008, Vardi 2015). However, multivariate models typically suffer from the curse of dimensionality: the number of model parameters or the computational cost can grow fast with number of assets, making estimation extremely challenging for large portfolio sizes. As in Chib et al. (2006), our factor SV specification addresses this problem in two ways. First, the number of parameters is a linear function of the number of assets, as in other factor models. Second, the factor structure leads to a convenient sampling scheme which reduces to parallel treatment of many univariate series after marginalization of the factors. With the use of the efficient simulation methods proposed in this paper, we are able to estimate our general specification for moderately large number of assets (up to 80 in our empirical application), despite the complexity of the model.

We organize the discussion as follows. Section 2 discusses the Nakajima and Omori (2012) model and introduces our new Bayesian estimation method for the univariate case. Section 3

extends the model and methodology of Section 2 to the factor SV setting. Section 4 presents simulation studies for the univariate model of Nakajima and Omori (2012) and our new the high-dimensional factor model. Section 5 presents empirical applications for two portfolios based on stock return for of the S&P100 and ASX50, including extensive comparison with alternative methods in terms of key metrics such as value-at-risk (VaR) accuracy and minimum-variance portfolio performance. Section 6 concludes.

2 Univariate estimation

In this section we develop an efficient Bayesian estimation method for the univariate SV model of Nakajima and Omori (2012). Our objective is to introduce key ideas and methods that will apply more generally to the factor SV model of Section 3. In particular, our MCMC algorithm for the multivariate model will follow from the efficient univariate particle Gibbs sampler of Section 2.2. Our simulation results in Section 4 suggest that our method can lead to large gains in efficiency and reliability compared to the MCMC algorithm proposed in Nakajima and Omori (2012), extending the scope of their flexible univariate SV model.

2.1 SV with leverage effects, skewness, and heavy tails

Nakajima and Omori (2012) introduce the following model

$$
\begin{aligned}
&y_t = \exp(h_t/2)\nu_t, && t = 1, ..., T, \\
&\nu_t = \alpha + \beta W_t \gamma + \sqrt{W_t}\gamma\epsilon_t, && t = 1, ..., T, \\
&h_{t+1} = \mu(1-\phi) + \phi h_t + \eta_t, && t = 2, ..., T-1, \\
&h_1 \sim \ \mathrm{N}\left(\mu, \frac{\sigma^2}{1-\phi^2}\right), && \\
&\begin{bmatrix} \epsilon_t \\ \eta_t \end{bmatrix} \sim N\Big(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & \rho\sigma \\ \rho\sigma & \sigma^2 \end{bmatrix} \Big), && t = 1, ..., T, \\
&W_t \sim IG\left(\frac{\zeta}{2}, \frac{\zeta}{2}\right), && t = 1, ..., T,
\end{aligned}
\tag{1}
$$

where y_t is a return series, h_t is the unobserved log-volatility (modeled as a stationary AR(1) process), ν_t is the return innovation, and η_t is the log-volatility innovation. The return innovation follows the (standardized and) generalized hyperbolic skew Student's t-distribution, which we write as a location-scale mixture of Gaussian variables for convenience. The mixing variable

W_t follows the inverse Gamma (IG) distribution. We choose $\alpha = -\beta\zeta/(\zeta-2)$ and $\gamma = (\mu_W + \beta^2\sigma_W^2)^{-1/2}$ with μ_W and σ_W^2 denoting the mean and variance of $IG(\zeta/2, \zeta/2)$-distributed W_t, respectively, so that ν_t has zero mean and unit variance. $\zeta > 4$ is imposed to ensure the existence of a finite variance. The asymmetry parameter β and the degrees of freedom ζ jointly determine the skewness and heavy-tailedness of ν_t.

Aas and Haff (2006) provide a detailed account of generalized hyperbolic skew Student's t-distribution including its density function f_ν, the p-th moment $\mathbb{E}(|\nu|^p)$, and an EM algorithm for parameter estimation. $\beta = 0$ corresponds to a symmetric Student's t-distribution for ν_t and a standard normal distribution if ζ further becomes large. As argued by Aas and Haff (2006), a unique feature of the distribution of ν_t is that in the tails

$$f_\nu(\nu) \propto |\nu|^{-\zeta/2-1} exp(-|\beta\nu| + \beta\nu) \quad \text{as } \nu \to \pm\infty.$$

This means that f_ν has semi-heavy tails, meaning that depends on the value of β and ζ one tail can decay polynomially (heavy tail) whereas the other exponentially (light tail), which is an appealing feature for financial data.

We allow the log-volatility innovation η_t and the normal component $\varepsilon)_t$ to have correlation ρ, leading to a leverage effect when $\rho < 0$. We can show that $\text{Corr}(\nu_t, \eta_t) = Le(\beta, \zeta)\rho$ where the multiplier $Le(\beta, \zeta)$ is

$$Le(\beta, \zeta) = \frac{\Gamma(\frac{\zeta-1}{2})}{\Gamma(\frac{\zeta}{2})}\sqrt{\frac{(\zeta-2)^2(\zeta-4)}{2\zeta^2 + (4\beta^2-12)\zeta + 16}}, \quad \zeta > 4. \tag{2}$$

The proof is in the supplementary appendix. Basic algebra shows that $Le(\beta, \zeta) \in (0, 1)$, $\forall \beta, \zeta \in \mathbb{R}$ with $\frac{\partial Le}{\partial \zeta} > 0$, $\frac{\partial^2 Le}{\partial \zeta^2} < 0$, $\frac{\partial Le}{\partial |\beta|} < 0$, and $\frac{\partial^2 Le}{\partial \beta^2} < 0$. Given β, when ζ becomes large the density of ν_t is less skewed and has lighter tails (Aas and Haff, 2006), so $Le(\beta, \zeta)$ tends to one or leverage effect tends to ρ, similar to the case of a standard SV model with normal error. Given $\zeta > 4$, the magnitude of leverage decreases to zero with $|\beta|$ even though $\rho \neq 0$. It means that if the return innovation ν_t puts a large weight on the mixing variable W_t (i.e. large $|\beta|$), leverage effect vanishes.

2.2 MCMC algorithm

Let $\theta = (\sigma, \rho, \phi, \mu, \beta, \zeta)$ collect the model parameters and $x_{1:T} = (x_1, x_2, \ldots, x_T)$ for a general process x. We propose an MCMC algorithm that iterates over

1. Sampling $(h_{1:T}, W_{1:T})|y_{1:T}, \theta$;

2. Sampling $\theta|y_{1:T}, h_{1:T}, W_{1:T}$,

as a Metropolis-within-Gibbs procedure (e.g., Gilks et al. 1995, Geweke and Tanizaki 2001 which Koop et al. 2007). The key feature of our algorithm is the joint sampling of the entire trajectories $h_{1:T}$ and $W_{1:t}$ as single block via the particle Gibbs algorithm. In comparison, Nakajima and Omori (2012) sample h_t using the multi-move sampler of Shephard and Pitt (1997) conditional on W_t (see also Watanabe and Omori 2004 and Takahashi et al. 2009); They subsequently sample W_t conditional on h_t using the Metropolis-Hastings algorithm.

2.2.1 Approximating $p(h_{1:T}, W_{1:T}|y_{1:T}, \theta)$ via efficient importance sampling

We turn to the EIS method of Richard and Zhang (2007) to build a proposal for the latent trajectories $h_{1:T}$ and $W_{1:T}$ conditional on the data $y_{1:T}$ and the parameters θ. We rewrite the model as

$$\begin{aligned} y_t &= (\alpha + \beta W_t \gamma + \sqrt{W_t}\gamma\epsilon_t)e^{h_t/2}, && t = 1, ..., T \\ h_{t+1} &= \mu(1-\phi) + \phi h_t + \rho\sigma\bar{\epsilon}_t + \sqrt{1-\rho^2}\sigma\eta_t^*, && t = 1, ..., T-1 \end{aligned}$$

where $\bar{\epsilon}_t = (y_t e^{-h_t/2} - \alpha - \beta W_t\gamma)/\sqrt{W_t}\gamma$, and η_t^* is standard normal independent on ϵ_t. Note that $\bar{\epsilon}_t \in \mathcal{F}_t$, where $\mathcal{F}_t$ is the filtration generated by both observables $y_{1:t}$ and unobservables $h_{1:t}$ and $W_{1:t}$ such that the model is Markovian and y_t forms a martingale difference sequence, allowing factorization of the likelihood.

Model (1) is a non-linear non-Gaussian state space model. Let $x_t = (h_t, W_t)'$ denote the state vector and $p(\cdot)$ a generic density function, possibly with a subscript to indicate a specific distribution. For notational simplicity, we suppress the dependence on θ. The likelihood is given by the integral

$$L(y_{1:T}) = \int p(y_{1:T}, x_{1:T})dx_{1:T} = \int p(y_1|x_1)p(x_1)\prod_{t=2}^{T} p(y_t|x_t)p(x_t|x_{t-1}, y_{t-1})dx_{1:T}, \tag{3}$$

where the transition density for $t = 2, ..., T$ follows

$$\begin{aligned} p(x_t|x_{t-1}, y_{t-1}) &= p_N(h_t|h_{t-1}, y_{t-1}, W_{t-1})p_{IG}(W_t) \\ &= N\big(h_t; \mu(1-\phi) + \phi h_{t-1} + \rho\sigma\bar{\epsilon}_{t-1}, (1-\rho^2)\sigma^2\big) \cdot IG\big(W_t; \frac{\zeta}{2}, \frac{\zeta}{2}\big). \end{aligned} \tag{4}$$

The EIS method suggests the following importance sampler:

$$q(x_{1:T}|y_{1:T}) = q(x_1|y_{1:T}) \prod_{t=2}^{T} q(x_t|x_{t-1}, y_{1:T}),$$

with the conditional density $q(x_t|x_{t-1}, y_{1:T})$ for $t = 2, ..., T$ written as

$$q(x_t|x_{t-1}, y_{1:T}) = \frac{k_q(x_t, x_{t-1}; \delta_t)}{\chi_q(x_{t-1}; \delta_t)} \quad \text{with} \quad \chi_q(x_{t-1}; \delta_t) = \int k_q(x_t, x_{t-1}; \delta_t) dx_t,$$

where $k_q(x_t, x_{t-1}; \delta_t)$ is a kernel in x_t with integration constant $\chi_q(x_{t-1}; \delta_t)$ and δ_t is a set of importance parameters (which are a function of $y_{1:T}$). At the initial period, the importance density is simply

$$q(x_1|, y_{1:T}) = \frac{k_q(x_1; \delta_1)}{\chi_q(\delta_1)} \quad \text{with} \quad \chi_q(\delta_t) = \int k_q(x_1, ; \delta_1) dx_1.$$

Using the above importance density, we can express the likelihood (3) as

$$\begin{aligned} &\int \frac{p(y_1|x_1)p(x_1)}{q(x_1|y_{1:T})} \prod_{t=2}^{T} \frac{p(y_t|x_t)p(x_t|x_{t-1}, y_{t-1})}{q(x_t|x_{t-1}, y_{1:T})} q(x_{1:T}|y_{1:T}) dx_{1:T} \\ =&\frac{1}{\chi_q(\delta_1)} \int \frac{p(y_1|x_1)p(x_1)}{k_q(x_1; \delta_1)/\chi_q(x_1; \delta_2)} \prod_{t=2}^{T} \frac{p(y_t|x_t)p(x_t|x_{t-1}, y_{t-1})}{k_q(x_t, x_{t-1}; \delta_t)/\chi_q(x_t; \delta_{t+1})} q(x_{1:T}|y_{1:T}) dx_{1:T}, \end{aligned} \tag{5}$$

with $\chi_q(x_T; \delta_{T+1}) = 1$.

