Sampling from VAR using Conditional SMC

Yuedan Huo

January 30, 2025

Vector Autoregression Model

Consider the following Vector Autoregression (VAR) model with N variables and lag p:

$$y_t = \Pi z_t + v_t, \tag{1}$$

$$v_t = A^{-1}\Lambda_t^{1/2}\epsilon_t, \quad \epsilon_t \sim N(0, I_N), \tag{2}$$

$$\log \lambda_t = \log \lambda_{t-1} + e_t, \quad e_t \sim N(0, \Phi). \tag{3}$$

where:

$z_t = \begin{bmatrix} 1 & y_{t-1}^T & \cdots & y_{t-p}^T \end{bmatrix}^T \in \mathbb{R}^{Np+1}$,

$\Pi \in \mathbb{R}^{N\times(Np+1)}$,

$\Lambda_t = \text{diag}(\lambda_t)$,

$A^{-1} \in \mathbb{R}^{N\times N}$, a lower triangular matrix with ones on the diagonal,

$\Phi \in \mathbb{R}^{N\times N}$, a diagonal matrix.

Aim

Draw from the posterior distribution of model parameters Π, A, Φ, and $\lambda_{1:T,1:N}$ by a Gibbs Sampler.

Bottle Neck

- With specific prior assumption, drawing from Π is drawing from a high dimensional Gaussian distribution, which is computationally complex.
- $\lambda_{1:T,1:N}$ is highly correlated and will result in slow mixing in a gibbs sampler.

We aim to resolve the second bottle neck by implementing CSMC, and compare with the method used in Carriero et al. [1].

Motivation

With the assumption that Φ is diagonal, the variance of the volatility for each variable follows an independent random walk:

$$\log \lambda_{j,t} = \log \lambda_{j,t-1} + e_{j,t}, \quad e_{j,t} \sim N(0, \Phi_{j,j}).$$

- Decompose the high-dimensional problem into a series of one-dimensional simulations, particle filter becomes a natural and heuristic choice.
- Compared to the method proposed in [2], which employs a Metropolis-Hastings algorithm to simulate each $\lambda_{t,j}$ sequentially for $t = 1, \ldots, T$ and $j = 1, \ldots, N$, using a particle filter effectively corresponds to a block update within a Gibbs sampler.

Metropolis-Hasting

Considering Markovianity,

$$\pi(\lambda_{t,j}|\lambda_{-t,j}, A, \Pi, y) \propto \left(\frac{1}{\lambda_{t,j}}\right)^{1.5} \exp\left\{-\frac{(A^{(j)T}(y_t - \Pi z_t))^2}{2\lambda_{t,j}}\right\}$$

$$\times \exp\left\{-\frac{(\log \lambda_{j,t} - m_{j,t})^2}{2\sigma^2_{j,t}}\right\},$$

where:

$$m_{j,t} = \frac{1}{2}(\log \lambda_{t-1,j} + \log \lambda_{t+1,j}),$$

$$\sigma^2_{j,t} = \frac{1}{2}\Phi_{j,j}.$$

Consider a Metropolis-Hasting sampler with target $\pi(\lambda_{t,j}|\lambda_{-t,j}, A, \Pi, y)$ and proposal

$$q(\lambda_{t,j}) \propto \frac{1}{\lambda^{1/2}_{t,j}} \exp\left\{-\frac{(A^{(j)T}(y_t - \Pi z_t))^2}{2\lambda_{t,j}}\right\}.$$

We sample from q in the log space.

Conditional SMC

With the assumption that Φ is a diagonal matrix, we can run a Conditional SMC for each $j = 1, \ldots, N$ independently. Consider the posterior distribution of $\log \lambda_{t,j}$:

$$\pi(\log \lambda_{:,j} | A, \Pi, y)$$
$$\propto \prod_{t=1}^{T} \exp\left\{ -\frac{(\log \lambda_{t,j} - \log \lambda_{t-1,j})^2}{2\Phi_{j,j}} \right\} \frac{1}{\lambda_{t,j}^{1/2}} \exp\left\{ -\frac{(A^{(j)T}(y_t - \Pi z_t))^2}{2\lambda_{t,j}} \right\}.$$

Using the bootstrap filter, in each time step, we sample $\log \lambda_{t,j} \sim N(\log \lambda_{t-1,j}, \Phi_{j,j})$ and then weight the particle by $\frac{1}{\lambda_{t,j}^{1/2}} \exp\left\{ -\frac{(A^{(j)T}(y_t - \Pi z_t))^2}{2\lambda_{t,j}} \right\}$.
We use the trajectory from the last iteration in Gibbs sampler as the fixed particle in CSMC, and choose a new trajectory with backwards sampling.

