WIRTSCHAFTS
UNIVERSITÄT
WIEN VIENNA
UNIVERSITY OF
ECONOMICS
AND BUSINESS

Approaches toward the Bayesian Estimation of the Stochastic Volatility Model with Leverage

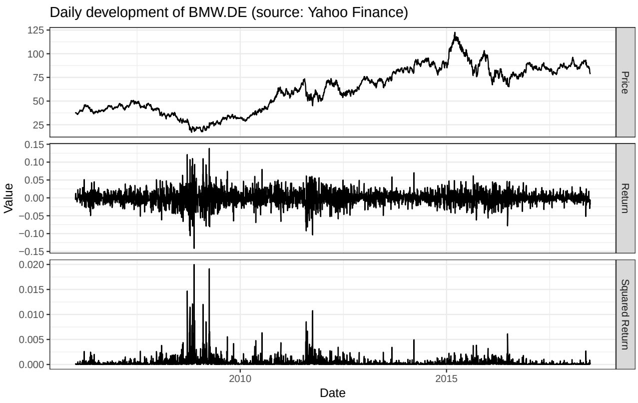
Darjus Hosszejni, Gregor Kastner

December 15, 2019

Motivation

Model volatility

- High autocorrelation
- (Usually) negative correlation with returns

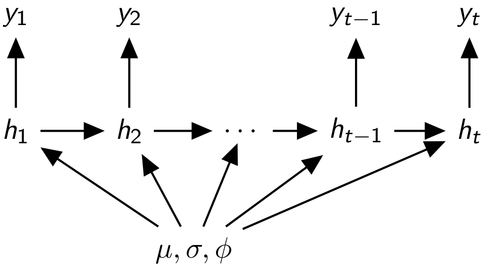


Formulation: SV [Taylor, 1986]

Stochastic volatility model without leverage:

$$
\begin{aligned}
y_t &= e^{\frac{h_t}{2}} \varepsilon_t, \\
h_{t+1} &= \mu + \phi(h_t - \mu) + \sigma \eta_t, \\
\mathrm{cor}(\varepsilon_t, \eta_t) &= 0,
\end{aligned}
$$

with $\varepsilon_t, \eta_t \sim$ i.i.d. $\mathcal{N}(0, 1)$.

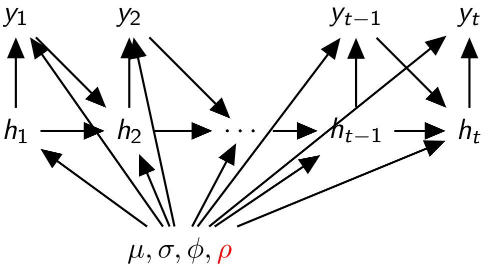


Formulation: SVL [Harvey, Shephard, 1996]

Stochastic volatility model with leverage:

$$
\begin{aligned}
y_t &= e^{\frac{h_t}{2}} \varepsilon_t, \\
h_{t+1} &= \mu + \phi(h_t - \mu) + \sigma \eta_t, \\
\mathrm{cor}(\varepsilon_t, \eta_t) &= \rho,
\end{aligned}
$$

with $\varepsilon_t, \eta_t \sim$ i.i.d. $\mathcal{N}(0, 1)$.



Problem

The non-linear relationship.

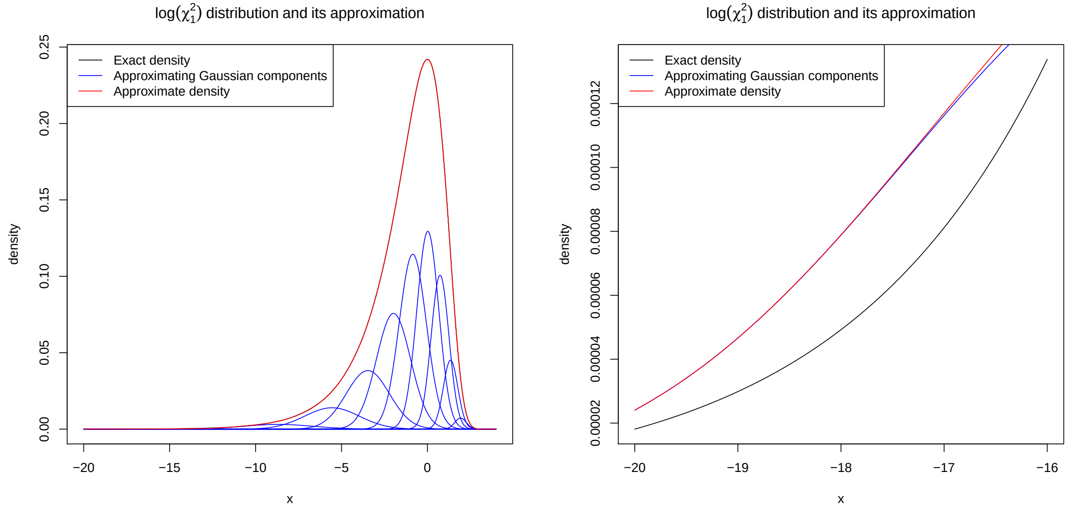
$$y_t = e^{\frac{h_t}{2}} \varepsilon_t,$$
$$h_{t+1} = \mu + \phi(h_t - \mu) + \sigma \eta_t.$$