The EIS method is particularly suitable to our problem since both the inverse Gamma and normal distributions are closed under multiplication. This implies that we can choose an importance kernel that is conjugate with the transition density (4),

$$k_q(x_t, x_{t-1}; \delta_t) = k(x_t, x_{t-1}; \delta_t) \cdot k_p(x_t, x_{t-1}; y_{t-1}),$$

where

$$k_p(x_t, x_{t-1}; y_{t-1}) = p(x_t|x_{t-1}, y_{t-1}) \chi_p(x_{t-1}; y_{t-1}), \text{ with } \chi_p(x_{t-1}; y_{t-1}) = \int k_p(x_t, x_{t-1}; y_{t-1}) dx_t.$$

The likelihood (5) then becomes

$$\chi_q(\delta_1) \int \frac{p(y_1|x_1) \frac{\chi_q(x_1; \delta_2)}{\chi_p(\cdot;)}}{k(x_1; \delta_1)} \prod_{t=2}^{T} \frac{p(y_t|x_t) \frac{\chi_q(x_t; \delta_{t+1})}{\chi_p(x_{t-1}; y_{t-1})}}{k(x_t, x_{t-1}; \delta_t)} q(x_{1:T}|y_{1:T}) dx_{1:T},$$

where $\chi_p(\cdot;)$ corresponds to the integration constant with respect to the unconditional distribution $N(h_1;\mu,\frac{\sigma^2}{1-\phi^2})$ and $IG(W_1;\frac{\zeta}{2},\frac{\zeta}{2})$. It follows from the transition density $p_N(h_t|h_{t-1},y_{t-1},W_{t-1})$ and $p_{IG}(W_t)$ in (4) that

$$k_p(x_t,x_{t-1};y_{t-1})=\exp\left(\frac{\mu(1-\phi)+\phi h_{t-1}+\rho\sigma\bar{\epsilon}_{t-1}}{(1-\rho^2)\sigma^2}h_t-\frac{h_t^2}{\frac{1}{2}(1-\rho^2)\sigma^2}\right)\cdot W_t^{-\frac{\zeta}{2}-1}\exp(-\frac{\zeta}{2}W_t^{-1}).$$

Let $\delta_t=(b_t,c_t,s_t,r_t)$. For conjugacy, we choose the kernel

$$k(x_t,x_{t-1};\delta_t)=\exp(b_th_t-\frac{1}{2}c_th_t^2)\cdot W_t^{s_t}\exp(r_tW_t^{-1}) \tag{6}$$

with the ratio of integration constant given by

$$\frac{\chi_q(x_t;\delta_{t+1})}{\chi_p(x_{t-1};y_{t-1})}=\sqrt{\frac{v_t}{(1-\rho^2)\sigma^2}}\exp\left(\frac{1}{2}(\frac{\mu_t^2}{v_t}-\frac{(\mu(1-\phi)+\phi h_{t-1}+\rho\sigma\bar{\epsilon}_{t-1})^2}{(1-\rho^2)\sigma^2})\right)\times \frac{\Gamma(\zeta/2+r_t)}{(\zeta/2+r_t)^{\zeta/2+s_t}}\frac{(\zeta/2)^{\zeta/2}}{\Gamma(\zeta/2)},$$

where

$$v_t=\frac{(1-\rho^2)\sigma^2}{1+(1-\rho^2)\sigma^2c_t},\quad\text{and}\quad\mu_t=v_t\left(b_t+\frac{\mu(1-\phi)+\phi h_{t-1}+\rho\sigma\bar{\epsilon}_{t-1}}{(1-\rho^2)\sigma^2}\right). \tag{7}$$

The choice of kernel (6) corresponds to an importance density that is the product of a normal density with mean μ_t and variance v_t as defined in (7) and an inverse Gamma density with shape parameter $\zeta/2+s_t$ and rate parameter $\zeta/2+r_t$, i.e.

$$q(x_t|x_{t-1},y_{1:T})=N(h_t;\mu_t,v_t)\cdot IG\left(W_t;\frac{\zeta}{2}+s_t,\frac{\zeta}{2}+r_t\right). \tag{8}$$

Kleppe and Liesenfeld (2014) originally considered this formulation in a continuous mixture setting.

We obtain the set of importance parameters δ_t iteratively via a sequence of auxiliary least square regressions. Briefly, given an initial set of importance parameters $\delta_t^{(n)}$, we can draw J trajectories of $x_t^{(j)}=(h_t^{(j)},W_t^{(j)})'$ for $j=1,...,J$ using (8). For each t, we update $\delta_t^{(n+1)}$ such that

$$\delta_t^{(n+1)}=\arg\min_{\delta_t}\sum_{j=1}^{J}\left[\left(\log p(y_t|x_t^{(j)})+\log\frac{\chi_q(x_t^{(j)};\delta_{t+1}^{(n+1)})}{\chi_p(x_{t-1}^{(j)};y_{t-1})}\right)-\left(\gamma_t+\log k(x_t^{(j)},x_{t-1}^{(j)};\delta_t)\right)\right]^2, \tag{9}$$

where γ_t is a normalizing constant. In essence, EIS finds an approximate minimiser δ_t for the

variance of the ratio or the importance weight

$$\frac{p(y_t|x_t)\frac{\chi_q(x_t;\delta_{t+1})}{\chi_p(x_{t-1};y_{t-1})}}{k(x_t,x_{t-1};\delta_t)} = \frac{p(y_t|x_t)p(x_t|x_{t-1},y_{t-1})}{k_q(x_t,x_{t-1};\delta_t)/\chi_q(x_t;\delta_{t+1})}. \tag{10}$$

Because exponential family kernels such as (6) are log-linear, the regression is a basic OLS with regressors $(h_t^{(j)}, {h_t^{(j)}}^2, -\log W_t^{(j)}, -1/W_t^{(j)})$. As shown by Richard and Zhang (2007) and Scharth and Kohn (2016), the backward-shift of the period $t+1$ integration constant $\chi_q(x_t;\delta_{t+1})$ is crucial for obtaining a globally efficient importance density as it depends on both the lagged and future states.

2.2.2 Sampling $h_{1:T}, W_{1:T}|y_{1:T},\theta$

We sample from $p(h_{1:T}, W_{1:T}|y_{1:T})$ by using the EIS state proposal within the particle Gibbs with ancestor sampling (PGAS) algorithm of Lindsten et al. (2014). We abbreviate this sampler as PGAS-EIS, as it belongs to the family of particle Gibbs algorithms (see e.g., Chopin et al., 2013) based on particle filtering or sequential Monte Carlo methods (see Pitt and Shephard, 1999b and Doucet et al., 2001 for a general discussion). Assuming at $t-1$, we have a particle system containing M particles $\{x_{1:t-1}^i\}_{i=1}^M$ and associated weights $\{\omega_{t-1}^i\}_{i=1}^M$ which approximates the filtering density $p(x_{1:t-1}|y_{1:t-1})$ by a sum of Dirac delta functions $\mathcal{D}(.)$, we have

$$\hat{p}(x_{1:t-1}|y_{1:t-1}) = \sum_{i=1}^{M} \frac{\omega_{t-1}^i}{\sum_{j=1}^M \omega_{t-1}^j} \mathcal{D}(x_{1:t-1} - x_{1:t-1}^i).$$

PGAS-EIS propagates the particle system by first sampling $\{a_t^i, x_t\}_{i=1}^M$ from

$$\mathcal{I}_t(a_t, x_t) = \frac{\omega_{t-1}^{a_t}}{\sum_{j=1}^M \omega_{t-1}^j} q(x_t|x_{t-1}^{a_t}, y_{1:T}),$$

with a_t indexing the ancestor particle, i.e. $x_{1:t}^i = (x_{1:t-1}^{a_t^i}, x_t^i)$. PGAS-EIS differs from a standard PG algorithm because it then samples the ancestor for the reference trajectory according to

$$\mathbb{P}(a_t^{M+1} = i) = \frac{\omega_{t-1}^i p(x_t^\star|x_{t-1}^i, y_{t-1})}{\sum_{j=1}^{M+1} \omega_{t-1}^j p(x_t^\star|x_{t-1}^j, y_{t-1})}, \tag{11}$$

and then we "rewrite" the history of the reference trajectory by setting $x_{1:t}^{M+1} = (x_{1:t-1}^{a_t^{M+1}}, x_t^{M+1})$.

We finish the recursion for each t by re-weighting the augmented system according to

$$\omega_t^i = \frac{p(y_t|x_t^i)p(x_t^i|x_{t-1}^i, y_{t-1})}{q(x_t^i|x_{t-1}^i, y_{1:T})}, \qquad \text{for } i = 1, ..., M+1. \tag{12}$$

Once the propagation reaches $t = T$, PGAS-EIS samples a new path $x_{1:T}^+$ from

$$\hat{p}(x_{1:T}|y_{1:T}) = \sum_{i=1}^{M+1} \frac{\omega_T^i}{\sum_{j=1}^{M+1} \omega_T^j} \mathcal{D}(x_{1:T} - x_{1:T}^i), \tag{13}$$

which serves as the reference trajectory $x_{1:T}^\star$ in the next MCMC iteration.

As shown in Lindsten et al. (2014), the AS step effectively breaks the reference trajectory into pieces, and as a result $x_{1:T}^+$ is substantially different from $x_{1:T}^\star$ with high probability. This improves mixing compared to standard PG. Our sampler further improves efficiency through the globally optimal EIS importance density. From Chopin et al. (2013) and Lindsten et al. (2014), it follows immediately that the Markov kernel implied by the PGAS-EIS sampler leaves the conditional posterior distribution $p(x_{1:T}|y_{1:T})$ invariant. Furthermore, assuming a boundedness condition for the importance weight ω_t^i for all i and t to hold, we can also show uniform ergodicity of the implied Markov kernel, which is an essential requirement for sampling the parameter vector θ.

2.2.3 Sampling $\theta|y_{1:T}, h_{1:T}, W_{1:T}$

We here focus on σ and ρ only. We sample the remaining parameters according to Nakajima and Omori (2012); See the appendix for details. Let $\pi_0(\cdot)$ and $\pi(|\cdot)$ denote prior and posterior distributions respectively. The joint posterior probability distribution $\pi(\sigma, \rho|\cdot)$ is

$$\pi(\sigma, \rho|\cdot) \propto \pi_0(\sigma, \rho)\sigma^{-T}(1-\rho^2)^{-\frac{T-1}{2}} \exp\left\{-\frac{(1-\phi^2)\bar{h}_1^2}{2\sigma^2} - \sum_{t=1}^{T-1} \frac{(\bar{h}_{t+1} - \phi\bar{h}_t - \rho\sigma\bar{\epsilon}_t)^2}{2\sigma^2(1-\rho^2)}\right\},$$

where $\bar{h}_t = h_t - \mu$.

We re-parameterize the likelihood in the above expression by $\vartheta = \rho\sigma$ and $\varpi = \sigma^2(1-\rho^2)$. By factorizing the joint prior as $\pi_0(\vartheta|\varpi)\pi_0(\varpi)$ and choosing $\pi_0(\varpi) = \text{IG}(s_0, r_0)$ and $\pi_0(\vartheta|\varpi) = \text{N}(\vartheta_0, v_\vartheta^2\varpi)$, i.e. a normal-inverse-gamma conjugate prior, we can efficiently generate new draws

from $\vartheta|\cdot \sim \mathrm{N}(\mu_\vartheta, \sigma_\vartheta^2\varpi)$, where $\varpi|\cdot \sim IG(r_1, s_1)$ and

$$\begin{aligned} \sigma_\vartheta^2 &= \left\{ \frac{1}{v_\vartheta^2} + \sum_{t=1}^{T-1} \bar{\epsilon}_t^2 \right\}^{-1}, \quad \mu_\vartheta = \sigma_\vartheta^2 \left\{ \frac{\vartheta_0}{v_\vartheta^2} + \sum_{t=1}^{T-1} \bar{\epsilon}_t(\bar{h}_{t+1} - \phi\bar{h}_t) \right\}, \\ s_1 &= s_0 + \frac{T}{2}, \qquad\qquad r_1 = r_0 + \frac{1}{2}\left\{ \sum_{t=1}^{T-1} (\bar{h}_{t+1} - \phi\bar{h}_t)^2 - \frac{\mu_\vartheta^2}{\sigma_\vartheta^2} + \frac{\vartheta_0^2}{v_\vartheta^2} \right\}. \end{aligned} \tag{14}$$

We update the Markov chain accordingly using $\sigma = \sqrt{\vartheta^2 + \varpi}$ and $\rho = \vartheta/\sigma$.

3 Factor SV with leverage effects, skewness, and heavy tails

3.1 Model specification

We propose the following factor SV model

$$\begin{aligned} &y_t = \Lambda f_t + u_t, \quad t = 1, ..., T, \\ &\{f_{j,t}\}_{t=1}^T \sim \text{Model (1)}, \quad \forall j \in \{1, ..., p\}, \\ &\{u_{i,t}\}_{t=1}^T \sim \text{Model (1)}, \quad \forall i \in \{1, ..., n\}, \end{aligned} \tag{15}$$

where $y_t \in \mathbb{R}^n$, $f_t \in \mathbb{R}^p$ and $\Lambda \in \mathbb{R}^{n\times p}$. Model (15) tells that each factor process $f_{j,t}$ and asset-specific process $u_{i,t}$ follow the univariate dynamics as in the univariate model (1). The above model is related to the specifications in Chib et al. (2006) and Nakajima (2015), but is considerably more flexible than these two. Chib et al. (2006) consider a factor structure, but do not allow for possible leverage, heavy-tailedness and skewness in the factors. Nakajima (2015) incorporate model (1), but do not achieve dimension reduction through a factor structure in asset returns.

