Efficient estimation of multivariate stochastic volatility models via Wishart processes

1 Efficient Importance Sampling Procedure

Multivariate stochastic volatility model via Wishart processes:

$$y_t = \epsilon_t, \qquad \epsilon_t \sim N_k(0_k, \Omega^{-1}), \tag{1}$$

$$\Omega_t | v, \Omega_{t-1} \sim W_k(v, S_{t-1}), \tag{2}$$

$$S_t = \frac{1}{v}(A^{1/2})(\Omega_t)^d (A^{1/2})'. \tag{3}$$

Likelihood function associated with observables $\underline{y}$:

$$\begin{aligned} L(\theta|\underline{y}) &= \int\int f(\underline{y}, \underline{\Omega}; \theta) d\underline{\Omega} \\ &= \int ... \int \prod_{t=1}^{T} f(y_t, \Omega_t | \underline{\Omega}_{t-1}, \underline{y}_{t-1}; \theta) d\Omega_1 ... d\Omega_T, \end{aligned} \tag{4}$$

Conditional densities in state space representation:

$$f(y_t, \Omega_t | \underline{\Omega}_{t-1}, \underline{y}_{t-1}; \theta) = g(y_t | \Omega_t, \underline{y}_{t-1}; \theta) \cdot p(\Omega_t | \underline{\Omega}_{t-1}, \underline{y}_{t-1}; \theta), \tag{5}$$

Likelihood function separated by measurement and transition densities:

$$L(\theta|\underline{y}) = \int \prod_{t=1}^{T} g(y_t | \Omega_t; \theta) \cdot p(\Omega_t | \Omega_{t-1}; \theta) d\underline{\Omega}, \tag{6}$$

Introduction of importance samplers to the likelihood:

$$L(\theta|\underline{y}) = \int \prod_{t=1}^{T} \underbrace{\left[\frac{f(y_t, \Omega_t | \underline{\Omega}_{t-1}, \underline{y}_{t-1}; \theta)}{m(\Omega_t | \underline{\Omega}_{t-1}; \gamma_t)} \right]}_{\varphi(y_t, \Omega_t; \theta, \gamma)} \prod_{t=1}^{T} m(\Omega_t | \underline{\Omega}_{t-1}; \gamma_t) d\underline{\Omega} \tag{7}$$

IS-MC estimate of $L(\theta|\underline{y})$:

$$\widetilde{L}_S(\theta|\underline{y}) = \frac{1}{S} \sum_{i=1}^{S} \left\{ \prod_{t=1}^{T} \left[\frac{f(y_t, \widetilde{\Omega}_t^i | \underline{\widetilde{\Omega}}_{t-1}^i, \underline{y}_{t-1}; \theta)}{m(\widetilde{\Omega}_t^i | \underline{\widetilde{\Omega}}_{t-1}^i; \gamma_t)} \right] \right\}, \tag{8}$$

Decomposition of $m(\cdot)$ into a kernel $k(\cdot)$ and its integrating constant $\chi(\cdot)$:

$$m(\Omega_t|\underline{\Omega}_{t-1};\gamma_t) = \frac{k(\underline{\Omega}_t;\gamma_t)}{\chi(\underline{\Omega}_{t-1};\gamma_t)}, \quad \chi(\underline{\Omega}_{t-1};\gamma_t) = \int k_t(\underline{\Omega}_t;\gamma_t)d\Omega_t. \tag{9}$$

EIS minimization problem:

$$\hat{\gamma}_t(\theta) = \min_{\gamma_t} \sum_{i=1}^{S} \left\{ \ln \left[f(y_t, \widetilde{\Omega}_t^i | \widetilde{\underline{\Omega}}_{t-1}^i, \underline{y}_{t-1}; \theta) \cdot \chi(\widetilde{\Omega}_t^i; \gamma_{t+1}) \right] - c_t - \ln k(\widetilde{\Omega}_t^i; \gamma_t) \right\}, \tag{10}$$

$$\hat{\gamma}_t(\theta) = \min_{\gamma_t} \sum_{i=1}^{S} \left\{ \ln \chi(\widetilde{\Omega}_t^i; \gamma_{t+1}) - c_t - \ln \zeta(\widetilde{\Omega}_t^i; \gamma_t) \right\}, \tag{11}$$