Numerical Experiment

- Use FRED data from [3] to compare the Gibbs sampler proposed in [1] with the Gibbs sampler incorporating CSMC.
- Consider a medium-sized model with $N = 18$, $p = 13$, T = 646.
- The theorectical complexity of both algorithms is O(MNT), where M = number of particles in CSMC, and M = number of MH steps in MHMC. We compare the performance by fixing the same M = 30 for both algorithms.

Comparison of Mixing

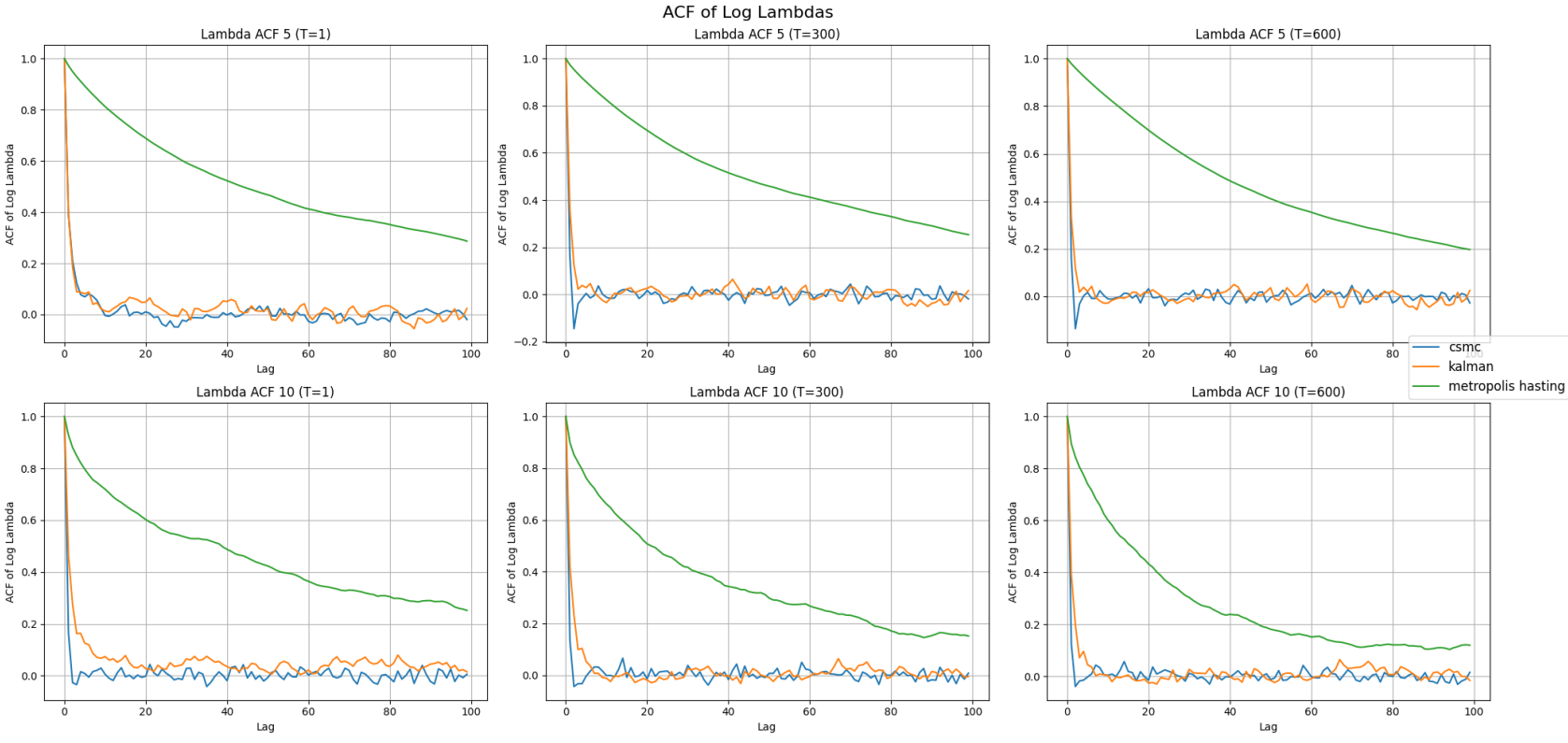


Figure: ACF of $\log \lambda_{1:T,1:N}$ in the full Gibbs samplers. Run for 2500 steps.