Let $h_{j,t}$ and $W_{j,t}$ denote the SV process and the inverse gamma mixing variable for the factor process $f_{j,t}$, $j = 1, ..., p$. Let $l_{i,t}$ and $Q_{i,t}$ denote those for the asset-specific process $u_{i,t}$, $i = 1, ..., n$. We use superscripts f_j and u_i for the parameters associated with these processes. That is, for the i-th return series the model reads

$$\begin{aligned} y_{i,t} = &\sum_{j=1}^{p} \Lambda_{ij}(\alpha^{f_j} + \beta^{f_j} W_{j,t}\gamma^{f_j} + \sqrt{W_{j,t}}\gamma^{f_j}\xi_{j,t}) e^{h_{j,t}/2} \\ &+ (\alpha^{u_i} + \beta^{u_i} Q_{i,t}\gamma^{u_i} + \sqrt{Q_{i,t}}\gamma^{u_i}\epsilon_{i,t}) e^{l_{i,t}/2}. \end{aligned} \tag{16}$$

Let h and l denote the set of SV processes corresponding to f_t and u_t respectively for

$t = 1, ..., T$; namely $h = \{h_1, ..., h_p\}$ where $h_j = \{h_{j,t}\}_{t=1}^T$ for $j = 1, ..., p$, and $l = \{l_1, ..., l_n\}$ where $l_i = \{l_{i,t}\}_{t=1}^T$ for $i = 1, ..., n$. We denote the set of mixture components by W and Q in a similar fashion. A model with n assets and p factors has $6(n+p) + np - (p^2+p)/2$ parameters with usual identification restrictions imposed on the factor loadings Λ.

3.2 Selection priors for leverage and skewness

The general formulation (15) allows for leverage effects and skewness across all factor and asset-specific components in a non-discriminatory fashion, which may lead to excessive flexibility. At the same time, there is no solid empirical evidence to guide us the specification leverage effects and skewness for particular model components. For example, are leverage effects are systematic, idiosyncratic, or both? This is an important and unanswered question that is amenable to investigation under our flexible multivariate framework.

We adopt selection priors to reduce model complexity and improve interpretability. As in Section 2.2.3, consider the convenient re-parametrization $\vartheta_k = \rho_k \sigma_k$ and $\varpi_k = \sigma_k^2(1-\rho_k^2)$, $k = 1, ..., n+p$. We modify the normal-inverse-gamma conjugate priors for (ϑ_k, ϖ_k) with sparsity priors similar to the Bayesian variable selection approach of Clyde and George (2004). The reparametrisation equips $\vartheta_k|\varpi_k$ with a normal prior as in Section 2.2.3,, so that $\pi_0^{\text{sparse}}(\vartheta_k|\varpi_k)$ takes the form

$$\Delta_\vartheta \mathcal{D}_0(\vartheta_k) + (1-\Delta_\vartheta) N(\vartheta_0, v_\vartheta^2 \varpi_k),$$

where $\mathcal{D}_0(.)$ denotes the Dirac delta function at zero, and $\mathrm{N}(\vartheta_0, v_\vartheta^2\varpi_k)$ is the (conditional) conjugate normal prior. This prior means that ϑ_k has Δ_ϑ point mass at zero and probability $1-\Delta_\vartheta$ of taking a value that follows the $N(\vartheta_0, v_\vartheta^2\varpi_k)$ distribution.

3.3 MCMC algorithm

As in Chib et al. (2006), the factor SV model (15) is a linear combination of $n+p$ univariate SV models specified as (1) due to independence structure conditional on factor process $\{f_t\}_{t=1}^T$ and loadings Λ. This model structure allows us to design an efficient MCMC algorithm that iterates over

1. sampling the factor loadings Λ by marginalizing the factors;

2. sampling the factors;

3. sampling the latent trajectories h, l, W, Q and associated parameters using the univariate method of Section 2.2,

as a simplified outline. In the third and most computationally intensive step, we separately work with $n+p$ separate univariate SV series, making our framework scalable and amenable to parallelization.

3.3.1 Sampling the factor loadings Λ

Factor models generally require identification on the factor loadings; for example, one can restrict the upper p-by-p sub-matrix of Λ to be lower triangular with ones on the diagonal. Due to the presence of the multiplicative term Λf_t in the likelihood, drawing Λ and f_t conditional on each other is likely to be inefficient. We follow Chib et al. (2006) and achieve efficiency by marginalizing f_t via a straightforward extension of the corresponding step in their algorithm to model (15).

We sample the matrix of factor loadings Λ conditional on the data, all the latent log-volatility trajectories $\{h_1, \ldots, h_p\}$ and $\{l_1, \ldots, l_n\}$, all the latent inverse Gamma series $\{W_1, \ldots, W_p\}$ and $\{Q_1, \ldots, Q_n\}$, and the parameters associated with all state variables. Define

$$\begin{aligned}
\tilde{y}_{i,t} &= y_{i,t} - (\alpha^{u_i} + \beta^{u_i} Q_{i,t} \gamma^{u_i}) e^{l_{i,t}/2}, \\
F_{j,t} &= (\alpha^{f_j} + \beta^{f_j} W_{j,t} \gamma^{f_j}) e^{h_{j,t}/2}, \\
V_t &= \mathrm{diag}(W_{1,t}(\gamma^{f_1})^2 e^{h_{1,t}}, ..., W_{p,t}(\gamma^{f_p})^2 e^{h_{p,t}}), \\
U_t &= \mathrm{diag}(Q_{1,t}(\gamma^{u_1})^2 e^{l_{1,t}}, ..., Q_{n,t}(\gamma^{u_n})^2 e^{l_{n,t}}),
\end{aligned} \tag{17}$$

$\tilde{y}_t = (\tilde{y}_{1,t}, ..., \tilde{y}_{n,t})'$, $F_t = (F_{1,t}, ..., F_{p,t})'$.

Given h, l, W and Q, the conditional log-likelihood function is

$$\begin{aligned}
l(y_{1:T}|\Lambda) &= \sum_{t=1}^{T} l_t(y_t|\Lambda) = \log N(\tilde{y}_t; \Lambda F_t, \Omega_t) \\
&= -\frac{1}{2} \sum_{t=1}^{T} \left\{ k \log 2\pi + \log|\Omega_t| + (\tilde{y}_t - \Lambda F_t)' \Omega^{-1} (\tilde{y}_t - \Lambda F_t) \right\},
\end{aligned} \tag{18}$$

where $\Omega_t = \Lambda V_t \Lambda' + U_t$. We apply the MH step of Chib et al. (2006) to sample $\mathrm{vec}(\Lambda)|\cdot$ using a multivariate Student's t-proposal density $\mathrm{T}(\mu_\Lambda, \Sigma_\Lambda, v)$ where μ_Λ is the mode of $l(y_{1:T}|\Lambda)$ and Σ_Λ equals minus the inverse of the approximate Hessian matrix of $l(y_{1:T}|\Lambda)$ around its mode. See the appendix for further details.

3.3.2 Sampling the factors f_t

Given the factor loadings Λ, we have that

$$\tilde{y}_t | f_t \sim N(\Lambda f_t, U_t), \qquad f_t \sim N(F_t, V_t),$$

where simplify the notation by suppressing the dependence on h, l, W and Q, as well as on all parameters.

Standard results for the multivariate Gaussian distribution imply that $f_t|\cdot \sim N(\mu_{f_t}, \Sigma_{f_t})$ where

$$\Sigma_{f_t} = (\Lambda' U_t^{-1} \Lambda + V_t^{-1})^{-1}, \qquad \mu_{f_t} = \Sigma_{f_t}(\Lambda' U_t^{-1} \tilde{y}_t + V_t^{-1} F_t),$$

3.3.3 Sampling h, l, W, Q and associated parameters

Given the previous step, we obtain p factor series $\{f_{j,t}\}_{t=1}^T$, $j = 1, \ldots, p$, and n idiosyncratic series $\{u_{i,t}\}_{t=1}^T$, $i = 1, \ldots, n$, all of which follow the univariate model (1). We use the univariate procedure of Section 2.2 for each of these components, except for a modification to account for the selection priors of Section 3.2.

The conditional distribution for ϑ_k as defined in 3.2 is given by

$$\vartheta_k|\cdot \sim \Delta_{\vartheta_k} \mathcal{D}_0(\vartheta_k) + (1 - \Delta_{\vartheta_k}) N(\mu_{\vartheta_k}, \sigma^2_{\vartheta_k} \varpi_k),$$

where μ_{ϑ_k} and $\sigma^2_{\vartheta_k}$ are defined in (14), and where

$$\Delta_{\vartheta_k} = \frac{1 - \Delta_\vartheta}{\Delta_\vartheta \tilde{\sigma}^2_{\vartheta_k} + 1 - \Delta_\vartheta}, \quad \text{with } \tilde{\sigma}^2_{\vartheta_k} = \frac{\sigma_{\vartheta_k}}{v_\vartheta} \exp\left(\frac{\mu^2_{\vartheta_k}}{2\sigma^2_{\vartheta_k} \varpi_k} \right).$$

The probability of a zero value Δ_ϑ has a beta conjugate prior, so we that can generate posterior draws from the number of non-zero ϑ_k's (i.e. non-zero ρ_k's) in the Markov chain. We treat the skewness parameter β_k in a similar way.

3.3.4 Initialization

The initialization of the Markov chain is important when the number of assets is large. Our experience has shown that careful initialization can save hours of computation time and accelerate the convergence of the Markov chain to its stationary distribution. See the appendix for our initialization procedure.

4 Simulation study

We implement univariate and multivariate simulation studies to investigate the efficiency of the MCMC algorithms of Sections 2.2 and 3.3. We highlight the contribution of the PGAS-EIS algorithm for sampling parameters and latent processes in comparison with the method developed by Nakajima and Omori (2012), and show that all the components of the methodology are fundamental for accurate estimation, relative to simpler implementations. In the factor SV setting, we also study the usefulness of the selection priors for leverage effect and skewness introduced in Section 3.2.

4.1 Univariate estimation

4.1.1 Simulation design

We simulate 500 series of length $T = 2000$ from model (1) with parameter values $\phi = 0.95$, $\sigma = 0.15$, $\rho = -0.5$, $\mu = -9$, $\beta = -0.5$, and $\zeta = 20$. We specify the following prior distributions:

$$\begin{aligned} &\frac{\phi+1}{2} \sim \text{Beta}(20, 1.5), \quad \varpi \sim \text{IG}(2.5, 0.025), \quad \vartheta|\varpi \sim \text{N}(0, 20\varpi), \\ &\mu \sim \text{N}(-10, 5), \qquad \beta \sim \text{N}(0, 1), \qquad \zeta \sim \text{Gamma}(20, 1.25)\mathbb{1}_{(\zeta>4)}, \end{aligned} \tag{19}$$

where $\varpi = (1-\rho^2)\sigma^2$, $\vartheta = \rho\sigma$ and where $\mathbb{1}_{(.)}$ is an indicator function.

The above prior distributions reflect popular choices in the literature of SV models. Table 1 shows alternative algorithms for comparison with the PGAS-EIS sampler:

Table 1: Sampling methods for the univariate SV model

Acronym	Algorithm	No. of particles
MM-MH	Sampler proposed by Nakajima and Omori (2012) – multi-move sampler of Pitt and Shephard (1999b) for h_t, conditional on which W_t is sampled via accept-reject Metropolis-Hastings algorithm	500
PG-EIS	Particle Gibbs with EIS importance density	10
PGAS-BF	Particle Gibbs with ancestor sampling using the bootstrap filter	2000

Due to the efficiency of the EIS proposal, we use only 10 particles for the PGAS-EIS and PG-EIS samplers . The PGAS-BF method needs to rely on a much larger number of particles since it is based on the bootstrap filter, which simply uses the state transition density as a proposal. We run 22000 Markov Chain iterations with a burn-in period of 2000. We base the comparisons on the inefficiency factor $IE(\theta) = 1 + 2\sum_{j=1}^{\infty} \rho_j(\theta)$ for a certain parameter θ, where

Table 2: Simulation results for the the univariate SV model

	PGAS-EIS				MM-MH			
θ	Mean	RMSE	95% CI.	IE(θ)	Mean	RMSE	95% CI	IE(θ)
ϕ	0.955	0.022	[0.939 0.978]	4.526	0.942	0.062	[0.921 0.973]	81.532
σ	0.151	0.012	[0.148 0.172]	11.046	0.167	0.048	[0.139 0.202]	159.406
ρ	-0.512	0.073	[-0.602 -0.415]	23.218	-0.542	0.112	[-0.731 -0.444]	79.157
μ	-8.979	0.411	[-9.081 -8.755]	4.326	-8.920	0.869	[-9.124 -8.795]	27.582
β	-0.573	0.186	[-0.790 -0.400]	16.027	-0.714	0.415	[-1.346 -0.318]	163.733
ζ	18.963	5.717	[16.462 26.348]	36.356	21.655	8.619	[16.067 37.881]	299.057
	PG-EIS				PGAS-BF			
θ	Mean	RMSE	95% CI	IE(θ)	Mean	RMSE	95% C.I.	IE(θ)
ϕ	0.965	0.083	[0.942 0.975]	64.057	0.872	0.0.180	[0.764 0.992]	16.746
σ	0.162	0.025	[0.154 0.187]	132.744	0.237	0.180	[0.167 0.304]	73.569
ρ	-0.522	0.129	[-0.712 -0.424]	92.246	-0.204	0.705	[-0.421 0.086]	52.74
μ	-9.343	0.651	[-9.547 -8.834]	15.321	-10.657	2.224	[-11.050 -8.891]	24.315
β	-0.688	0.286	[-0.842 -0.460]	93.682	-0.137	0.727	[-1.143 0.722]	51.985
ζ	22.462	8.482	[16.785 30.114]	123.37	32.864	18.196	[19.674 65.049]	96.781

[1] We report the average posterior mean (across the simulation replications), root mean squared error, average lower and upper bounds of 95% credible interval (CI), and average inefficiency factors ($IE(\theta)$).
[2] True DGP: $\phi = 0.95$, $\sigma = 0.15$, $\rho = -0.5$, $\mu = -9$, $\beta = -0.5$ and $\zeta = 20$.

