Efficient Estimation of the Stochastic Volatility Model with Leverage

Hosszejni, Darjus; Kastner, Gregor

Published: 01/01/2018

Document Version
Publisher's PDF, also known as Version of record

Efficient Estimation of the Stochastic Volatility Model with Leverage

Darjus Hosszejni, Gregor Kastner

Institute for Statistics and Mathematics

WIRTSCHAFTSUNIVERSITÄT WIEN VIENNA UNIVERSITY OF ECONOMICS AND BUSINESS

Stochastic Volatility Model with Leverage

Stochastic volatility (SV) models are an increasingly popular choice for modelling financial return data. The basic SV model assumes an autoregressive structure for the log-volatility, and it is able to match the empirically observable low serial autocorrelation in the return series, but high serial autocorrelation in the squared return series. The SV model with leverage (SVL) extends the SV model by allowing the return series and the increment of the log-volatility series to correlate. This correlation models a real world phenomenon, the asymmetric relationship between returns and their volatility. Formally, in the centered parameterization, for $t = 1, \dots, T$,

$$
\begin{aligned}
y_t &= \exp(h_t/2)\varepsilon_t, \\
h_{t+1} &= \mu + \varphi(h_t - \mu) + \sigma\eta_t, \qquad \text{(C)} \\
\operatorname{cor}(\varepsilon_t, \eta_t) &= \rho,
\end{aligned}
$$

with $(\varepsilon_t)_{t=1}^T, (\eta_t)_{t=1}^T \sim \mathcal{N}_T(0_T, \mathcal{I}_T)$ vectors. An equivalent specification can be obtained by substituting $\tilde{h}_t = (h_t - \mu)/\sigma$ into (C),

$$
\begin{aligned}
y_t &= \exp((\mu + \sigma\tilde{h}_t)/2)\varepsilon_t, \qquad \text{(NC)} \\
\tilde{h}_{t+1} &= \varphi\tilde{h}_t + \eta_t.
\end{aligned}
$$

Project

The goal of this project is to find an MCMC sampler for SVL that has a high effective sample size per unit time, and that can be used as a subroutine for more sophisticated hierarchical models. Adaptive methods and adaptation phases are hence not preferred.

Auxiliary sampler

In terms of sampling vol. efficiently, the state-of-the-art MCMC sampler is based on an auxiliary model developed in [3]. It transforms the observation equation to a linear form, and then approximates $(\log(\varepsilon_t^2), \eta_t)$ by a ten component mixture of bivariate Gaussian distributions. With the mixture components denoted by $s_t \in \{1, \dots, 10\}$, the auxiliary model in NC is

$$
\begin{aligned}
\log(y_t^2) &= \mu + \sigma\tilde{h}_t + m_{s_t}^{(1)} + v_{s_t}^{(1)} w_t, \\
\tilde{h}_{t+1} &= \varphi\tilde{h}_t + \sqrt{1-\rho^2} z_t \qquad \text{(A)} \\
&\quad + \operatorname{sgn}(y_t)\rho\left(m_{s_t}^{(2)} + v_{s_t}^{(2)} w_t\right),
\end{aligned}
$$

where $w_t, z_t \sim$ i.i.d. $\mathcal{N}(0,1)$, and $m_{s_t}^{(i)}, v_{s_t}^{(i)}$ are precalculated constants. An independent Metropolis-Hastings (MH) step is used for $(\varphi, \rho, \sigma^2)$, and then Gaussian simulation smoothing for μ and the vector $\boldsymbol{h}$.

Direct sampler

Direct estimation of (C) or (NC) is also possible since $p(\varphi, \rho, \sigma^2, \mu \mid \boldsymbol{y}, \boldsymbol{h})$, and also its derivatives can be evaluated. Due to the issues with the independent MH sampler for (A), we tried the random-walk MH (RWMH) approach, and the Metropolis adjusted Langevin algorithm (MALA) for parameter sampling, and, as an approximation, stayed with the efficient simulation smoother from the (A) sampler. This approximation can be corrected by an MH acceptance-rejection step.

As already shown for the basic SV model [2], samplers based on different parameterizations can have substantially different sampling efficiency on the same data set due to the altered dependence structure. To exploit this phenomenon, the ancillarity-sufficiency interweaving strategy (ASIS) [5] can utilize samplers of both C and NC, and thus ASIS may be able to deliver a markedly higher effective sample size than C or NC samplers. ASIS affects only dependent MH algorithms, hence we took advantage of it in the RWMH and the MALA samplers.

Stan & JAGS

For completeness, Stan [1] and JAGS [4] were also tried out through their R interface in both the C and the NC parameterizations.