- Sampling $\boldsymbol{h}$, either
  - Statistically inefficient (e.g., single-move Gibbs), or
  - Really slow (e.g., particle filter)

Linearization:

$$\log(y_t^2) = h_t + \log(\varepsilon_t^2),$$
$$h_{t+1} = \mu + \phi(h_t - \mu) + \sigma \eta_t.$$

Now problems with the $\log(\chi_1^2)$ distribution


log(χ_1^2) distribution and its approximation
Exact density
Approximating Gaussian components
Approximate density
density
x
0.00
0.05
0.10
0.15
0.20
0.25
−20
−15
−10
−5
0
log(χ_1^2) distribution and its approximation
Exact density
Approximating Gaussian components
Approximate density
density
x
0.00002
0.00004
0.00006
0.00008
0.00010
0.00012
−20
−19
−18
−17
−16

Auxiliary Mixture Repres. [Omori, et al., 2007]

Let $s_t \in \{1, \dots, 10\}$ denote the mixture components, then

$$\log(y_t^2) = h_t + m_{s_t}^{(1)} + v_{s_t}^{(1)} w_t,$$
$$h_{t+1} = \mu + \varphi(h_t - \mu) + \sigma\sqrt{1-\rho^2} z_t + \sigma \operatorname{sgn}(y_t)\rho \left(m_{s_t}^{(2)} + v_{s_t}^{(2)} w_t \right),$$

where $w_t, z_t \sim$ i.i.d. $\mathcal{N}(0,1)$, and $m_{s_t}^{(i)}, v_{s_t}^{(i)}$ are model-independent constants.

- 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 Laplace approx. of $p(\varphi, \rho, \sigma \mid \boldsymbol{y}, \boldsymbol{s})$ as a proposal
  - Includes Kalman filter, **numerical optimization** and differentiation
- Draw $\boldsymbol{h}, \mu \mid \boldsymbol{y}, \boldsymbol{s}, \varphi, \rho, \sigma$
  - Using Gaussian simulation smoothing

Pros & Cons

Pro: good proposal distribution

- High effective sample size for all parameters

Con: numerical optimization

- Sensitive to optimizer
- Sensitive to data generating process (DGP)
- Up to 80% of the runtime (in C++)

Solution idea

Step back, and use simple but computationally cheap proposals

- Random-walk Metropolis-Hastings (RWMH)
- Metropolis Adjusted Langevin Algorithm (MALA)
- Ancillarity-Sufficiency Interweaving Strategy (ASIS) [Yu, Meng, 2011]
  - Sample the parameters twice, based on two different parameterizations

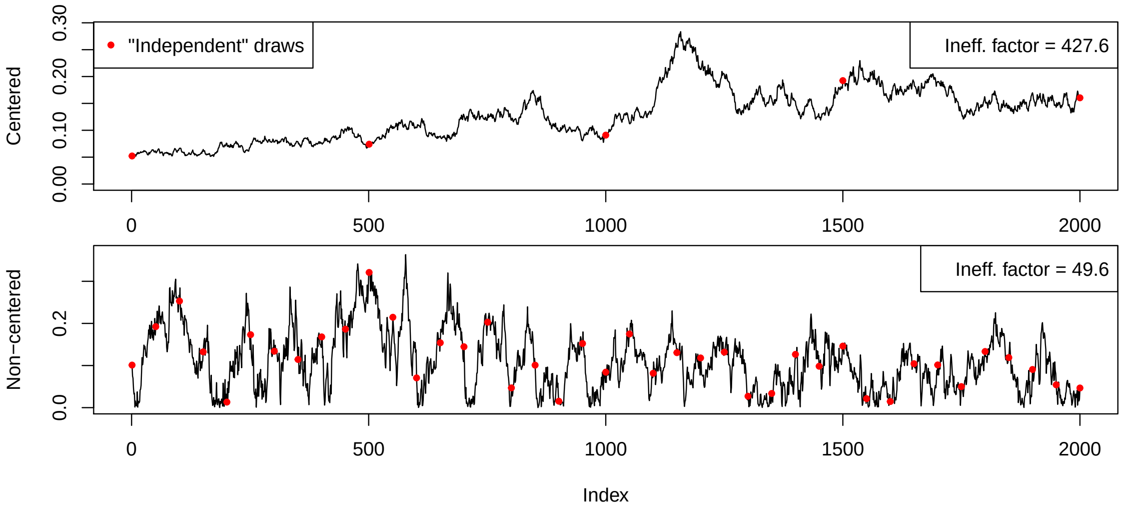
Centered parameterization:

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

And the non-centered, by substituting $\tilde{h}_t = (h_t - \mu)/\sigma$ into (C), i.e.,

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

Traceplot of σ, same data, different samplers. Bad case for the centered.