$\rho_j(\theta)$ is the j-th sample autocorrelation. We choose Parzen window with bandwidth 1000 for computing the inefficiency factors.

4.1.2 Results

Table 2 reports the estimated posterior means together other statistics measuring accuracy and stability under the four sampling methods, averaged across all simulated series. The results show that the the PGAS-EIS sampler outperforms the alternative methods. The average posterior mean is close to the true DGP values for the PGAS-EIS, PG-EIS, MM-MH methods and closest for the PGAS-EIS sampler; the latter is also the stablest sampler according to the root mean squared error criterion. The inefficiency factors for the PGAS-EIS sampler range from low to moderate, and are substantially lower than those from other samplers. The comparison between PGAS-EIS with PG-EIS suggests that the ancestor sampling step employed in particle Gibbs algorithm is the main driver of efficiency relative to the MM-MH sampler. However, note the PGAS method requires a proposal for the state trajectory, which is enabled by the EIS method. The poor performance of the PGAS-BF sampler suggests that a tailored proposal such as in the EIS method is crucial for reliable estimation.

4.2 Factor SV simulation

4.2.1 Simulation design

We simulate the factor SV model 15 with $n = 50$ assets and $p = 8$ factors as in Chib et al. (2006). We consider a range of scenarios for the presence of leverage effects and skewness in the model components, as outlined in Table 3.

Table 3: Different datasets simulated from the factor SV model

Acronym	Data generating process
sLE_sSK	For some $k \in \{1, ..., n+p\}$, $\rho_k \neq 0$ and $\beta_k \neq 0$. That is, leverage effect and return asymmetry are present in some processes among factors $\{f_{j,t}\}_{j=1}^{p}$ and asset-specific components $\{u_{i,t}\}_{i=1}^{n}$.
sLE_aSK	For some k, $\rho_k \neq 0$; and for all k, $\beta_k \neq 0$.
aLE_sSK	For all k, $\rho_k \neq 0$; and for some k, $\beta_k \neq 0$.
aLE_aSK	For all k, $\rho_k \neq 0$ and $\beta_k \neq 0$.
nLE_nSK	For all k, $\rho_k = 0$ and $\beta_k = 0$.

Our empirical results in the next section suggest that the scenario in which only a subset of factors f_t and asset-specific processes u_t display leverage effect or skewness (sLE_sSK) is the most relevant one. When a dataset has leverage effects or skewness, we generate an $n+p$-dimensional vector from the binomial distribution with 0.5 probability of success indicating which series among $\{f_{j,t}\}_{j=1}^{p}$ and $\{u_{i,t}\}_{i=1}^{n}$ have leverage effect or skewness. We choose the Beta$(2, 2)$ prior for the probability parameters in the selection priors of Section 3.2. We assume a flat $N(0, 10)$ prior for each free element of Λ, and generate the true loadings from the $N(1, 1)$ distribution. We generate the other parameters from the prior distributions given in (19), except that we only allow for zero or negative skewness.

4.2.2 Results

Table 4 summarizes the correlation between the posterior means of all parameters and their DGP values under different samplers and scenarios, a well as the mean absolute deviations in brackets. The PGAS-EIS method achieves the highest correlations and lowest mean absolute deviations, with only two correlations below 0.9 and none smaller than 0.8, followed by PG-EIS. This relative outperformance is especially pronounced for the degrees of freedom parameter ζ, which seems to be hardest parameter to estimate across all methods and datasets. As in the univariate setting, the the PGAS-BF does not lead to reliable estimation, displaying much lower correlations much and higher mean absolute deviations much higher than the other three

Table 4: Accuracy comparisons of different methods under different datasets

	ϕ	σ	ρ	μ	β	ζ	Λ
				PGAS-EIS			
sLE_sSK	.94 [.01]	.94 [.01]	.98 [.07]	.95 [.21]	.98 [.28]	.85 [3.89]	.99 [.11]
sLE_aSK	.94 [.01]	.92 [.01]	.96 [.04]	.95 [.18]	.99 [.35]	.91 [5.34]	.99 [.13]
aLE_sSK	.95 [.02]	.97 [.02]	.99 [.08]	.97 [.28]	.97 [.24]	.88 [3.07]	.98 [.09]
aLE_aSK	.97 [.01]	.96 [.02]	.99 [.06]	.98 [.18]	.99 [.26]	.95 [4.41]	.99 [.12]
nLE_nSK	.96 [.01]	.91 [.02]	.95 [.01]	.97 [.29]	.92 [.08]	.82 [4.27]	.98 [.06]
				MM-MH			
sLE_sSK	.87 [.09]	.81 [.02]	.91 [.15]	.92 [.33]	.88 [.38]	.73 [9.01]	.97 [.24]
sLE_aSK	.84 [.11]	.96 [.02]	.97 [.07]	.95 [.84]	.92 [.67]	.67 [11.91]	.93 [.34]
aLE_sSK	.91 [.03]	.88 [.01]	.89 [.12]	.88 [1.11]	.93 [.56]	.81 [8.49]	.98 [.19]
aLE_aSK	.93 [.03]	.92 [.02]	.87 [.11]	.89 [.78]	.91 [.44]	.74 [5.76]	.97 [.40]
nLE_nSK	.93 [.16]	.90 [.02]	.97 [.03]	.93 [.27]	.95 [.12]	.83 [7.63]	.99 [.09]
				PG-EIS			
sLE_sSK	.93 [.01]	.92 [.02]	.99 [.08]	.94 [.27]	.97 [.22]	.88 [4.12]	.99 [.07]
sLE_aSK	.96 [.02]	.91 [.01]	.95 [.09]	.94 [.31]	.98 [.52]	.93 [8.69]	.99 [.03]
aLE_sSK	.91 [.04]	.94 [.02]	.94 [.10]	.96 [.29]	.94 [.32]	.84 [6.11]	.99 [.10]
aLE_aSK	.97 [.02]	.92 [.03]	.98 [.06]	.97 [.62]	.98 [.43]	.96 [5.57]	.99 [.07]
nLE_nSK	.97 [.01]	.96 [.01]	.94 [.03]	.93 [.46]	.95 [.11]	.79 [6.79]	.99 [.09]
				PGAS-BF			
sLE_sSK	.77 [.19]	.64 [.04]	.51 [.18]	.87 [.57]	.78 [.86]	.24 [15.04]	.86 [.98]
sLE_nSK	.82 [.14]	.77 [.06]	.62 [.26]	.84 [1.34]	.69 [.67]	.31 [9.60]	.87 [1.64]
nLE_sSK	.84 [.09]	.68 [.05]	.41 [.33]	.76 [.88]	.74 [.93]	.24 [11.24]	.79 [1.44]
nLE_nSK	.91 [.10]	.81 [.04]	.57 [.41]	.63 [.76]	.84 [.34]	.47 [10.07]	.85 [.67]
aLE_aSK	.84 [.12]	.83 [.02]	.45 [.43]	.81 [1.21]	.62 [.88]	.35 [8.65]	.89 [1.27]

The table reports the correlations between posterior means true DGP values under four sampling methods and different datasets. The sample mean of absolute deviations are in brackets.

methods.

Table 5 reports the median inefficiency factors under the scenario in which all components display leverage and skewness (aLE_aSK). The results for the other scenarios are similar. The efficiency is comparable to that for the univariate case in Table 2, highlighting the scalability of the factor approach. The PGAS-EIS achieves the smallest inefficiency factors. In particular, the results suggest that it greatly improves performance for the skew-t parameters β and ζ relative to the MM-MH produces $IE(\mu)$ that is six times larger than PGAS-EIS. Similar to the univariate case, ancestor sampling contributes to the efficiency of our MCMC sampler. PGAS-EIS is at least twice as efficient as PG-EIS, and for β it is more than five times more efficient. The four methods perform similarly for Λ, since they all marginalize the factors when sampling Λ in similar way.

Table 5: Inefficiency factors under leverage effects and skewness for all parameters

	PGAS-EIS	MM-MH	PG-EIS	PGAS-BF
ϕ	8.12 [4.97 14.86]	24.68 [18.94 29.52]	14.69 [10.64 19.48]	18.87 [14.31 22.49]
σ	21.33 [8.62 27.53]	124.20 [111.37 134.79]	78.54 [34.73 94.39]	64.61 [48.03 92.59]
ρ	22.74 [18.40 28.69]	107.56 [84.16 147.33]	84.57 [41.58 106.34]	87.53 [55.21 104.46]
μ	7.48 [5.97 9.06]	42.48 [37.64 48.76]	16.06 [10.55 23.74]	27.45 [14.62 44.78]
β	19.73 [14.82 26.45]	214.78 [149.60 307.43]	117.41 [89.65 134.80]	54.73 [39.06 82.36]
ζ	43.80 [31.69 67.81]	371.81 [256.14 504.26]	108.56 [86.17 134.89]	53.29 [45.88 67.21]
Λ	41.43 [24.74 51.12]	37.85 [27.84 48.57]	38.94 [32.50 57.93]	46.28 [37.66 63.18]

The table reports the median inefficiency factor with 10-th and 90-th percentiles in the bracket.

Table 6: Posterior zero probability of leverage effect and skewness

	$\min_k \mathbb{P}(\rho_k = 0)$ where $\rho_k^{DGP} = 0$				$\min_k \mathbb{P}(\beta_k = 0)$ where $\beta_k^{DGP} = 0$			
	PGAS-EIS	MM-MH	PG-EIS	PGAS-BF	PGAS-EIS	MM-MH	PG-EIS	PGAS-BF
sLE_sSK	0.91	0.86	0.93	0.56	0.95	0.92	0.95	0.81
sLE_aSK	0.94	0.87	0.90	0.66	–	–	–	–
aLE_sSK	–	–	–	–	0.95	0.94	0.94	0.76
nLE_nSK	0.94	0.92	0.90	0.79	0.98	0.96	0.97	0.91
	$\max_k \mathbb{P}(\rho_k = 0)$ where $\rho_k^{DGP} \neq 0$				$\max_k \mathbb{P}(\beta_k = 0)$ where $\beta_k^{DGP} \neq 0$			
sLE_sSK	0.77	0.86	0.76	0.92	0.52	0.74	0.61	0.90
sLE_aSK	0.71	0.74	0.81	0.88	0.21	0.22	0.18	0.64
aLE_sSK	0.44	0.51	0.46	0.71	0.57	0.68	0.60	0.86
aLE_aSK	0.37	0.52	0.41	0.68	0.18	0.21	0.19	0.58

The table reports thee lowest or highest posterior probability of leverage effect ρ_k and skewness parameter β_k being zero for $k \in A$ and $A \subset \{1, ..., n+p\}$ such that $\rho_k^{DGP} = 0$ (top left section), $\beta_k^{DGP} = 0$ (top right), $\rho_k^{DGP} \neq 0$ (bottom left) and $\beta_k^{DGP} \neq 0$. The superscript DGP indicates the corresponding true values. $\mathbb{P}(.)$ denotes a posterior probability.

Table 6 investigates the performance of the selection priors for skewness and leverage. The table reports the lowest posterior probability of a parameter being zero when associated DGP values are zero, and the highest probability when the DGP values are not zero. The PGAS-EIS, PG-EIS, and MM-MH methods produce reliable zero probability when the true parameters are zero, with the PGAS-EIS algorithm slightly outperforming the other two. When leverage effect and skewness are absent, the PGAS-EIS method consistently produces a posterior probability of above 0.9 for the parameter being zero, confirming the usefulness of the selection priors. At the same time, the estimated posterior can be conservative when leverage effects and skewness are present, regardless of the sampling algorithm.