Setup

The samplers below were run on an extensive grid of parameters, altogether 91500 different MCMC chains were produced. The length of the burn-in was 5000, and 50000 samples were drawn afterwards. The initial values were the true ones in all cases, and the priors were always

$$
\begin{aligned}
(\varphi + 1)/2 &\sim \text{Beta}(20, 1.5), \\
(\rho + 1)/2 &\sim \text{Beta}(3, 5), \\
\sigma^2 &\sim \text{Gamma}(0.5, 0.5), \\
\mu &\sim \mathcal{N}(-10, \sqrt{10}^2), \\
h_1 &\sim \mathcal{N}(\mu, \sigma^2/(1-\varphi^2)).
\end{aligned}
$$

Auxiliary model

- Draw $\boldsymbol{s} \mid \boldsymbol{y}, \boldsymbol{h}, \varphi, \rho, \sigma, \mu$
  - Using inverse transform sampling
- Draw $\varphi, \rho, \sigma^2 \mid \boldsymbol{y}, \boldsymbol{s}$
  - Collapsed sampler: $\boldsymbol{h}$ and μ are integrated out
  - MH step with a 3D truncated Gaussian proposal that approximates $p(\varphi, \rho, \sigma \mid \boldsymbol{y}, \boldsymbol{s})$ around its mode
  - Includes Kalman filter, numerical optimization and differentiation
- Draw $\boldsymbol{h}, \mu \mid \boldsymbol{y}, \boldsymbol{s}, \varphi, \rho, \sigma$
  - Using Gaussian simulation smoothing

Exact model & ASIS

- Draw $\boldsymbol{h} \mid \boldsymbol{y}, \varphi, \rho, \sigma$
  - Using the AUX model
- Draw $\varphi, \rho, \sigma^2, \mu \mid \boldsymbol{y}, \boldsymbol{h}$
  - RWMH: 4D uncorr. Gaussian random walk in an unbounded space
  - MALA: RWMH's proposal shifted by the posterior's scaled gradient
- If ASIS
  - Calculate $\tilde{\boldsymbol{h}}$ using the new σ^2, μ values
  - Redraw $\varphi, \rho, \sigma^2, \mu \mid \boldsymbol{y}, \tilde{\boldsymbol{h}}$
  - Move back to $\boldsymbol{h}$ using the new σ^2, μ values

Results

Figure: Boxplot showing the execution times of the samplers after the burn-in. Different samplers react differently to changes in the parametereziation or changes in the true parameter.

