Supplementary appendix for "Leverage, asymmetry and heavy tails in the high-dimensional factor stochastic volatility model"

Blinded*

Blinded

Blinded†

Blinded

January 31, 2019

'

*Blinded
†Blinded

S.1 The leverage effect multiplier

The leverage effect for the univariate SV model is $\text{Corr}(\nu_t, \eta_t) = \text{Cov}(\nu_t, \eta_t)/\sqrt{\text{Var}(\nu_t)\text{Var}(\eta_t)}$, where the numerator

$$\text{Cov}(\nu_t, \eta_t) = \mathbb{E}(\sqrt{W_t})\rho\sigma.$$

Since $W_t \sim IG(\frac{\zeta}{2}, \frac{\zeta}{2})$, $\frac{1}{\zeta}W_t$ is $IG(\frac{\zeta}{2}, \frac{1}{2})$-distributed or $Inv - \chi^2(\zeta)$ distributed. Let $\tilde{W}_t = \sqrt{W_t}$, we have $\frac{1}{\zeta}\tilde{W}_t^2 \sim Inv - \chi^2(\zeta)$ with Jacobian $\frac{2}{\zeta}\tilde{W}_t$. It follows that

$$\begin{aligned}
\mathbb{E}(\tilde{W}_t) &= \int_0^\infty \frac{2}{\zeta}\tilde{W}_t^2 \frac{2^{-\zeta/2}}{\Gamma(\zeta/2)} \zeta^{\frac{2}{\zeta+2}} \tilde{W}_t^{-(\zeta+2)} \exp(\frac{-\zeta}{2\tilde{W}_t^2}) d\tilde{W}_t \\
&= \frac{1}{\sqrt{\zeta}}\tilde{W}_t \int_0^\infty \frac{1}{2^{\zeta/2-1}\Gamma(\zeta/2)} \left(\frac{1}{\sqrt{\zeta}}\tilde{W}_t\right)^{-\zeta-1} \exp(-\frac{1}{2}\frac{1}{\sqrt{\zeta}}\tilde{W}_t)^{-2} d\tilde{W}_t \\
&= \frac{\sqrt{\zeta}}{2^{\zeta/2-1}\Gamma(\zeta/2)} \int_0^\infty y^{-\zeta} \exp(\frac{1}{-2y^2}) dy \\
&= \frac{\sqrt{\zeta}}{2^{\zeta/2-1}\Gamma(\zeta/2)} \int_0^\infty 2^{\zeta/2-3/2} z^{\zeta/2-1} \exp(-z) dz \\
&= \frac{\sqrt{\zeta}\Gamma\big((\zeta-1)/2\big)}{\sqrt{2}\Gamma(\zeta/2)},
\end{aligned}$$

where we use substitution $y \equiv \frac{1}{\sqrt{\zeta}}\tilde{W}_t$ and $z \equiv \frac{1}{2}y^{-2}$. In the denominator, the variance of the generalised hyperbolic skew Student's t-distributed error ν_t is given by Aas and Haff (2006) (in their parametrisation δ^2 and v are both equivalent to our ζ), i.e.

$$\text{Var}(\nu_t) = \frac{2\beta^2\zeta^2}{(\zeta-2)^2(\zeta-4)} + \frac{\zeta}{\zeta-2}.$$

With these quantities, the unconditional leverage effect multiplier can be shown to be to one given in Section 2.1.

S.2 Marginal likelihood estimation

S.2.1 Importance sampling squared for marginal likelihood estimation

For model comparisons, one needs to calculate the marginal likelihood $p(y_{1:T}|\mathcal{M})$ under a certain model $\mathcal{M}$, so that the Bayes factor $p(y_{1:T}|\mathcal{M}_1)/p(y_{1:T}|\mathcal{M}_2)$ can be computed[1]. Let us suppress

[1] See Chib (2001) and Chib and Jeliazkov (2005) for alternative methods that deal with models of less complexity, using the so-called "reduced MCMC" run based on likelihood identity.