Figure 1 shows correlations between the posterior means of selected factors, SV series with their inverse gamma mixture components and their DGP values. The PGAS-EIS sampler is the best performing method across all datasets, followed by PG-EIS in most cases. This suggests that ancestor sampling contributes to precision on top of the EIS proposal. The PGAS-BF method again suffers from inaccuracy. The correlations between many estimated latent processes and their DGP values are lower than 0.5 when bootstrap filter is used.

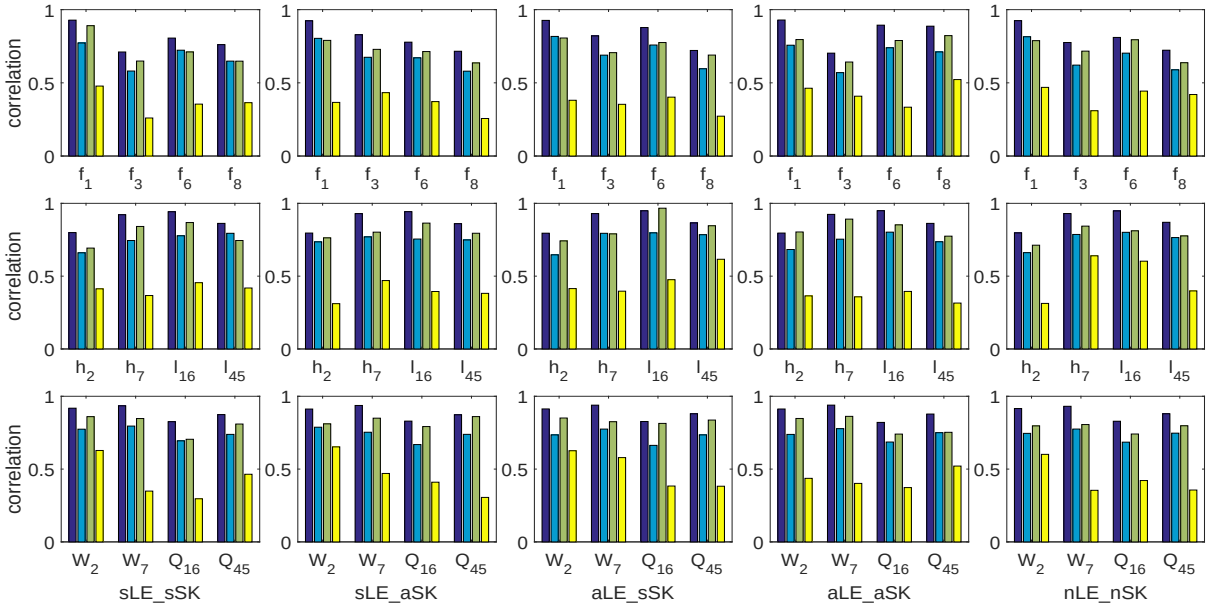

Figure 1: Correlations between posterior mean estimate of latent processes and DGP values. Purple: PGAS-EIS; Blue: MM-MH; Green: PG-EIS; Yellow: PGAS-BF; Top row: the 1st, 3rd, 6th and 8th factor; Middle row: the 2nd and 7th log-volatility process of factors (h), 16th and 45th log-volatility process of asset-specific processes (l); Bottom row: the 2nd and 7th inverse gamma mixing components of factors (W), 16th and 45th inverse gamma mixing components of asset-specific processes (Q); Columns from the left to the right indicate different simulated datasets.

5 Empirical study

We consider an empirical application to a sample of daily returns for 80 stock constituents of the S&P100 index. We exclude the other 20 stock components of the index due to their short trading history. Our sample period has a total of $T = 5663$ trading days from January 1996 to June 2018. Section 5.2 provides detailed estimation results for the high-dimensional factor SV model of Section 3.1 under two specifications of factor structure. Section 5.3 studies the forecasting performance of the two models in terms of evaluation of value-at-risk and predictive density score, including a comparison with other models.

5.1 Selecting the number of factors

Selecting the number of factors is an important and challenging problem for dynamic factor modeling. We perform model selection by designing a feasible marginal likelihood estimation procedure for the high-dimensional factor SV model (15), based on the importance sampling squared (IS^2) method of Tran et al. (2014); see the supplemental appendix for the description. This marginal likelihood criterion leads us to select four or five factors for two specifications discussed in this empirical application. As a comparison, this choice is consistent with the (non-Bayesian) IC_p1 and IC_p2 criteria of Bai and Ng (2002).

5.2 Estimation results

Because factor models as in (15) are identified up to scaling and rotation of factors, identification restrictions are usually imposed. We consider two specifications in this study. Let HFSV_1 and HFSV_2 denote model (15) with Λ restricted to be lower-triangular with ones on its diagonal and the one with the three Fama-French (FF) factors (Fama and French, 1993), i.e. $f_t = ((\text{Rm-Rf})_t, \text{SMB}_t, \text{HML}_t)'$ stacked onto the observation vector y_t and the upper-left block of Λ restricted to have identity matrix of dimension three, respectively. Notice HFSV_1 gives exact identification whereas HFSV_2 has overidentification for $p = 4$ and exact identification for $p = 5$; for $p \geq 6$, HFSV_2 requires further restrictions. HFSV_2 has "economically meaningful" factors that correspond to the seminal FF factors. The latter is also used by other factor models that we compare with in section 5.3.

The Bayes factor analysis based on the feasible marginal likelihood method developed for the proposed models (see the supplementary appendix) points to decisive evidence favoring 4 factors for HFSV_1. Although there is either strong or decisive evidence supporting HFSV_2 with 4 or 5 factors against other specifications, comparisons between these two are ambiguous. To be conservative, we choose 5 factors for $HFSV_2$.

Table 7 summarizes selected posterior mean and standard deviation (s.d.) for a selection of model parameters. The sample mean and sample s.d. (in parenthesis) of these posterior statistics across all factor and idiosyncratic components, for both HFSV_1 and HFSV_2, are reported. The mean of autoregressive parameter ϕ is close to unity with a small standard deviation under both specifications, suggesting persistent time-varying log-volatility for latent components. We find that only 5 assets with ϕ_i^l smaller than 0.9, and 21 assets smaller than 0.95 for HFSV_1. In general, estimation results of two specifications are similar, except for more heterogeneous d.o.f.

Table 7: Estimated posterior (summary)

	$HFSV_1$				$HFSV_2$			
posterior	mean		s.d.		mean		s.d.	
sample	mean	s.d.	mean	s.d.	mean	s.d.	mean	s.d.
ϕ	0.98	0.09	0.02	0.01	0.97	0.12	0.03	0.01
σ	0.20	0.10	0.03	0.01	0.22	0.13	0.03	0.01
μ	-10.61	1.45	0.33	0.09	-9.86	1.68	0.41	0.11
ζ	16.20	8.58	1.03	0.46	12.64	16.11	1.12	0.61

The table shows the sample mean and standard deviations of posterior mean and standard deviation estimates of selected parameters for both $HFSV_1$ and $HFSV_2$.

parameters from $HFSV_2$ as seen by the large sample s.d. of posterior means.

Table 8 summarizes the inefficiency factors (IE) for the Markov chain of parameters in terms of median (med), minimum (min), maximum (max) and, interquartile range (IQR) across parameters of all model components. We conclude that the MCMC algorithm of Section 3.3 is highly efficient for the proposed high-dimensional model, leading to moderate (often low) inefficiency factors despite the complex model structure of $HFSV_1$ and $HFSV_2$. In particular, we note that these inefficiency factors are comparable to those we obtained for the univariate SV model in the simulation study. Except for $IE(\sigma)$, parameters of $HFSV_2$ are estimated with slightly more efficient Markov chain than $HFSV_2$. The last five columns show the inefficiency factors for the loadings. Interestingly, under $HFSV_1$ compared to $IE(\Lambda_2)$ and $IE(\Lambda_4)$, the IQR of $IE(\Lambda_1)$ and $IE(\Lambda_3)$ are tighter due to many near-zero loadings on the 2-nd and 4-th factor. This is however not the case for $HSFV_2$, where we have comparably small IEs across all factor loadings, suggesting identified systematic content of the 4-th and 5-th factors on top of what is captured by the factors aligned with the FF factors; though $IE(\Lambda_4)$ and $IE(\Lambda_5)$ are clearly larger than those of factors identified using FF factors.

Figure 2 shows the posterior estimates for ρ, β and ζ, sorted in ascending order. Both $HFSV_1$ and $HFSV_2$ show that leverage effects and skewness or return asymmetry are associated with common factors, especially the latter. All assets share the non-zero leverage effects of some factors, while the third and fourth factor, from $HFSV_1$ and $HFSV_2$ respectively, contributes the most to the observed return asymmetry for individual returns. Nevertheless, asset-specific leverage effects are still present for many assets. From the right panels we find evidence supporting factor heavy-tailedness regardless of factor structure, which explains the stylized fact of tail events correlation, particularly with a negative factor skewness, downturn correlation among

Table 8: Inefficiency factors (summary)

HFSV_1	ϕ	σ	μ	ζ	Λ_1	Λ_2	Λ_3	Λ_4	Λ_5
med	11.3	27.8	6.4	67.2	14.9	45.2	33.6	69.7	-
min	5.7	22.4	3.8	32.9	8.9	25.5	21.7	42.6	-
max	21.3	52.5	13.4	84.2	19.6	87.0	50.4	108.7	-
IQR	13.7	30.6	8.1	45.2	6.3	51.9	18.7	57.8	-
HFSV_2									
med	8.1	31.2	4.6	58.5	10.1	11.7	8.6	34.6	27.8
min	2.3	14.6	2.2	24.0	6.2	5.3	4.8	18.9	19.5
max	12.4	61.4	14.7	109.6	13.6	16.1	17.2	88.2	51.7
IQR	6.5	31.3	3.8	43.8	5.0	8.3	8.3	28.7	22.6

The table summarizes the inefficiency factors for the posterior draws by averaging over the model components. Λ_j stands for the loadings on the j-th factor. IQR is the interquartile range.

assets.

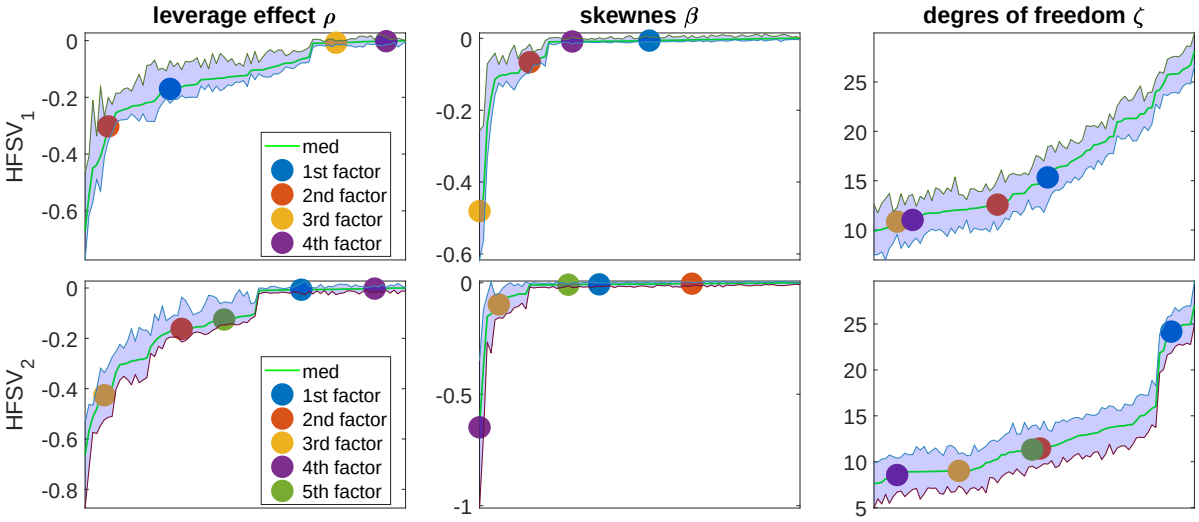

Figure 2: Posterior estimates of ρ, β and ζ. Graphs show posterior median with 10-th and 90-th percentile. From the left to the right are the estimates of leverage effect, return asymmetry (skewness) and, d.o.f. parameters of mixing components governing heavy-tailedness, rearranged in ascending order. Colored dots indicate estimates corresponding to factors. Top and bottom panels refer to HFSV_1 and HFSV_2, respectively.

Figure 3 illustrates estimates of some time series of interest from HFSV_1 and HFSV_2. The top-left panel shows the average factor log-volatility $\frac{1}{p}\sum_{i=1}^{p} h_{i,t}$ whereas the top-right panel shows the average asset-specific log-volatility $\frac{1}{n}\sum_{i=1}^{n} l_{i,t}$. Both specifications indicate similar evolution of factor and asset-specific volatilities. Interestingly, we find that the average of asset-specific volatilities explains on average over 70% of their variation (similar to the first principal component), which suggests strong second-order comovement on top of stochastic

volatility coming from market factors as in HFSV_2 or statistical factors as in HFSV_1. This result strengthens the findings in Campbell et al. (2001) and Herskovic et al. (2016) which also find commonality in idiosyncratic volatilities[1].