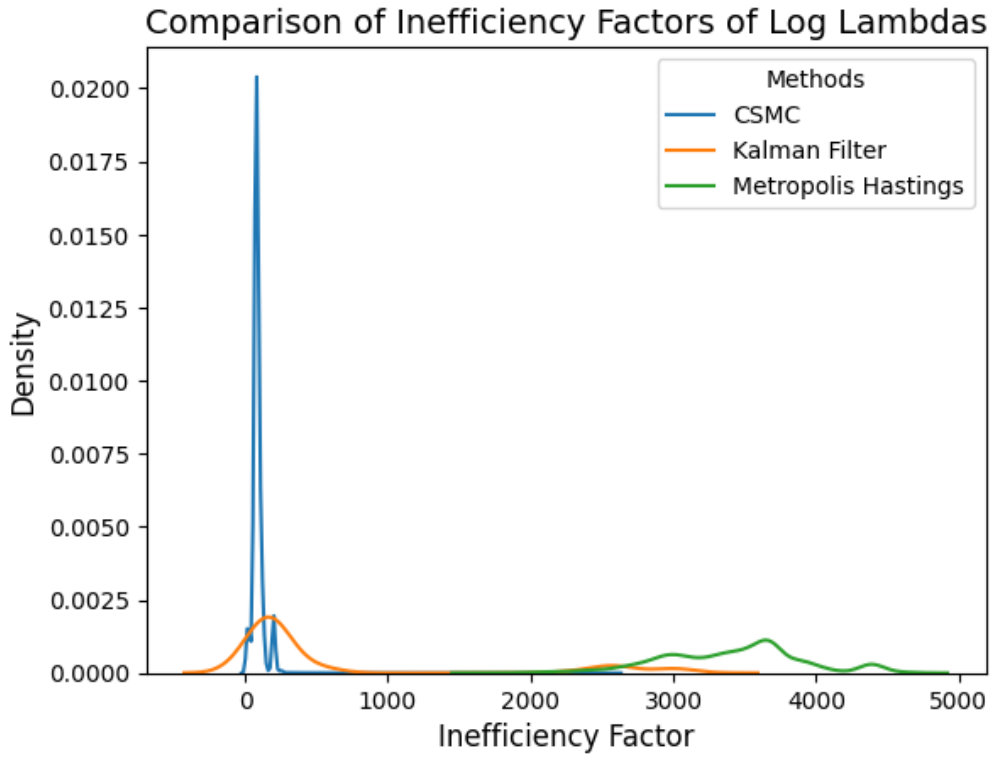
Comparison of Inefficiency

The inefficiency factor (IF) in MCMC quantifies the redundancy of correlated samples relative to independent samples. It is defined as:

$$\text{IF} = \frac{n_{samples}}{\text{ESS}}$$

where:

$\text{ESS} = \frac{n_{samples}}{1+2\sum_{k=1}^{\infty}\rho_k}$, ρ_k is the autocorrelation at lag k.