Density kernels of $g(\cdot)$, $p(\cdot)$ and $k(\cdot)$:

$$g(y_t|\Omega_t;\theta) \propto |\Omega_t|^{\frac{1}{2}} \exp\left\{-\frac{1}{2} y_t' \Omega_t y_t\right\}, \tag{12}$$

$$p(\Omega_t|\Omega_{t-1};\theta) \propto |\Omega_t|^{\frac{v-k-1}{2}} \exp\left\{-\frac{1}{2} \operatorname{tr}\left[S_{t-1}^{-1} \cdot \Omega_t\right]\right\}, \tag{13}$$

$$k(\underline{\Omega}_t;\gamma) = g(y_t|\Omega_t;\theta) \cdot p(\Omega_t|\Omega_{t-1};\theta) \cdot \zeta(\Omega_t;\gamma_t), \tag{14}$$

$$\zeta(\Omega_t;\gamma_t) \propto |\Omega_t|^{\frac{\gamma_{1,t}}{2}} \exp\left\{-\frac{1}{2} \operatorname{tr}\left[\Gamma_t \cdot \Omega_t\right]\right\}, \tag{15}$$

$$k(\underline{\Omega}_t;\gamma) \propto |\Omega_t|^{\frac{v+1+\gamma_{1,t}-k-1}{2}} \exp\left\{-\frac{1}{2} \operatorname{tr}\left[(y_t \cdot y_t' + S_{t-1}^{-1} + \Gamma_t) \cdot \Omega_t\right]\right\}, \tag{16}$$

Degrees of freedom and scale matrix of $m(\cdot)$:

$$v^* = v + 1 + \gamma_{1,t}, \quad \text{and} \quad S_{t-1}^* = [y_t \cdot y_t' + S_{t-1}^{-1} + \Gamma_t]^{-1}, \tag{17}$$

Analytical expression for the integrating constant of the period t EIS sampler:

$$\chi(\Omega_{t-1};\gamma_t) \propto \frac{|S_{t-1}^*|^{\frac{v^*}{2}}}{|S_{t-1}|^{\frac{v}{2}}}, \tag{18}$$

2 Code Steps

2.1 Code $lik\_KChi\_initial$

In order to analyze statistical gains from using the EIS optimization method, an initial sampler can be constructed based on the kernel (14) by setting $\gamma_t = 0 \,\forall\, t$, which yields the following initial degrees of freedom and scale matrix of $m(\cdot)$:

$$v^{*INI} = v + 1, \quad \text{and} \quad S_{t-1}^{*INI} = [y_t \cdot y_t' + S_{t-1}^{-1}]^{-1}.$$

Note that S_{t-1}^{*INI} incorporates contemporaneous information on the observables y_t (through $y_t \cdot y_t'$), but it can not explore information from EIS optimizations, which are not transferred to the importance sampler $m(\cdot)$. More specifically, the initial sampler does not incorporate information on Ω_t contained in y_{t+1}, because EIS parameters γ_t can not transfer information related to the EIS smoother through matrix Γ_t since $\gamma_t = 0 \,\forall\, t$.

Thus, the initial sampler is used to draw N trajectories of the latent precision matrices $\{\{\widetilde{\Omega}_t^i\}_{t=1}^T\}_{i=1}^N$ which do not incorporate information on the dynamics of Ω_t contained in $f(y_{t+1}, \Omega_{t+1}|\underline{\Omega}_t, \underline{y}_t; \theta)$. For this reason, the initial sampler is used with comparison purposes.

Code $lik\_KChi\_initial$ approximates the log-likelihood function described in Eq. (8) using the precision matrices $\{\{\widetilde{\Omega}_t^i\}_{t=1}^T\}_{i=1}^N$ sampled with the initial sampler:

a. The initial sampler is initialized using initial values for the precision matrix Ω_0 at time $t = 1$, the lower triangular matrix $A^{\frac{1}{2}}$ containing the intertemporal sensitivity parameters of Ω_t, the persistence parameter d and the degrees of freedom v from the k-dimensional Wishart transition density. For example, the initial sampler can use values for $A^{\frac{1}{2}}$, d and v which follow the parameterization stated in Table 4 from Philipov and Glickman (2006);

b. N draws of S_t, the scale matrix of the Wishart transition density, are computed $\forall\, t = 1, ..., T$, using the initial parameterization defined above. Note that for $t = 1$, the N computed scale matrices are equal once the initial Ω_0 is the same for the entire MC sample. Next, N draws of S_{t-1}^{*INI}, the scale matrix of the EIS sampler, can be computed via Eq. (19);

c. Once a sample of size N of scale matrices S_t, $\{\widetilde{S}_t^i\}_{t=1}^T\}_{i=1}^N$, are computed, the computation of S_{t-1}^{*INI} is straightforward and one can sample N trajectories of $\{\{\widetilde{\Omega}_t^i\}_{t=1}^T\}_{i=1}^N$ from the EIS sampler $\Omega_t|v^{*INI}, \Omega_{t-1} \sim W_k(v^{*INI}, S_{t-1}^{*INI})$;

d. Ultimately, code $lik\_KChi\_initial$ computes the IS-MC estimate of the log-likelihood of IS ratio $\varphi(y_t, \Omega_t; \theta, \gamma)$ for the entire sample period and through the MC sample, using:

   (i) $\{\{\widetilde{\Omega}_t^i\}_{t=1}^T\}_{i=1}^N$ for the measurement density;

   (ii) $\{\{\widetilde{\Omega}_t^i\}_{t=1}^T\}_{i=1}^N$, $\{\widetilde{S}_t^i\}_{t=1}^T\}_{i=1}^N$ and v for the state transition density;

   (iii) $\{\{\widetilde{\Omega}_t^i\}_{t=1}^T\}_{i=1}^N$, $\{\widetilde{S}_t^{*INI^i}\}_{t=1}^T\}_{i=1}^N$ and v^{*INI} for the importance sampler density.

2.2 Code lik_KChi_R

Code lik_KChi_R performs the EIS optimization by solving the backward sequence of ordinary least-square problems in (10):

2.2.1 Analyzing the first double loop *for* when $t = 1$:

a. First, N equal draws of S_{t_1} are computed for $t = 1$ using the same initial parameterization defined for $lik\_KChi\_initial$[1]. For the first iteration, the code also defines zero values to EIS parameters $\forall\, t$. That is, the first iteration consider samples of $\{\{\widetilde{\Omega}_t^i\}_{t=1}^T\}_{i=1}^N$ which still do not incorporate information on Ω_t contained in y_{t+1}. Thus, effects of EIS optimizations spread with the number of iterations. Thereupon, N draws of S_{t-1}^*, the scale matrix of the complete EIS sampler, are computed for period $t = 1$ via Eq. (17);

b. The knowledge of $\{\widetilde{S}_{t_0}^{*i}\}_{i=1}^N$ and v^* allows one to sample N trajectories of $\{\widetilde{\Omega}_{t_1}^i\}_{i=1}^N$ from the EIS sampler $\Omega_{t_1}|v^*, \Omega_{t_0} \sim W_k(v^*, S_{t_0}^*)$.

   Obs: (i) parameter v^* is equal to $v + 1$ for the first EIS iteration since $\gamma_{1,t} = 0\,\forall\, t$; and (ii) the subscript t_0 in $S_{t_0}^*$ is used to show that S_t^* at time $t = 1$ depends on the precision matrix Ω_{t_0}, which is the variable $V0 = VINI$ set in the code;

[1] It is noteworthy to mention that depending on the optimization function, one has to define "natural" interval limits for the persistence parameter d and the degrees of freedom v (d can be equal to $\frac{d^2}{1+d^2}$, and v are added to the number of observed variables K) in order to avoid optimization searching between -Inf and +Inf.

c. Set EIS regressors using the N simulated precision matrices $\{\widetilde{\Omega}_{t_1}^i\}_{i=1}^N$ and the natural logarithm of its determinant: $XX(i, :, t)$ for $t = 1$:

$$XX_{N \times Nb+1} = \begin{bmatrix} \widetilde{\Omega}_{11}^1 & \widetilde{\Omega}_{21}^1 & \widetilde{\Omega}_{31}^1 & \widetilde{\Omega}_{22}^1 & \widetilde{\Omega}_{32}^1 & \widetilde{\Omega}_{33}^1 & \ln|\widetilde{\Omega}^1| \\ \widetilde{\Omega}_{11}^2 & \widetilde{\Omega}_{21}^2 & \widetilde{\Omega}_{31}^2 & \widetilde{\Omega}_{22}^2 & \widetilde{\Omega}_{32}^2 & \widetilde{\Omega}_{33}^2 & \ln|\widetilde{\Omega}^2| \\ \vdots & & & \vdots & & & \vdots \\ \widetilde{\Omega}_{11}^N & \widetilde{\Omega}_{21}^N & \widetilde{\Omega}_{31}^N & \widetilde{\Omega}_{22}^N & \widetilde{\Omega}_{32}^N & \widetilde{\Omega}_{33}^N & \ln|\widetilde{\Omega}^N| \end{bmatrix}, \tag{19}$$

where Nb is the number of estimated parameters from the Γ_t matrix filled with $(K^2 + K)/2$ auxiliary parameters. Obs: $\ln|\widetilde{\Omega}^i|$ and the main diagonal parameters of the simulated precision matrices $\widetilde{\Omega}_{11}^i$, $\widetilde{\Omega}_{22}^i$ and $\widetilde{\Omega}_{33}^i$ are also compensated by 0.5.