The middle panels show the role of factors in explaining return dynamics, ignoring the inverse gamma mixing component $W_{j,t}$, $j = 1, ..., p$ and, $Q_{i,t}$, $i = 1, ..., N$ given in (16). Denote asset i's volatility at time t as

$$\sigma_{i,t} = \left(\sum_{j=1}^{p} \Lambda_{ij}^2 \exp(h_{j,t}) + \exp(l_{i,t}) \right)^{\frac{1}{2}}, \quad i = 1, ..., n, \quad t = 1, ..., T.$$

The middle-left and -right panels show the average share of return explained by factors $\frac{1}{n} \sum_{i=1}^{n} Share_{i,t}$ and the average correlation $\frac{2}{n(n+1)} \sum_{i \neq j} Corr_{ij,t}$ among returns, respectively, where

$$Share_{i,t} = \frac{\sum_{j=1}^{p} \Lambda_{ij}^2 \exp(h_{j,t})}{\sigma_{i,t}^2}, \qquad Corr_{ij,t} = \frac{\sum_{k=1}^{p} \Lambda_{ik} \Lambda_{jk} \exp(h_{k,t})}{\sigma_{i,t} \sigma_{j,t}}. \tag{20}$$

It is clear that both average share and correlation peak during 2008 crisis, highlighting the systematic influence of this event. On the contrary, though both factor and asset-specific volatilities increase during 2000 'high-tech bubble', factors only explains 35% of return variation and average correlation dives to nearly zero in 2001. Ignoring mixing components amounts to treating factors and asset-specific components as Gaussian, which discards tail dependence. The bottom panels of Figure 3 show the counterparts of (20) with mixing components taken into account. Not surprisingly, mixing components in factors amplify the common volatility during market turmoil and thus explains more variations, leading to higher correlation among asset returns.

5.3 Value-at-risk

We now carry out a brief study on the performance of our model for value-at-risk (VaR) evaluation. We consider $S = 3600$ out-of-sample periods and estimate all competing models using a rolling window of $T = 2000$ days. We focus our comparison on alternative factor models proposed in the literature, since they are the most computationally viable approach for the dimensionality of our dataset.

[1]The authors find significant comovement among idiosyncratic volatilities after projecting asset returns onto the space spanned by market returns; however they do not model factor volatility explicitly. The presence of commonality in idiosyncratic volatilities on top of factor volatility has consequence on asset pricing and diversification, which we leave for future research.

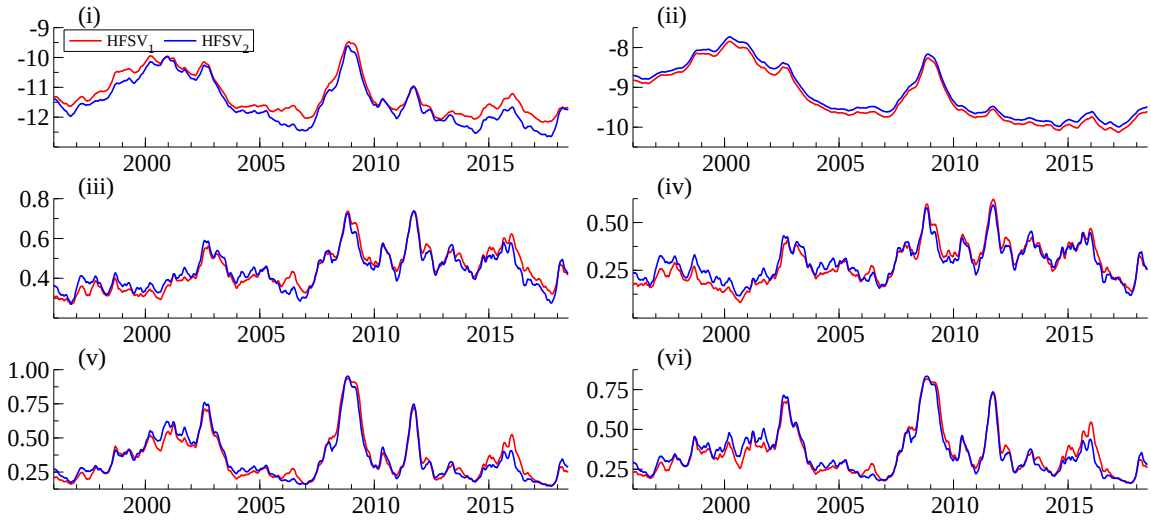

Figure 3: Posterior mean estimate of volatility and correlation trajectories. Left: SV associated with the 1st and 4th factor; Middle: SV associated with three randomly chosen assets; Right: Correlations.

5.3.1 Alternative models

The first model is the factor stochastic volatility model (FSV) of Chib et al. (2006), which is based on Gaussian SV models for the factors. The second model, also from Chib et al. (2006), augments the previous specification with stochastic jumps (FSV-J), $y_t = \Lambda f_t + K_t q_t + u_t$, where they model each factor $f_{j,t}$ as standard Gaussian SV model. The asset-specific $u_{i,t}$ process follows the Student's-t SV model. K_t records jump size at time t with Bernoulli jump variable $q_{i,t}$. We use the PGAS-EIS procedure for estimation. Subscripts 1 and 2 indicate factors identified via restricting the loading matrix or the FF factors, as in our proposed HFSV_1 and HFSV_2.

The third model, introduced by Chan et al. (1999) (CKL), is given by $y_t = \Lambda f_t + u_t$ with covariance matrix $\Omega_t = \Lambda V_t \Lambda' + U_t$. Here f_t is directly observable, e.g. the three FF factors. We compute the covariance matrix V_t from rolling-windows, and estimate covariance matrix U_t from the residuals of asset-by-asset regressions.

The fourth model is the dynamic factor multivariate GARCH (DFMG) model of Santos and Moura (2014), which is also based on observed factors. They specify $y_t = \Lambda_t f_t + u_t$ with covariance matrix $\Omega_t = \Lambda_t V_t \Lambda_t' + U_t$ and time-varying loadings $\lambda_{k,t+1} = \lambda_{k,t} + \eta_t$, where $\lambda_{k,t}$ is the k-th element of $\text{vec}(\Lambda_t)$, $k = 1, ..., p \times n$. V_t and U_t are diagonal matrices with each element evolving according to a standard GARCH dynamics. We find V_t and U_t by fitting GARCH to factors and regression residuals. Conditional on V_t and U_t, we obtain $\text{vec}(\Lambda_t)$ for all t via the Kalman filter.

The last specification is the factor copula (FCO and FCOt) model of Oh and Patton (2017b). Their model separates the modeling of the marginals (assets) and the dependence structure (factors). Refer to the original paper for a detailed description. We estimate GARCH/GARCH-t marginals for FCO/FCOt and a copula implied by Gaussian bi-factor model.

5.3.2 Value-at-risk

We consider an equally weighted portfolio and investigate the one-day and five-day ahead forecast of the portfolio VaR, i.e. $\text{VaR}_{t+h|t}$, $h = 1, 5$. The evaluation of VaR requires the simulation of predictive density. To this end, for each rolling-window $\mathfrak{M}$ draws of parameter vector θ are generated from their posterior samples, say $\hat{\theta}^{(i)}$, $i = 1, ..., \mathfrak{M}$; given $\hat{\theta}^{(i)}$, $\mathfrak{N}$ trajectories with length h of latent processes including stochastic volatility series and inverse-gamma mixing components are simulated, say $\hat{h}^{(j)}_{t+h}$, $\hat{l}^{(j)}_{t+h}$, $\hat{W}^{(j)}_{t+h}$ and, $\hat{Q}^{(j)}_{t+h}$, $j = 1, ..., \mathfrak{N}$; given each set of trajectories, $\mathfrak{K}$ predicted return vectors are simulated. This means that a forest of size $\mathfrak{M}\mathfrak{N}\mathfrak{K}$ of return vectors are simulated to construct the empirical predictive distribution of the portfolio, denoted by $\tilde{F}_{t+h|t}(.)$. We choose $\mathfrak{M} = \mathfrak{N} = \mathfrak{K} = 30$. h-step ahead VaR at α is given by $\text{VaR}_{t+h|t}(\alpha) = \tilde{F}^{-1}_{t+h|t}(\alpha)$, which is the 100α-th percentile of the predictive distribution function of the portfolio. We can readily approximate the predictive distribution of $y_{t+h|t}$ for the HFSV_i, FSV_i and FSV-J_i, $i = 1, 2$ based on particle system at time t (see also the appendix). For the other models, we derive the predictive densities in a similar fashion to GARCH-type models.

Define the following binary sequence $I_t(\alpha)$

$$I_t(\alpha) = \begin{cases} 1 & \text{if } \frac{1}{n}\sum_{i=1}^{n} y_{i,t+h} < \text{VaR}_{t+h|t} \\ 0 & \text{if } \frac{1}{n}\sum_{i=1}^{n} y_{i,t+h} \geq \text{VaR}_{t+h|t} \end{cases},$$

where $I_t(\alpha) = 1$ denote hits or violations. Well-behaved VaR estimates mean that the sequence $I_t(\alpha)$ is serially independent, i.e. $I_t(\alpha) \perp I_s(\alpha), \forall t \neq s$; and has correct unconditional coverage ratio, i.e. $\mathbb{P}[I_t(\alpha)] = \mathbb{E}[I_t(\alpha)] = \alpha$. Christoffersen (1998) develops likelihood ratio (LR) tests based on the violation process $I_t(\alpha)$ and hit rate (HR) $\frac{1}{T}\sum_{t=1}^{T} I_t(\alpha)$.

Table 9 reports the p-values of likelihood ratio test for unconditional coverage (LR_{uc}), independence (LR_{ind}) and, conditional coverage (based on the combined test statistic $LR_{cc} = LR_{uc} + LR_{ind}$). Shaded cells indicate a rejection at the 5% level. In general, the proposed HFSV_1 and HFSV_2 perform the best with the former marginally violating independence for $h = 5$ at $\alpha = 0.01$. Followed by the performance of FSV-J_1 and FSV-J_1, these models make

Table 9: Comparisons of VaR estimates

$h=1$	α=0.05				α=0.01			
	HR	LR_{uc}	LR_{ind}	LR_{cc}	HR	LR_{uc}	LR_{ind}	LR_{cc}
FSV_1	0.107	0.00	0.43	0.15	0.024	0.03	0.00	0.01
FSV_2	0.093	0.00	0.28	0.06	0.043	0.00	0.00	0.00
FSV-J_1	0.061	0.27	0.60	0.34	0.011	0.84	0.03	0.63
FSV-J_2	0.048	0.46	0.49	0.33	0.009	0.30	0.07	0.08
$HFSV_1$	0.047	0.35	0.77	0.56	0.012	0.47	0.08	0.18
$HFSV_2$	0.044	0.60	0.61	0.52	0.009	0.29	0.11	0.08
CKL	0.141	0.01	0.19	0.03	0.051	0.00	0.07	0.00
DFMG	0.086	0.01	0.42	0.14	0.018	0.08	0.14	0.02
FCO	0.080	0.03	0.04	0.00	0.074	0.00	0.00	0.00
FCOt	0.061	0.17	0.37	0.13	0.022	0.03	0.10	0.01
$h=5$								
FSV_1	0.112	0.00	0.44	0.16	0.014	0.18	0.04	0.03
FSV_2	0.127	0.00	0.34	0.09	0.044	0.01	0.21	0.03
FSV-J_1	0.075	0.04	0.02	0.00	0.013	0.26	0.67	0.41
FSV-J_2	0.044	0.08	0.01	0.01	0.008	0.07	0.42	0.15
$HFSV_1$	0.056	0.07	0.63	0.33	0.015	0.21	0.03	0.04
$HFSV_2$	0.048	0.14	0.48	0.20	0.011	0.61	0.06	0.31
CKL	0.091	0.00	0.71	0.43	0.052	0.00	0.01	0.00
DFMG	0.055	0.24	0.62	0.35	0.020	0.02	0.32	0.08
FCO	0.104	0.00	0.12	0.01	0.030	0.00	0.17	0.02
FCOt	0.051	0.61	0.84	0.74	0.028	0.01	0.27	0.06

The table shows p-values of coverage ratio tests for the equally weighted portfolio constructed using 80 stocks included in the S&P 100 index. α is the nominal level of VaR. Shaded cells indicate rejections in the coverage ratio tests at the 5% level. Daily ($h=1$) and weekly ($h=5$) forecasts are considered.

clear that VaR estimation depends on the modeling of mixing variables with sudden swings or jumps. This can also be seen if we compare FCOt which incorporates Student's t marginals with FCO. However, that $HFSV_1$ and $HFSV_2$ outperform the latter three provides evidence on the importance of modeling factors with richer features such as skewness and leverage effect. Failure to do so, as in the other competing models, lead to violations of unconditional coverage and independence, suggesting inaccuracy and misspecification. It is worth noting that models with subscript 1 performs very similarly to their counterpart with subscript 2 (see also Figure 3). So regardless of the rotation of factors, as long as econometricians treat them as latent processes and model them similarly, VaR forecasting performances are expected to be comparable. Lastly, we conjecture using observed factors directly without careful modeling may ignore important local information of market comovements and result in poor performance.