Comparison of CPU time

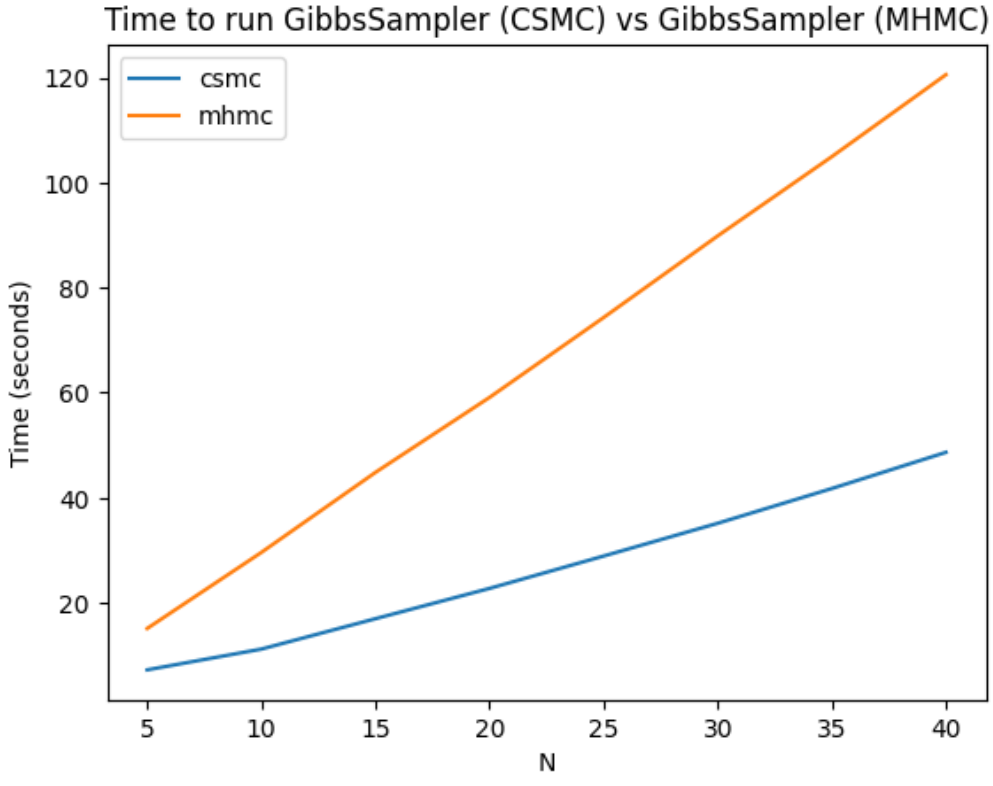


Figure: Compare the CPU time of drawing 10 iterations for both algorithms. p = 3, N range from 5 to 40, T = 646. Each run ten times and take average.

- Under same theorectical complexity, MHMC takes more CPU time to run. And the difference increases as N increases.

Comment on Number of Particles

In fact, even smaller number of particles can already result in fairly good updating rate in CSMC, in this specific model. Meaning that CSMC is even computationally cheaper than MHMC.

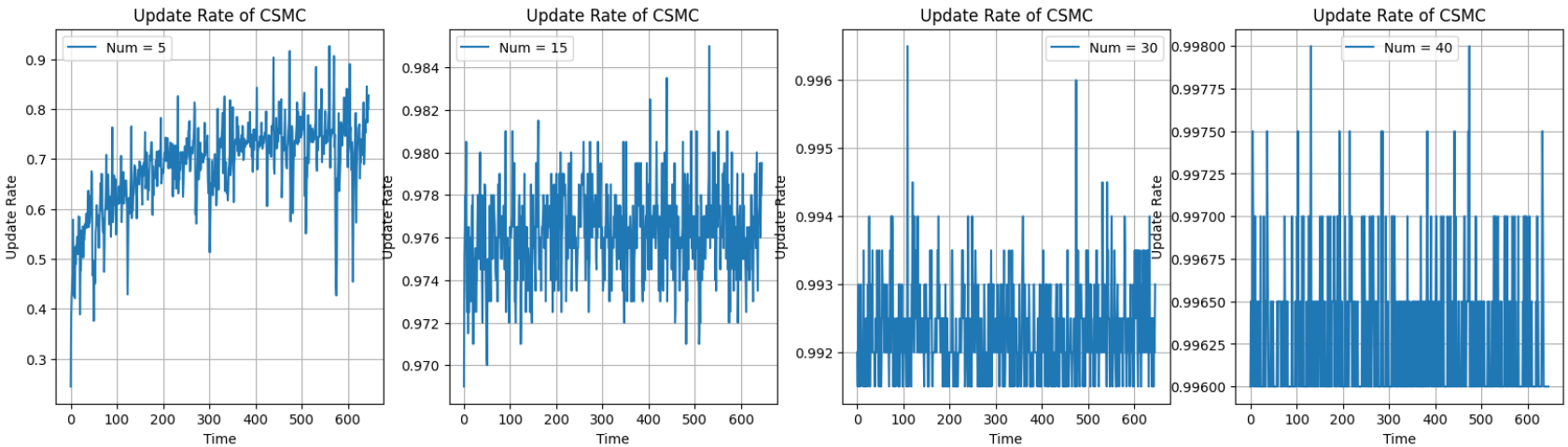


Figure: Running CSMC step alone for 2000 iterations. Compute the rate that fixed particles are not chosen in every time step. N = 1, T = 646.