the dependence on model $\mathcal{M}$. We can write the marginal likelihood as

$$p(y_{1:T}) = \int p(y_{1:T}, \theta) d\theta = \int p(y_{1:T}|\theta)\pi_0(\theta) d\theta$$
$$= \int \frac{p(y_{1:T}|\theta)\pi_0(\theta)}{q(\theta|y_{1:T})} q(\theta|y_{1:T}) d\theta,$$

where $\pi_0(\theta)$ is the prior, and $q(\theta|y_{1:T})$ is an importance density mimicking the posterior $\pi(\theta|y_{1:T}) \propto p(y_{1:T}|\theta)\pi_0(\theta)$. The above integral can be computed via Monte Carlo simulation. It follows

$$\hat{p}(y_{1:T}) = \frac{1}{S}\sum_{s=1} w(\theta^s), \quad \text{where } w(\theta^s) = \frac{p(y_{1:T}|\theta^s)\pi_0(\theta^s)}{q(\theta^s|y_{1:T})} \text{ and } \theta^s \sim q(\theta|y_{1:T}). \tag{S1}$$

The is straightforward to implement if the likelihood $p(y_{1:T}|\theta)$ is available in closed form, which is not our case due to many latent processes. Tran et al. (2014) show that under mild conditions that if there exists an unbiased estimate of the likelihood, i.e. $E(\tilde{p}(y_{1:T}|\theta)) = p(y_{1:T}|\theta)$, averaging importance weights to compute the marginal likelihood as formula (S1) is still valid with $p(y_{1:T}|\theta)$ replaced by $\tilde{p}(y_{1:T}|\theta)$[2].

The particle efficient importance sampling (PEIS) developed by Scharth and Kohn (2016) provide a particle MCMC method (Andrieu et al., 2010) that we can use within the IS^2 framework [3]. PEIS is similar to the PGAS-EIS sampler which builds a sequential but globally optimal importance density $q(x_t|x_{t-1}, y_{1:T})$. Conceptually, the global optimality of PEIS which minimises the variance of importance weights as in (5) and (10) is what makes it efficient for evaluating marginal likelihood.

The algorithm is straightforward at first sight: (i) construct $q(\theta|y_{1:T})$ based on posterior samples; (ii) draw $\theta^s \sim q(\theta|y_{1:T})$; (iii) construct an importance density based on θ^s; (iv) apply PEIS to compute $\tilde{p}(y_{1:T}|\theta^s)$ and $w(\theta^s)$; (v) average out $w(\theta^s)$ for $s = 1, ..., S$. Unfortunately, this IS^2 algorithm is infeasible. The construction of an importance density that targets the $2\times(n+p)\times T$-variate density $p(h_{1:T}, l_{1:T}, W_{1:T}, Q_{1:T}|y_{1:T}, \theta^s)$ is nearly impossible. To circumvent this problem, we propose a feasible IS^2 alternative.

Suppose we have stored the importance parameters corresponding to the $n+p$ proposals when applying the PGAS-EIS sampler within each iteration of the Markov chain. We can start the

[2]For many models, an unbiased estimate of likelihood is readily available using particle marginal Metropolis-Hastings (PMMH) algorithm (Andrieu et al. 2010 and Del Moral and Formulae 2004). For example, the paper of factor SV model by Chib et al. (2006) applies the celebrated auxiliary particle filter (APF) of Pitt and Shephard (1999) to compute the posterior ordinate for the evaluation of Bayes factors based on reduced MCMC run.

[3]According to them, PEIS significantly outperforms PMMH and particle MCMC in terms of variance reduction for the Monte Carlo estimate of likelihood which is critical for efficient computation of marginal likelihood.

feasible IS^2 algorithm by firstly constructing $q(\theta|y_{1:T})$. Suppose we believe that after a chosen burn-in period, the chain has converged to its stationary distribution. We have a posterior sample of θ as shown in Figure S.1 (for simplicity we illustrate the vector as one-dimensional). An m-component Gaussian mixture is fit to the posterior sample of θ via standard EM algorithm[4]

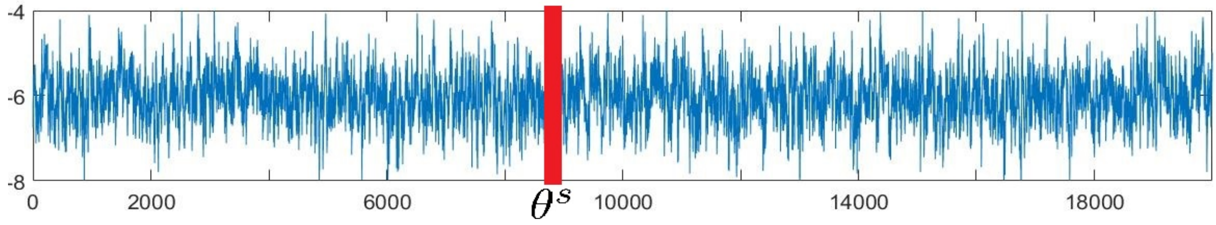

Figure S.1: The posterior sample of θ after convergence. The red bar indicates a certain θ^s.

with m determined by information criterion.