6 Conclusion

In this paper, we leveraged advances in Monte Carlo methods such as the efficient importance sampling (EIS) algorithm of Richard and Zhang (2007) and the particle Gibbs method of Andrieu et al. (2010) to design an MCMC algorithm that enables the estimation of a highly flexible factor stochastic volatility model in moderately high dimensions (up to 80 stocks in our empirical application). Our framework bridges contributions from authors such as Ang and Chen (2002) and Patton (2004), who have long documented that correlations between stocks are substantially higher for downside moves and during downturns, among other findings, and the literature on multivariate financial time series, which aims to address the curse of dimensionality and computational challenges towards the estimation of increasingly more accurate models in higher dimensions.

References

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

Andrieu, C., A. Doucet, and R. Holenstein (2010). Particle markov chain monte carlo methods. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 72*(3), 269–342.

Andrieu, C. and G. Roberts (2009). The pseudo-marginal approach for efficient monte carlo computations. *The Annals of Statistics 37*(2), 697–725.

Ang, A. and J. Chen (2002). Asymmetric correlations of equity portfolios. *Journal of financial Economics 63*(3), 443–494.

Asai, M., M. McAleer, and J. Yu (2006). Multivariate stochastic volatility: a review. *Econometric Reviews 25*(2-3), 145–175.

Bai, J. and S. Ng (2002). Determining the number of factors in approximate factor models. *Econometrica 70*(1), 191–221.

Beine, M., A. Cosma, and R. Vermeulen (2010). The dark side of global integration: Increasing tail dependence. *Journal of Banking & Finance 34*(1), 184–192.

Campbell, J. Y., M. Lettau, B. G. Malkiel, and Y. Xu (2001). Have individual stocks become more volatile? An empirical exploration of idiosyncratic risk. *The Journal of Finance 56*(1), 1–43.

Chan, J. C., R. Leon-Gonzales, and R. W. Strachan (2013). Invariant inference and efficient computation in the static factor model.

Chan, L. K., J. Karceski, and J. Lakonishok (1999). On portfolio optimization: Forecasting covariances and choosing the risk model. *Review of Financial Studies 12*(5), 937–974.

Chib, S., F. Nardari, and N. Shephard (2006). Analysis of high dimensional multivariate stochastic volatility models. *Journal of Econometrics 134*(2), 341–371.

Chib, S., Y. Omori, and M. Asai (2009). Multivariate stochastic volatility. In *Handbook of Financial Time Series*, pp. 365–400. Springer.

Chopin, N., S. S. Singh, et al. (2013). *On the particle Gibbs sampler*. CREST.

Christoffersen, P. F. (1998). Evaluating interval forecasts. *International economic review*, 841–862.

Clyde, M. and E. I. George (2004). Model uncertainty. *Statistical science*, 81–94.

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

Dempster, M. A. H., G. Pflug, and G. Mitra (2008). *Quantitative Fund Management*. Chapman and Hall/CRC.

Doucet, A., N. De Freitas, and N. Gordon (2001). An introduction to sequential Monte Carlo methods. In *Sequential Monte Carlo methods in practice*, pp. 3–14. Springer.

Doz, C., D. Giannone, and L. Reichlin (2011). A two-step estimator for large approximate dynamic factor models based on Kalman filtering. *Journal of Econometrics 164*(1), 188–205.

Durbin, J. and S. J. Koopman (1997). Monte Carlo maximum likelihood estimation for non-gaussian state space models. *Biometrika 84*(3), 669–684.

Durbin, J. and S. J. Koopman (2012). *Time series analysis by state space methods*. Number 38. Oxford University Press.

Fama, E. F. and K. R. French (1993). Common risk factors in the returns on stocks and bonds. *Journal of financial economics 33*(1), 3–56.

Geweke, J. and H. Tanizaki (2001). Bayesian estimation of state-space models using the Metropolis–Hastings algorithm within Gibbs sampling. *Computational Statistics & Data Analysis 37*(2), 151–170.

Gilks, W. R., N. Best, and K. Tan (1995). Adaptive rejection Metropolis sampling within Gibbs sampling. *Applied Statistics*, 455–472.

Grothe, O., T. S. Kleppe, and R. Liesenfeld (2017). The gibbs sampler with particle efficient importance sampling for state-space models.

Herskovic, B., B. Kelly, H. Lustig, and S. Van Nieuwerburgh (2016). The common factor in idiosyncratic volatility: Quantitative asset pricing implications. *Journal of Financial Economics 119*(2), 249–283.

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.

Kleppe, T. S. and R. Liesenfeld (2014). Efficient importance sampling in mixture frameworks. *Computational Statistics & Data Analysis 76*, 449–463.

Koop, G., D. J. Poirier, and J. L. Tobias (2007). *Bayesian econometric methods*. Cambridge University Press.

Koopman, S. J. and E. Hol Uspensky (2002). The stochastic volatility in mean model: empirical evidence from international stock markets. *Journal of applied Econometrics 17*(6), 667–689.

Liesenfeld, R. and J.-F. Richard (2006). Classical and bayesian analysis of univariate and multivariate stochastic volatility models. *Econometric Reviews 25*(2-3), 335–360.

Lindsten, F., M. I. Jordan, and T. B. Schön (2014). Particle gibbs with ancestor sampling. *Journal of Machine Learning Research 15*(1), 2145–2184.

Nakajima, J. (2015). Bayesian analysis of multivariate stochastic volatility with skew return distribution. *Econometric Reviews*, 1–23.

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.

Oh, D. H. and A. J. Patton (2017a). Modeling dependence in high dimensions with factor copulas. *Journal of Business & Economic Statistics 35*(1), 139–154.

Oh, D. H. and A. J. Patton (2017b). Modeling dependence in high dimensions with factor copulas. *Journal of Business & Economic Statistics 35*(1), 139–154.

Patton, A. J. (2004). On the out-of-sample importance of skewness and asymmetric dependence for asset allocation. *Journal of Financial Econometrics 2*(1), 130–168.

Pitt, M. and N. Shephard (1999a). Time varying covariances: a factor stochastic volatility approach. *Bayesian statistics 6*, 547–570.

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

Richard, J.-F. and W. Zhang (2007). Efficient high-dimensional importance sampling. *Journal of Econometrics 141*(2), 1385–1411.

Ruiz, E. (1994). Quasi-maximum likelihood estimation of stochastic volatility models. *Journal of econometrics 63*(1), 289–306.

Santos, A. A. and G. V. Moura (2014). Dynamic factor multivariate garch model. *Computational Statistics & Data Analysis 76*, 606–617.

Scharth, M. and R. Kohn (2016). Particle efficient importance sampling. *Journal of Econometrics 190*(1), 133–147.

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

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.

Tran, M.-N., M. Scharth, M. K. Pitt, and R. Kohn (2014). Importance sampling squared for bayesian inference in latent variable models. *Available at SSRN 2386371*.

Vardi, N. (2015). Top quant hedge funds stand out with good 2015.

Watanabe, T. and Y. Omori (2004). A multi-move sampler for estimating non-gaussian time series models: Comments on shephard & pitt (1997). *Biometrika*, 246–248.

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

Appendix

A MCMC algorithm details for the univariate process

A.1 Sampling the autoregressive coefficient ϕ

Given the remaining variables, the conditional posterior distribution for ϕ is

$$\begin{aligned}\pi(\phi|\cdot) &\propto \pi_0(\phi)\sqrt{1-\phi^2}\exp\left\{-\frac{(1-\phi^2)\bar{h}_1^2}{2\sigma^2}-\sum_{t=1}^{T-1}\frac{(\bar{h}_{t+1}-\phi\bar{h}_t-\rho\sigma\bar{\epsilon}_t)^2}{2\sigma^2(1-\rho^2)}\right\}\\ &\propto \pi_0(\phi)\sqrt{1-\phi^2}\exp\left\{-\frac{(\phi-\mu_\phi)^2}{2\sigma_\phi^2}\right\},\end{aligned}$$

where $\bar{h}_t = h_t - \mu$ and

$$\mu_\phi = \frac{\sum_{t=1}^{T-1}(\bar{h}_{t+1}-\sigma\rho\bar{\epsilon}_t)\bar{h}_t}{\rho^2\bar{h}_1^2+\sum_{t=2}^{T-1}\bar{h}_t^2}, \qquad \sigma_\phi^2 = \frac{\sigma^2(1-\rho^2)}{\rho^2\bar{h}_1^2+\sum_{t=2}^{T-1}\bar{h}_t^2}.$$

We employ the Metropolis-Hastings (MH) algorithm to sample from the above posterior. We draw a candidate ϕ^* from $N(\mu_\phi, \sigma_\phi^2)$ truncated within $(-1, 1)$ to ensure stationarity. We accept the proposed value with probability

$$\min\left\{\frac{\pi_0(\phi^*)\sqrt{1-\phi^{*2}}}{\pi_0(\phi)\sqrt{1-\phi^2}}, 1\right\}.$$

A.2 Sampling the unconditional mean μ of h_t.

Let the prior distribution of the unconditional mean μ be $N(\mu_0, v_\mu^2)$. The conditional posterior distribution is given by

$$\pi(\mu|\boldsymbol{\cdot}) \propto \exp\left\{-\frac{(\mu-\mu_0)^2}{2v_\mu^2}-\frac{(1-\phi^2)\bar{h}_1^2}{2\sigma^2}-\sum_{t=1}^{T-1}\frac{(\bar{h}_{t+1}-\phi\bar{h}_t-\rho\sigma\bar{\epsilon}_t)^2}{2\sigma^2(1-\rho^2)}\right\}.$$

We can generate a new draw $\mu|\boldsymbol{\cdot} \sim N(\mu_\mu, \sigma_\mu^2)$ with

$$\begin{aligned}\sigma_\mu^2 &= \left\{\frac{1}{v_\mu^2}+\frac{(1-\rho^2)(1-\phi^2)+(T-1)(1-\phi)^2}{\sigma^2(1-\rho^2)}\right\}^{-1},\\ \mu_\mu &= \sigma_\mu^2\left\{\frac{\mu_0}{v_\mu^2}+\frac{(1-\rho^2)(1-\phi^2)h_1+(1-\phi)\sum_{t=1}^{T-1}(h_{t+1}-\phi h_t-\rho\sigma\bar{\epsilon}_t)}{\sigma^2(1-\rho^2)}\right\}.\end{aligned}$$

A.3 Sampling the skewness parameter β

.

Let the prior distribution of the skewness parameter β be $N(\beta_0, v_\beta^2)$. Denoting $\bar{W}_t = W_t - \frac{\zeta}{\zeta-2}$, the conditional posterior distribution follows

$$\pi(\beta|\centerdot) \propto \exp\left\{ -\frac{(\beta-\beta_0)^2}{2v_\beta^2} - \sum_{t=1}^{T} \frac{(y_t - \beta\bar{W}_t\gamma e^{\frac{h_t}{2}})^2}{2W_t e^{h_t}\gamma^2} - \sum_{t=1}^{T-1} \frac{(\bar{h}_{t+1} - \phi\bar{h}_t - \rho\sigma(y_t e^{\frac{-h_t}{2}} - \beta\bar{W}_t\gamma)/\sqrt{W_t}\gamma)^2}{2\sigma^2(1-\rho^2)} \right\},$$

from which we can generate a new draw $\beta|\centerdot \sim N(\mu_\beta, \sigma_\beta^2)$ with

$$\sigma_\beta^2 = \left\{ \frac{1}{v_\beta^2} + \frac{1}{1-\rho^2} \sum_{t=1}^{T-1} \frac{\bar{W}_t^2}{W_t} + \frac{\bar{W}_T^2}{W_T} \right\}^{-1},$$

$$\mu_\beta = \sigma_\beta^2 \left\{ \frac{\beta_0}{v_\beta^2} + \frac{1}{1-\rho^2} \sum_{t=1}^{T-1} \frac{y_t\bar{W}_t}{W_t e^{\frac{h_t}{2}}\gamma} + \sum_{t=1}^{T-1} \frac{(\bar{h}_{t+1} - \phi\bar{h}_t)\rho\bar{W}_t}{\sigma(1-\rho^2)\sqrt{W_t}} \right\}.$$

A.4 Sampling the W_t degrees of freedom parameter ζ

.