For $t = 2, ..., T$:

d. Repeat procedure (a.) using the N precision matrices simulated at period $t-1$, $\{\widetilde{\Omega}_{t-1}^i\}_{i=1}^N$, to compute the N draws of S_t from subsequent period;

e. The natural logarithm of the integrating constant $\chi(\Omega_{t_1}; \gamma_{t_2})$ defined in Eq. (18) is computed as from $t = 2$. The integrating constant $\chi(\Omega_{t_0}; \gamma_{t_1})$ is not computed for $t = 1$ because only $\chi(\Omega_{t_1}; \gamma_{t_2})$ is required to smooth estimates of $\chi(\Omega_{t_0}; \gamma_{t_1})$. Remembering that $\gamma_{1,t} = 0$ holds $\forall\, t$ to compute $\chi(\Omega_{t-1}; \gamma_t)$. As the integrating constant $\chi(\Omega_{t-1}; \gamma_t)$ is used as regressand for EIS regressions, matrix $Y(i, t-1)_{N \times T-1}$ stores its values for $t = 2, ..., T$ following the structure below for $\forall\, t$:

$$Y_{t,\, N \times 1} = \begin{bmatrix} \ln\left(\chi(\widetilde{\Omega}_{t-1}^1; \gamma_t^1)\right) \\ \ln\left(\chi(\widetilde{\Omega}_{t-1}^2; \gamma_t^2)\right) \\ \vdots \\ \ln\left(\chi(\widetilde{\Omega}_{t-1}^N; \gamma_t^N)\right) \end{bmatrix}. \tag{20}$$

2.2.2 Analyzing the second loop *for*:

- Using matrices $X = [1_{N \times 1}\ XX]$ and Y defined above for $\forall\, t = 1, ..., T$ and $\forall\, t = 2, ..., T$, respectively, this loop performs the backward EIS regressions described in (10) to estime EIS parameters collected by the $Nb + 1 \times 1$ vector $beta_t$,

$$beta_t = \begin{bmatrix} \hat{\gamma}_{2,t} \\ \hat{\gamma}_{3,t} \\ \vdots \\ \hat{\gamma}_{7,t} \\ \hat{\gamma}_{1,t} \end{bmatrix}. \tag{21}$$

Note that EIS regressions are performed for $T - 1$ periods in order to estimate $\{beta_t\}_{t=1}^{T-1}$. The EIS regressions are represented as follows:

$$\begin{aligned} Y_{t_2,\, N \times 1} &= XX_{t_1,\, N \times Nb+1} \cdot beta_{t_1,\, Nb+1 \times 1} + \varepsilon_{t_1,\, N \times 1} \\ &\dots \\ Y_{T,\, N \times 1} &= XX_{T-1,\, N \times Nb+1} \cdot beta_{T-1,\, Nb+1 \times 1} + \varepsilon_{T-1,\, N \times 1} \end{aligned} \tag{22}$$

The process of transferring back the integrating constant works like a smoother bringing information on Ω_t in y_{t+1}, because the aforementioned regressions relate period-$t+1$ integrating constant to period-t matrix XX. Thus, parameters $\hat{\gamma}_t$ contain information on $\ln\left(\chi(\widetilde{\Omega}_t; \gamma_{t+1})\right) \propto |S_t^*|^{\frac{v^*}{2}} / |S_t|^{\frac{v}{2}}$, which in turn carries information on Ω_t contained within y_{t+1} through $S_t^* = [y_{t+1} \cdot y_{t+1}' + S_t^{-1} + \Gamma_{t+1}]^{-1}$.

2.2.3 Convergence of EIS estimates are checked in order to continue or stop iterations. EIS iterations continues while *diff*>*tol*, where *diff* equals the sum of individual differences between estimated parameters $\hat{\gamma}_{2,t}, ..., \hat{\gamma}_{7,t}$ and their values at time $t-1$. A maximum number of iterations, represented by *itmax*, is also set.

2.2.4 If EIS iterations continue, the next i^{th} iteration does not use zero values to EIS auxiliary parameters $(\gamma_{1,t}, ..., \gamma_{7,t})$ anymore; it uses the estimated values from the last iteration. That is, both the degrees of freedom v^* and the scale matrix of the EIS sampler incorporate the $i-1^{th}$ EIS parameters $\hat{\gamma}_{1,t}^{i-1}, ..., \hat{\gamma}_{7,t}^{i-1}$.