Comments on Backwards Sampling

- To mitigate the effects of path degeneracy and allow particles to explore the state space effectively, backward sampling is often employed.
- In this specific example, the update rate remains notably high even without backward sampling.
- CPU time might be further reduced without the weighting and resampling steps in Backwards Sampling.

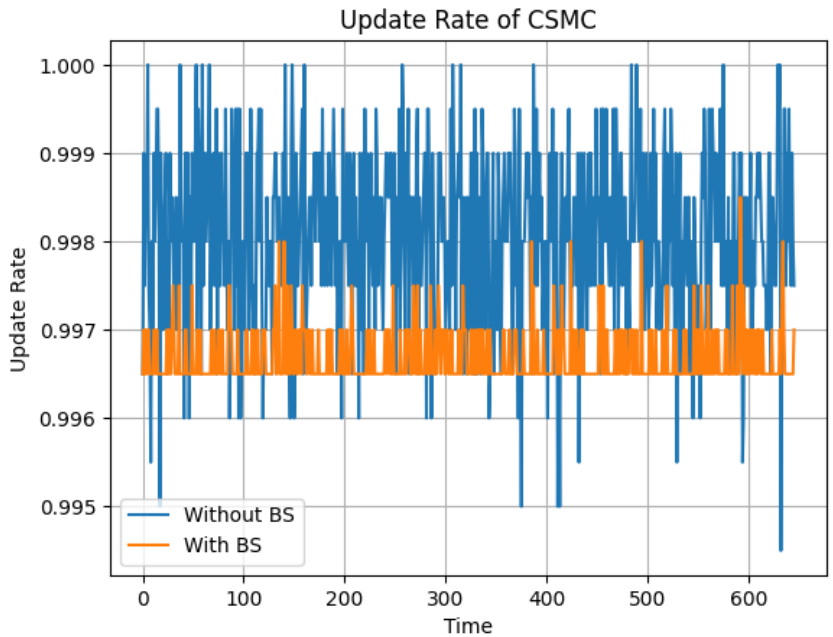


Figure: Running CSMC step alone for 2000 iterations. particles $= 30$.

Kalman Filter

- Approximate the distribution of log volatility by a mixture of Gaussian and do Kalman Filter. Mainly used in the case where Φ is not diagonal.
- Similar mixing, but less efficient compared to CSMC.
- computational complexity is $\mathbf{O}(TN^2)$. In contrast, using CSMC, the complexity reduces to $\mathbf{O}(MNT)$, where M is the number of particles.


Carriero, A., Clark, T. E., & Marcellino, M. (2019). Large Bayesian vector autoregressions with stochastic volatility and non-conjugate priors. *Journal of Econometrics, 212*(1), 137–154. https://doi.org/10.1016/j.jeconom.2019.04.024


Cogley, T., & Sargent, T. J. (2005). Drifts and volatilities: monetary policies and outcomes in the post WWII US. Review of Economic Dynamics, 8(2), 262–302. https://doi.org/10.1016/j.red.2004.10.009


McCracken, M. W., Ng, S. (2015). FRED-MD: a monthly database for macroeconomic research. Journal of Business and Economic Statistics, 34(4), 574–589. https://doi.org/10.1080/07350015.2015.1086655


Kim, S., Shepherd, N., Chib, S. (1998). Stochastic Volatility: Likelihood Inference and Comparison with ARCH Models. The Review of Economic Studies, 65(3), 361–393. https://doi.org/10.1111/1467-937x.00050


Carter, C. K., Kohn, R. (1994). On Gibbs sampling for state space models. Biometrika, 81(3), 541. https://doi.org/10.2307/2337125