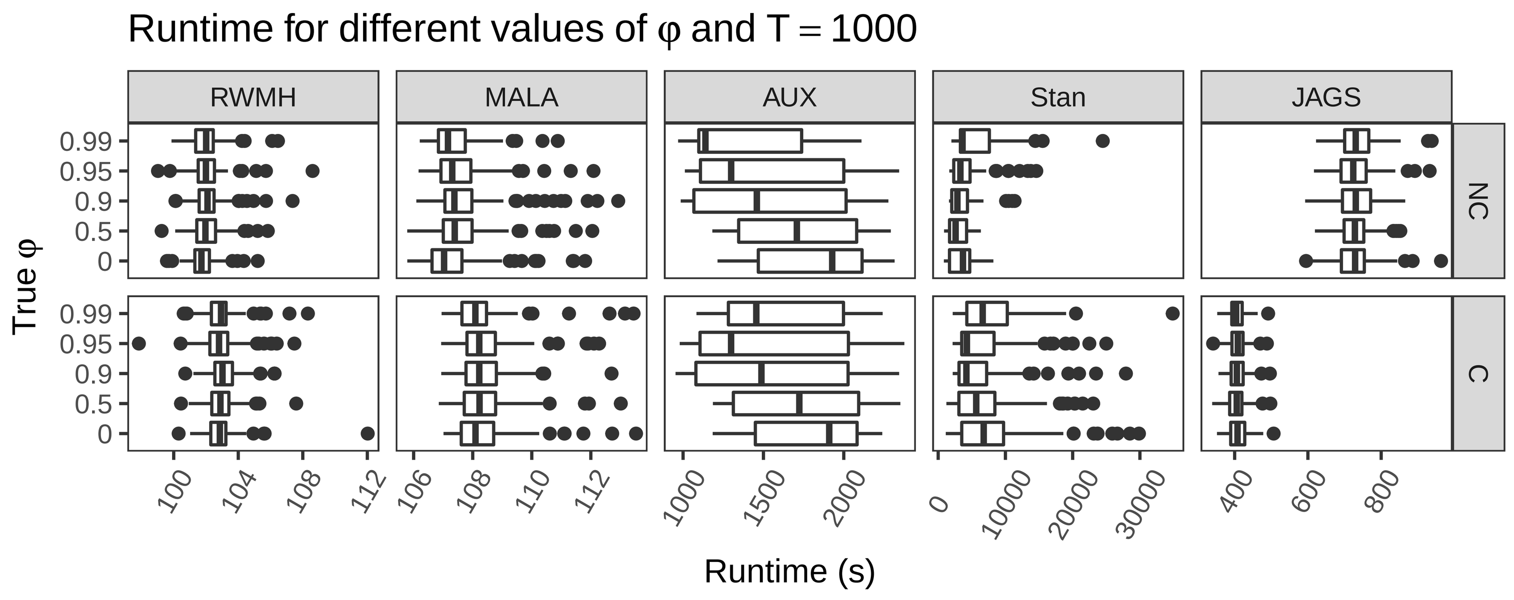
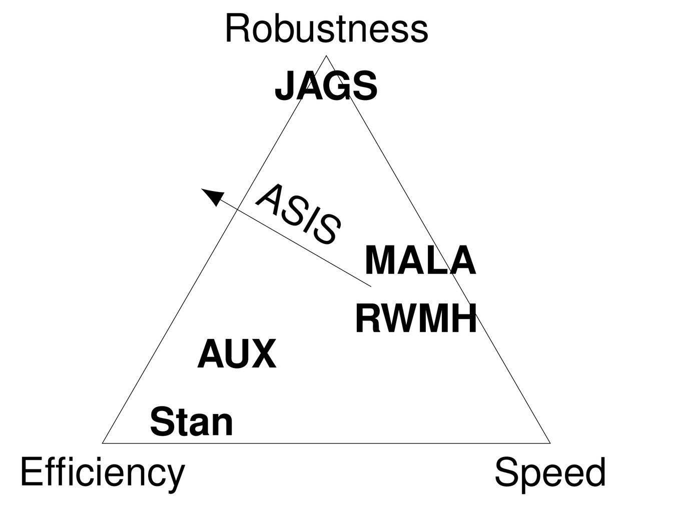


Figure: Triangle of sampler properties: Efficiency in terms of ESS, Speed in terms of runtime, and Robustness in terms of insensitiveness to different data.