Importantly, we then draw θ^s non-parametrically from $q(\theta|y_{1:T})$. That is, θ^s is a vector in the posterior sample as the red bar in Figure S.1, with the probability of being drawn proportional to $q(\theta^s|y_{1:T})$. After this step, we can simply retrieve the corresponding importance parameters from the $n + p$ PGAS-EIS samplers. This enables us to apply PEIS to compute $\tilde{p}(y_{1:T}|\theta^s)$ and thus the marginal likelihood. This procedure completely avoids the construction of an intimidatingly large importance density by utilising the $n + p$ individual proposal densities constructed when applying the PGAS-EIS samplers. We leave details of applying PEIS in the IS^2 framework to the factor SV model in the appendix. Also, the appendix discusses forecasting and filtering methods following Chib et al. (2006) which we use in our empirical study.

S.2.2 Particle EIS and IS^2 details

In the following, we slightly abuse the use of notations. PEIS finds an unbiased estimate for the likelihood function $p(y_{1:T}|\theta) = p(y_1|\theta)\prod_{t=2}^{T} p(y_t|y_{1:t-1}, \theta)$ by propagating the particle system with forward weights resampling (Shephard and Pitt 1997 and Scharth and Kohn 2016). Suppose at time t, one has the particle system $\{x^i_{1:t}, \omega^i_t\}^M_{i=1}$, where

$$x^i_t = \{\{h^i_{j,t}\}^p_{j=1}, \{W^i_{j,t}\}^p_{j=1}, \{l^i_{k,t}\}^n_{k=1}, \{Q^i_{k,t}\}^n_{k=1}\}.$$

[4]For parameters that take restricted values, say $\sigma_k > 0$ for $k = 1, ..., n+p$, we consider transformation such that $\log \sigma_k \in \mathbb{R}$.

Suppressing the dependence on θ^s which is a specific draw from the posterior sample of θ, the forward weights are calculated according to

$$\overrightarrow{\omega}_t^i = \varpi_t^i \frac{\chi_q(x_t^i; \delta_{t+1}^s)}{\chi_p(x_{t-1}^i; y_{t-1})}, \qquad i = 1, ..., M,$$

where ϖ_t^i is the normalised weight $\varpi_t^i = \omega_t^i / \sum_{i=1}^M \omega_t^i$. $\chi_q(x_t^i; \delta_{t+1}^s)$ is the integration constant of the importance density $q(x_{t+1}^i | x_t^i, y_{1:T})$, the product of $n+p$ individual EIS importance densities with parameters δ_{t+1}^s with kernel $k_q(x_{t+1}^i, x_t^i; \delta_{t+1}^s)$; whereas $\chi_p(x_{t-1}^i; y_{t-1})$ is the integration constant of the transition density $p(x_t | x_{t-1}^i, y_{t-1})$. Next, with the normalised forward weights

$$\overrightarrow{\varpi}_t^i = \frac{\overrightarrow{\omega}_t^i}{\sum_{j=1}^M \overrightarrow{\omega}_t^j},$$

one calculates the effective sample size $ESS = 1/\sum_{i=1}^M (\overrightarrow{\varpi}_t^i)^2$. If ESS drops below a predetermined threshold, resampling is applied to M particles $\{x_t^i\}_{i=1}^M$ with probability $\{\overrightarrow{\varpi}_t^i\}_{i=1}^M$, and all normalised weights ϖ_t^i are set to be $1/M$ for $i = 1, ..., M$. At time $t+1$, M new particles $\{x_{t+1}^i\}_{i=1}^M$ need to be generated from the importance density $q(x_{t+1} | x_t^i, y_{1:T})$, which requires M draws from $n+p$ individual importance densities with each as in (8). To do this, we have to recover the factor process, which is needed to propagate the SV processes due to the presence of leverage effects. We take the conditional posterior mean of the factor process, i.e.