2.2.5 Once EIS iterations stop, samples from the optimized EIS sampler are simulated. That is, new N trajectories of the latent precision matrices for $\forall\, t$, $\{\{\widetilde{\Omega}_t^i\}_{t=1}^T\}_{i=1}^N$, are simulated using the optimized values for $\{\hat{\gamma}_t\}_{t=1}^T$. Ultimately, as described in item (d.) of subsection 2.1, code lik_KChi_R computes the mean of the log-likelihood of IS ratio $\varphi(y_t, \Omega_t; \theta, \gamma)$ from Eq. 7 for the entire sample period T and through the MC sample.

2.2.6 The processes described by functions $lik\_KChi\_initial$ and lik_KChi_R are computed j multiple times to analyze the MC sampling variance of $L(\theta|\underline{y})$ described in (8), where the seed value is the only parameter that changes between each simulation j.

2.3 Model Estimation

The estimation of parameters from model (1)-(3) can be performed using the *fminsearch* or *fmincon* function to minimize the negative of the log-likelihood computed in lik_KChi_R. The minimization problem attempts to find a minimizer vector containing parameters d, v, $A_i^{1/2}$ to the output of function lik_KChi_R.

2.4 Overflow and Underflow Problems

The last step of $lik\_KChi\_initial$ and lik_KChi_R codes is to compute the approximated log-likelihood described in Eq. (8). In order to overcome numerical problems such as overflow and underflow, the log-likelihood ratio is adjusted due to exponentiation. Thus, $\widetilde{L}_S(\theta|\underline{y}) = \frac{1}{N}\sum_{j=1}^N \sum_{t=1}^T \ln(r_t^j)$ from (8) is computed through:

- lik = mean(exp(sum(ratio-adj)));
- loglik = log(lik)+T·adj;

$$\text{(ratio-adj)}_{T\times N} = \begin{bmatrix} \ln(r_{t_1}^1) & \ln(r_{t_1}^2) & \cdots & \ln(r_{t_1}^N) \\ \ln(r_{t_2}^1) & \ln(r_{t_2}^2) & \cdots & \ln(r_{t_2}^N) \\ \vdots & \vdots & \vdots & \vdots \\ \ln(r_T^1) & \ln(r_T^2) & \cdots & \ln(r_T^N) \end{bmatrix} - \text{adj}_{T\times N},$$

where r_t^j is the likelihood ratio computed at time t for the j^{th} MC sample.

$$\text{sum(ratio-adj)}_{1\times N} = \left[\sum_{t=1}^T \ln(r_t^1) - T\cdot\text{adj} \quad \sum_{t=1}^T \ln(r_t^2) - T\cdot\text{adj} \quad \cdots \quad \sum_{t=1}^T \ln(r_t^N) - T\cdot\text{adj}\right]$$

$$\text{exp(sum(ratio-adj))}_{1\times N} = \left[e^{\sum_{t=1}^T \ln(r_t^1)}\cdot e^{-T\cdot\text{adj}} \quad \ldots \quad e^{\sum_{t=1}^T \ln(r_t^N)}\cdot e^{-T\cdot\text{adj}}\right]$$

$$\text{lik = mean(exp(sum(ratio-adj)))} = \frac{1}{N} e^{-T\cdot \text{adj}} \sum_{j=1}^{N} e^{\sum_{t=1}^{T} \ln(r_t^j)}$$

$$\text{loglik} = \ln\left(\frac{1}{N} e^{-T\cdot \text{adj}} \sum_{j=1}^{N} e^{\sum_{t=1}^{T} \ln(r_t^j)}\right) + T \cdot \text{adj}$$

$$\text{loglik} = \ln(e^{-T\cdot \text{adj}}) + \ln\left(\frac{1}{N} \sum_{j=1}^{N} e^{\sum_{t=1}^{T} \ln(r_t^j)}\right) + T \cdot \text{adj}$$

$$\text{loglik} = -T \cdot \text{adj} + \ln\left(\frac{1}{N} \sum_{j=1}^{N} e^{\sum_{t=1}^{T} \ln(r_t^j)}\right) + T \cdot \text{adj}$$

$$\text{loglik} = \ln\left(\frac{1}{N} \sum_{j=1}^{N} e^{\sum_{t=1}^{T} \ln(r_t^j)}\right)$$

References

Philipov, A. and M. E. Glickman (2006). Multivariate stochastic volatility via wishart processes. *Journal of Business & Economic Statistics 24*(3), 313–328.