We adopt a Gamma prior $\pi_0(\zeta) \equiv G(s_\zeta, r_\zeta)$ for the d.o.f parameter of the mixture process W_t. The conditional posterior distribution of ζ involves the full joint likelihood,

$$\pi(\zeta|\centerdot) \propto \pi_0(\zeta) \prod_{t=1}^{T} IG(W_t; \frac{\zeta}{2}, \frac{\zeta}{2}) \exp\left\{ -\sum_{t=1}^{T} \frac{(y_t - \beta\bar{W}_t\gamma e^{h_t/2})^2}{2W_t e^{h_t}\gamma^2} - \sum_{t=1}^{T-1} \frac{(\bar{h}_{t+1} - \phi h_t - \rho\sigma\bar{\epsilon}_t)^2}{2\sigma^2(1-\rho^2)} \right\}.$$

The we use the MH algorithm to draw $\delta = \log(\zeta - 4)$ based on a normal approximation of the logarithm of the transformed posterior density function $\log\tilde{\pi}(\delta|\centerdot)$, whose mode and the second derivative around the model are μ_δ and σ_δ^2, respectively. We accept the draw with probability

$$\min\left\{ \frac{\pi(\zeta^*|\centerdot)N(\delta; \mu_\delta, -\sigma_\delta^{-2})\exp(\zeta)}{\pi(\zeta|\centerdot)N(\delta^*; \mu_\delta, -\sigma_\delta^{-2})\exp(\zeta^*)}, 1 \right\}.$$

B MCMC algorithm details for the factor SV model

B.1 Sampling the factor loadings

We sample the loadings matrix Λ using the MH algorithm based on a Laplace approximation and a multivariate Student's t-proposal distribution with arbitrarily chosen degrees of freedom v. To find the mode, we propose a Hessian-free optimisation routine (e.g. any quasi-Newton methods), based on the score function $\partial l(y_{1:T}|\Lambda)/\partial\Lambda_{ij} = \sum_{t=1}^{T} \partial l_t(y_t|\Lambda)/\partial\lambda_{ij}$ with λ_{ij} denoting the ij-th free element of Λ and

$$\begin{aligned}\frac{\partial l_t(y_t|\Lambda)}{\partial\lambda_{ij}} =& -\frac{1}{2}\left\{\frac{\partial \log|\Omega_t|}{\partial\lambda_{ij}} + \frac{\partial}{\partial\lambda_{ij}}(\tilde{y}_t'\Omega_t^{-1}\tilde{y}_t - 2\tilde{y}_t'\Omega_t^{-1}\Lambda F_t + F_t'\Lambda'\Omega_t^{-1}\Lambda F_t)\right\} \\ =& -\operatorname{tr}\left(\Omega_t^{-1}\Lambda V_t\frac{\partial\Lambda'}{\partial\lambda_{ij}}\right) + \tilde{y}_t'\Omega_t^{-1}\frac{\partial\Lambda}{\partial\lambda_{ij}}V_t\Lambda'\Omega_t^{-1}\tilde{y}_t + \tilde{y}_t'\Omega_t^{-1}\frac{\partial\Lambda}{\partial\lambda_{ij}}(I_p - 2V_t\Lambda'\Omega_t^{-1}\Lambda)F_t \\ &+ F_t'\left(\Lambda'\Omega_t^{-1}\frac{\partial\Lambda}{\partial\lambda_{ij}}V_t\Lambda'\Omega_t^{-1}\Lambda - \frac{1}{2}(\frac{\partial\Lambda'}{\partial\lambda_{ij}}\Omega_t^{-1}\Lambda + \Lambda'\Omega_t^{-1}\frac{\partial\Lambda}{\partial\lambda_{ij}})\right)F_t,\end{aligned}$$

where $\Omega_t^{-1} = U_t^{-1} - U_t^{-1}\Lambda(V_t^{-1} + \Lambda'U_t^{-1}\Lambda)^{-1}\Lambda'U_t^{-1}$. After some convergence criterion is met, we compute the inverse of observed information matrix, i.e. $\Sigma_\Lambda = (G(\mu_\Lambda)G(\mu_\Lambda)')^{-1}$ where $G(\mu_\Lambda)$ is the gradient matrix whose t-th column equals $\operatorname{vec}(\{\partial l_t(\tilde{y}_t|\Lambda)/\partial\lambda_{ij}\}_{i,j})$ with i, j running through all free elements of Λ. We then draw a candidate $\operatorname{vec}(\Lambda^*)$ from the proposal, with acceptance probability

$$\min\left\{\frac{\pi_0(\operatorname{vec}(\Lambda^*))\exp(l(y_{1:T}|\Lambda^*))T(\operatorname{vec}(\Lambda);\mu_\Lambda,\Sigma_\Lambda,v)}{\pi_0(\operatorname{vec}(\Lambda))\exp(l(y_{1:T}|\Lambda))T(\operatorname{vec}(\Lambda^*);\mu_\Lambda,\Sigma_\Lambda,v)}, 1\right\}.$$

B.2 Initialization

We initialize our model using principal components (PC). Let us rewrite model (15) as $Y = F\Lambda' + u$ where $Y \in \mathbb{R}^{T\times n}$, $F \in \mathbb{R}^{T\times p}$ and $u \in \mathbb{R}^{T\times n}$. So the t-th row in of Y, F and u are y_t', f_t' and u_t' respectively, and f_t is the PC's. Or equivalently we have

$$\boldsymbol{y} = (I_n \otimes F)\boldsymbol{\lambda} + \boldsymbol{u}, \tag{21}$$

where $\boldsymbol{y} = \operatorname{vec}(Y)$, $\boldsymbol{\lambda} = \operatorname{vec}(\Lambda')$, and $\boldsymbol{u} = \operatorname{vec}(u)$. Under conditions specified by Doz et al. (2011), PCs are consistent estimators of the factors. We apply the criterion in Bai and Ng (2002) to choose the preliminary number of factors. Because we impose identification restriction on Λ, the matrix of eigenvectors in relation to PCs cannot initialize Λ. Notice that (21) is a linear

regression model in $\boldsymbol{\lambda}$. So the identification restriction implies a linear constraint of the form

$$R\boldsymbol{\lambda} = r.$$

This gives us the constraint OLS estimate $\hat{\boldsymbol{\lambda}}_{cols}$ to initialize Λ, which is

$$\hat{\boldsymbol{\lambda}}_{cols} = \hat{\boldsymbol{\lambda}}_{ols} - (I_n \otimes (F'F)^{-1})R'(R(I_n \otimes (F'F)^{-1})R')^{-1}(R\hat{\boldsymbol{\lambda}}_{ols} - r), \tag{22}$$

where $\hat{\boldsymbol{\lambda}}_{ols} = (I_n \otimes (F'F)^{-1}F')\boldsymbol{y}$. Given $\hat{\boldsymbol{\lambda}}_{cols}$, Doz et al. (2011) suggest that the estimate of factors $E(f_t|y_{1:T})$ can be obtained by

$$f_t = (\Lambda_{cols}{}'\Lambda_{cols})^{-1}\Lambda_{cols}{}'y_t, \qquad t = 1, ..., T. \tag{23}$$

We complete the initialization of factors f_t for $t = 1, ..., T$ and loadings Λ with iterations over (22) and (23) until convergence. The above procedure delivers sound initialization, especially for the loading matrix Λ with identification restrictions. Chan et al. (2013) show that there exists a unique mapping which rotates the PCs towards the factors under certain identification scheme imposed on the loadings.

Given the initialized Λ and f_t, we compute the residuals as $u_t = y_t - \Lambda f_t$. Hence, we obtain $n + p$ univariate series $z_{j,t} = f_{j,t}$ for $j = 1, ..., p$ and $z_{p+i,t} = u_{i,t}$ for $i = 1, ..., n$. For any $k \in \{1, ..., n + p\}$, we treat $\{z_{k,t}\}_{t=1}^T$ as a basic SV model re-parametrized according to Ruiz (1994). We can then efficiently implement a quasi-maximum likelihood (QML) estimator based on the following approximate linear Gaussian state space model

$$\begin{aligned} \log(z_{k,t}^2) &= \log(2) + \psi(1/2) + h_{k,t} + \sqrt{\psi'(1/2)}\epsilon_{k,t}, \quad && t = 1, ..., T, \\ h_{k,t+1} &= \mu_k(1 - \phi_k) + \phi_k h_{k,t} + \sigma_k \eta_{k,t}, && t = 1, ..., T - 1, \end{aligned} \tag{24}$$

where $\psi(\cdot)$ is the digamma function and $\psi'(\cdot)$ is its first order derivative. $\epsilon_{k,t}$ and $\eta_{k,t}$ are i.i.d normal with correlation coefficient ρ_k. Maximizing the log-likelihood via Kalman filter (Durbin and Koopman 2012) gives the QML estimate of ϕ_k, σ_k, ρ_k and μ_k, which serve as initializations for $k = 1, ..., n + p$. We choose the initial value of the skewness parameter β_k to be zero and the d.o.f ζ_k to be 20 for all k.

We initialize the SV process $\{h_{j,t}\}_{t=1}^T$ and $\{l_{i,t}\}_{t=1}^T$ for all i and j by applying the simulation smoother of De Jong and Shephard (1995) to model (24). We initialize the inverse gamma mixing component $W_{j,t}$ and $Q_{i,t}$ for all i and j by drawing from $IG(s_k, r_{k,t})$, where $s_k = \zeta_k/2 + 1$

and $r_{k,t} = \zeta_k/2 + z_{k,t}^2 \exp(-h_{k,t})/2$.

C Forecasting and filtering

The filtering algorithm mainly follows from Chib et al. (2006). Keeping θ at some posterior estimate, at time T the particles $\{x_T^k\}_{k=1}^K$ with normalized weights $\{\overline{\omega}_T^k\}_{k=1}^K$ are propagated 1-period forward based on their transition dynamics. The 1-period ahead forecast $\dot{y}_{T+1}$ is given by

$$\dot{y}_{T+1} = \sum_{k=1}^{K} \overline{\omega}_T^k \hat{y}_{T+1}^k,$$

where for $k = 1, ..., K$, $\hat{y}_{T+1}^k$ is imputed from

$$\hat{y}_{T+1}^k \sim N(\tilde{y}_{T+1}^k + \Lambda F_{T+1}^k, \Omega_{T+1}^k), \qquad \Omega_{T+1}^k = \Lambda V_{T+1}^k \Lambda' + U_{T+1}^k.$$

We can obtain the S-period ahead forecast $\dot{y}_{T+S}$ similarly. The predicted total return over S periods $\sum_{s=1}^S \dot{y}_{T+s}$ thus follows a mixture Gaussian distribution

$$\sum_{s=1}^{S} \dot{y}_{T+s} \sim \sum_{k=1}^{K} \overline{\omega}_T^k N(\sum_{s=1}^{S} \tilde{y}_{T+s}^k + \Lambda F_{T+s}^k, \sum_{s=1}^{S} \Omega_{T+s}^k),$$

we can use to estimate return moments over S periods. This is essential for portfolio management, and for computing risk metrics such as the tail index and the VaR.

The filtered mean return and covariance matrix are

$$\mu_{t|t-1} = E(y_t | y_{1:t-1}, \theta), \qquad \Omega_{t|t-1} = E(\Omega_t | y_{1:t-1}, \theta),$$

which we can estimate by

$$\hat{\mu}_{t|t-1} = \sum_{k=1}^{K} \overline{\omega}_{t-1}^k (\tilde{y}_t^k + \Lambda F_t^k), \qquad \hat{\Omega}_{t|t-1} = \sum_{k=1}^{K} \overline{\omega}_{t-1}^k \Omega_t^k,$$

where $\Omega_t^k = \Lambda V_t^k \Lambda' + U_t^k$ and $\{\overline{\omega}_{t-1}^k\}_{k=1}^K$ are the normalized weights pertaining to the particle system at time $t-1$. Chib et al. (2006) express the filtered correlation as

$$R_{t|t-1} = E(\Upsilon_t | y_{1:t-1}, \theta).$$

Υ_t is thus the conditional correlation matrix for $y_t|\{h_{j,t}\}_{j=1}^p, \{W_{j,t}\}_{j=1}^p, \{l_{i,t}\}_{i=1}^n, \{Q_{i,t}\}_{i=1}^n$, or

$$\Upsilon_t = D(\Omega_t)^{-\frac{1}{2}} \Omega_t D(\Omega_t)^{-\frac{1}{2}},$$

where $D(\Sigma)$ denotes the matrix with diagonal elements equal to those of Σ and zero off-diagonal elements. We can thus estimate $R_{t|t-1}$ by

$$\hat{R}_{t|t-1} = \sum_{k=1}^{K} \overline{\omega}_{t-1}^k \Upsilon_t^k = \sum_{k=1}^{K} \overline{\omega}_{t-1}^k D(\Omega_t^k)^{-\frac{1}{2}} \Omega_t^k D(\Omega_t^k)^{-\frac{1}{2}}.$$