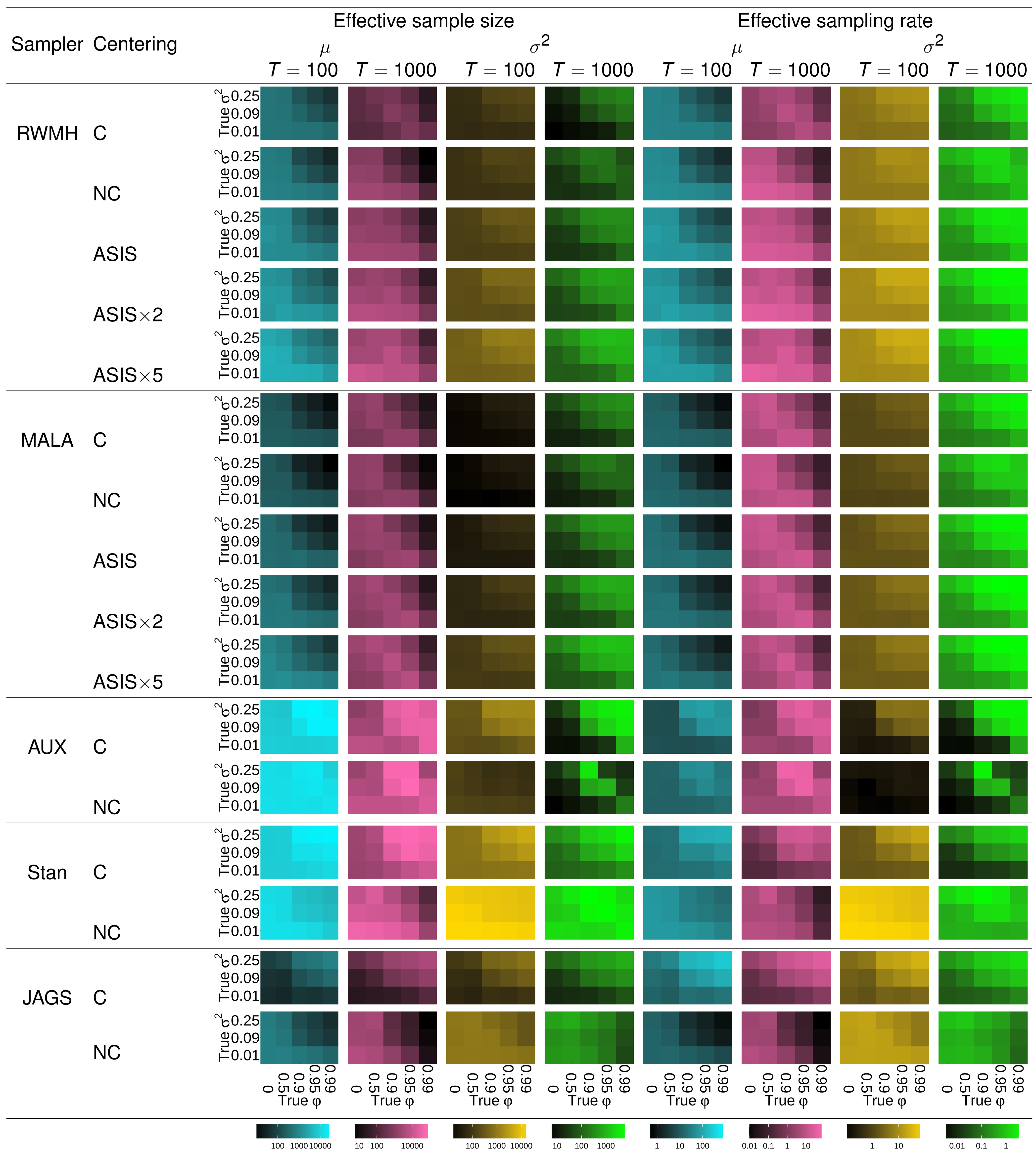


Table: 8 heatmaps are shown for each sampler and parameterization strategy. The first 4 illustrate effective sample sizes (ESS) of the drawn μ and σ^2 chains, respectively, grouped by the size of the data set. The second 4 columns of heatmaps showcase the effective sampling rates (ESR) in a similar fashion. Intuitively, the ESS is the number of i.i.d. draws, and it should ideally be 50000 here. Analogously, ESR is the number of i.i.d. draws per second. The plots are parceled out in a grid of the true data generating values of ϕ and σ^2. Finally, the grid point colors are on a logarithmic scale, and they map to the median of the 50 corresponding values.

Efficiency of φ and ρ

In general, the picture looks similar to the case of μ and σ^2. Stan outperforms all other choices in terms of ESS. If runtime is also of concern, then RWMH is the strongest choice for small data sets, while MALA together with RWMH show the best performance for larger data sets.

Efficiency of the volatility

Interestingly, the general framework of Stan is able to deliver the highest ESSs, slightly outperforming on average even the model-specific, optimized AUX sampler. RWMH with ASISx5 has 4 to 8 times smaller ESS for the latent vector than Stan in most cases. In terms of ESR, Stan and JAGS are the least favorable, while RWMH and MALA without ASIS perform ca. 10 times better than other choices. Since the Gaussian simulation smoother is a highly efficient algorithm both in terms of speed and sampling efficiency, the computation times of the $(\varphi, \rho, \sigma^2, \mu)$ draws becomes a crucial factor, in which RWMH excels the most.

Effect of reparameterization and ASIS

According to the Table, the sampling efficiency of JAGS, Stan, and AUX greatly depends on the parameterization and the data generating process. This effect is observable only in a weaker form at RWMH and MALA. On the other hand, ASIS increases both the ESS and the ESR, and we recommend ASIS for more reliable performance as well.

Future

The most promising algorithms will be included in the R package `stochvol` as a computationally highly efficient, compiled extension to the basic SV model sampler.

References

[1] B. Carpenter, A. Gelman, M. Hoffman, D. Lee, B. Goodrich, M. Betancourt, M. Brubaker, J. Guo, P. Li, and A. Riddell. Stan: A Probabilistic Programming Language. *Journal of Statistical Software, Articles*, 76(1):1–32, 2017.

[2] G. Kastner and S. Frühwirth-Schnatter. Ancillarity-sufficiency interweaving strategy (ASIS) for boosting MCMC estimation of stochastic volatility models. *Computational Statistics and Data Analysis*, 76:408–423, 2014.

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

[4] M. Plummer. JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling. In *Proceedings of the 3rd international workshop on distributed statistical computing*, volume 124, page 125, 2003.

[5] Y. Yu and X.-L. Meng. To Center or Not to Center: That Is Not the Question—An Ancillarity–Sufficiency Interweaving Strategy (ASIS) for Boosting MCMC Efficiency. *Journal of Computational and Graphical Statistics*, 20(3):531–570, 2011.

darjus.hosszejni@wu.ac.at