$$f_t^i = (\Lambda'(U_t^i)^{-1}\Lambda + (V_t^i)^{-1})^{-1}(\Lambda'(U_t^i)^{-1}\tilde{y}_t^i + (V_t^i)^{-1}F_t^i), \tag{S2}$$

where $\tilde{y}_t^t$, F_t^i, V_t^i and U_t^i are as in (17). The idiosyncratic noise is thus

$$u_t^i = y_t - \Lambda f_t^i.$$

With both f_t^i and u_t^i, leverage effect can be accounted for. For example, $h_{j,t+1}^i$ for $i = 1, ..., M$ can be obtained by

$$h_{j,t+1}^i = \mu^{f_j}(1 - \phi^{f_j}) + \phi^{f_j} h_{j,t}^i + \frac{\rho^{f_j}\sigma^{f_j}}{W_{j,t}^i}(f_{j,t}^i - \alpha^{f_j} - \beta^{f_j} W_{j,t}^i) + \sqrt{1 - {\rho^{f_j}}^2}\sigma^{f_j} {\eta_{j,t}^*}^i,$$

with ${\eta^*_{j,t}}^i \sim N(0,1)$ for $j = 1, ..., p$. Other latent processes propagate similarly[5]. Once the prorogation of all particles is finished, the importance weights are recalculated as

$$\omega^i_{t+1} = \begin{cases} \overline{\omega}^i_t \times p(y_{t+1}|x^i_{t+1})p(x^i_{t+1}|x^i_t, y_t)/k_q(x^i_{t+1}, x^i_t; \delta_{t+1}), & \text{if resampling} \\ \overline{\omega}^i_t \times p(y_{t+1}|x^i_{t+1})p(x^i_{t+1}|x^i_t, y_t)/q(x^i_{t+1}|x^i_t, y_{1:T}), & \text{otherwise.} \end{cases}$$

The last step at time t is to record the estimate of the likelihood contribution via

$$\tilde{p}(y_{t+1}|y_{1:t}) = \begin{cases} (\sum_{i=1}^M \overrightarrow{\omega}^i_t)(\sum_{i=1}^M \omega^i_{t+1}), & \text{if resampling} \\ \sum_{i=1}^M \omega^i_{t+1}, & \text{otherwise.} \end{cases}$$

Once the propagation of particle system reaches time T, the unbiased estimate of likelihood is simply given by $\tilde{p}(y_{1:T}|\theta) = \tilde{p}(y_1|\theta) \prod_{t=2}^T \tilde{p}(y_t|y_{1:t-1}, \theta)$ with obvious modification to $\tilde{p}(y_1|\theta)$.

S.2.3 Simulation study

For model selection, we compute marginal likelihood used in Bayes factor via the feasible IS^2 method introduced in Section 3. We here focus on the ability of feasible IS^2 to select the correct number of factors.

Table S.1 shows conditional average log-likelihood or posterior ordinate with hyperparameters evaluated at their posterior means $\hat{\theta}$, using the modified PEIS method (see appendix), i.e. $\frac{1}{T}\tilde{p}(y_{1:T}|\hat{\theta})$. We report evaluations with different number of particles which provides a guideline as for how many particles are needed for empirical use[6]. From Table S.1 we see that, the log-likelihood estimates for aLE_aSK converge with 100 particles. For nLE_nSK and sLE_aSK with the number of particles larger than 200, there is no major difference in the log-likelihood. For aLE_sSK and sLE_sSK, more than 300 particles lead to converged log-likelihood.

It is reasonable to believe that the number of particles needed does not change significantly across different parameter values when one computes the likelihood (Tran et al., 2014). So when applying IS^2 to calculate the marginal likelihood, we use 300 particles for each draw of parameters. We simulate 30 times to obtain 30 different sLE_sSK datasets with 8 factors as

[5]Antithetic variables are used to reduce Monte Carlo noise during particle propagation. In particular, pairs of perfectly negatively correlated Gaussian variables are generated for all SV processes (Durbin and Koopman 2000 and Scharth and Kohn 2016), and pairs of inverse gamma variables are generated using a Gaussian copula with perfect negative correlation.

[6]Scharth and Kohn (2016) detail an algorithm to choose the number of particles using PEIS based on the trade-off of overhead cost for constructing the EIS importance density and the Monte Carlo variance of log-likelihood. But this procedure becomes prohibitively time-consuming for our high-dimensional model, we thus use a simple heuristic.

Table S.1: PEIS Log-likelihood Evaluation

Dataset	Number of particles 100	200	300	500	1000
sLE_sSK	-1864.79	-1834.26	-1833.68	-1833.27	-1833.34
sLE_aSK	-1824.61	-1819.57	-1819.24	-1819.66	-1819.29
aLE_sSK	-1841.27	-1828.03	-1830.86	-1830.94	-1830.62
aLE_aSK	-1812.46	-1811.74	-1811.09	-1811.42	-1812.00
nLE_nSK	-1923.95	-1920.53	-1921.81	-1921.86	-1921.44

Reported are average log-likelihood evaluated at posterior mean estimates of hyperparameters using different number of particles.

Table S.2: Frequency(%) Of Bayes Factors With Different Number Of Factors

sLE_sSK	DGP: 8 factors 1-3.2	3.2-10	10-100	>100	Total>10
8/6	0	0	0	100.00	100.00
8/7	0	0	0	100.00	100.00
8/9	0	0	3.33	96.67	100.00
8/10	0	0	0	100	100.00
7/6	0	16.67	40.00	43.33	83.33
7/9	0	6.67	56.67	36.67	93.33
7/10	0	3.33	6.67	0.90	96.67
9/6	0	0	0	100.00	100.00
9/10	0	16.67	36.67	46.67	83.33

The choice of range for Bayes factors is according to the Jeffrey's scale. Frequency distribution is determined across 30 simulated replications. The leftmost column indicates the comparison between two specifications. For example, 8/6 corresponds to the Bayes factor of a model with 8 factors against the one with 6 factors.

before. Out of the 30 simulated replications, the IC_p1 criterion of Bai and Ng (2002) chooses 8 factors 21 times, and 6, 7, 9, and 10 factors twice, twice, 4 times and once respectively.

Table S.2 shows the model comparisons based on Bayes factor. The Jeffrey's scale suggests decisive evidence in favor of the model with 8 factors against all cases. It can be concluded that if the true DGP follows sLE_sSK, Bayes factor based on the feasible IS^2 method of computing marginal likelihood is more convincing than the criteria of Bai and Ng (2002). Furthermore, we emphasize that the proposed feasible IS^2 procedure is much more easy to implement than the reduced MCMC run method in Chib and Greenberg (1994) and Chib et al. (2006).

Table S.3: Number of Factor Based on Marginal Likelihood

Jeffrey's scale	Number of factors 4/2	4/3	4/5	4/6	3/5	3/6	6/5
1-3.2	–	–	–	–	–	√	–
3.2-10	–	–	–	–	√	–	√
10-100	√	√	–	–	–	–	–
> 100	–	–	√	√	–	–	–

Shaded cell indicates the Bayes factor using IS^2 marginal likelihood for one choice of number of factors against another falls into a certain category given by the Jeffrey's scale.

S.2.4 Empirical application

Table S.3 shows the Bayes factor calculated via IS^2 marginal likelihood for model specifications with different number of factors. The number of factors under consideration are between 2 and 6, in line with other literature. Model with 4 factors is preferred over all other specifications, in particular over model with 5 and 6 factors, which is the choice made by the IC_p3 of Bai and Ng (2002). Also the IC_p1 delivers almost equal values for specification with 5 and 6 factors. Via the use of IS^2 for calculating the marginal likelihood, we can safely choose a model with 4 factors. Other comparisons show that the model with 3 factors is slightly preferred over 6-factor model, and evidently preferred over the model with 5 factors.

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.

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

Chib, S. (2001). Markov chain Monte Carlo methods: computation and inference. *Handbook of econometrics 5*, 3569–3649.

Chib, S. and E. Greenberg (1994). Bayes inference in regression models with ARMA (p, q) errors. *Journal of Econometrics 64*(1-2), 183–206.

Chib, S. and I. Jeliazkov (2005). Accept–reject metropolis–hastings sampling and marginal likelihood estimation. *Statistica Neerlandica 59*(1), 30–44.

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

Del Moral, P. and F.-K. Formulae (2004). Genealogical and interacting particle systems with applications, Probability and Its Applications.

Durbin, J. and S. J. Koopman (2000). Time series analysis of non-Gaussian observations based on state space models from both classical and Bayesian perspectives. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 62*(1), 3–56.

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

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.

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*.