Exact Samplers & ASIS

Let $\boldsymbol{\vartheta} = (\varphi, \rho, \sigma^2, \mu)'$.

- Draw $\boldsymbol{h} \mid \boldsymbol{y}, \boldsymbol{\vartheta}$
  - Using the auxiliary model as proposal, then correcting with an MH acceptance-rejection step
- Draw $\boldsymbol{\vartheta} \mid \boldsymbol{y}, \boldsymbol{h}$
  - 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 values of σ^2, μ
  - Redraw $\boldsymbol{\vartheta} \mid \boldsymbol{y}, \tilde{\boldsymbol{h}}$
  - Move back to $\boldsymbol{h}$ using the new values of σ^2, μ

Draw the latent vector

Let $\boldsymbol{h}$ be the current latent vector, and $\boldsymbol{h}'$ be the proposed one by the auxiliary model. One has to calculate

$$\min\left\{1, \frac{p_{\text{orig}}\left(\boldsymbol{h}' \mid \boldsymbol{y}, \vartheta\right)}{p_{\text{orig}}\left(\boldsymbol{h} \mid \boldsymbol{y}, \vartheta\right)} \frac{p_{\text{aux}}\left(\boldsymbol{h} \mid \boldsymbol{y}, \vartheta\right)}{p_{\text{aux}}\left(\boldsymbol{h}' \mid \boldsymbol{y}, \vartheta\right)}\right\},$$

where $p_{\text{aux}}\left(\boldsymbol{h} \mid \boldsymbol{y}, \vartheta\right) = \sum_{\boldsymbol{s}\in\{1,\ldots,10\}^T} p_{\text{aux}}\left(\boldsymbol{h} \mid \boldsymbol{y}, \vartheta, \boldsymbol{s}\right)$, a sum of 10^T elements.
BUT actually, the auxiliary model is "not yet fully Bayesian":

- $\text{sgn}(\boldsymbol{y})$ is handled as a model-indep. constant,
- $p_{\text{aux}}(\boldsymbol{s} \mid \text{sgn}(\boldsymbol{y}), \vartheta)$ is not specified $\implies$ user input.

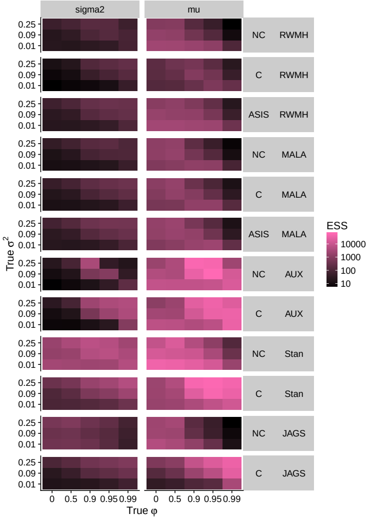
With a good choice, $O(10^T)$ can be reduced to $O(T)$.

Simulation study

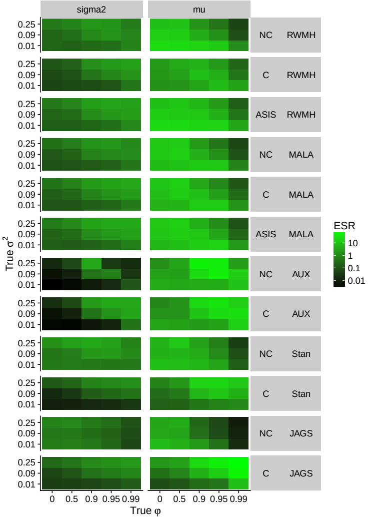
Goal: a sampler that works well on all reasonable datasets, and include it in `stochvol` [Kastner, 2016].

- 91500 MCMC chains on a grid of DGP setups
- RWMH, MALA, AUX for $(\varphi, \sigma^2, \rho, \mu)$
- Both parameterizations and ASIS
- Auxiliary model for $\boldsymbol{h}$
- Stan and JAGS as benchmarks
- Burn-in 5000 (enough), then chain 50000
- Initial values: true values
- Standard priors, the same for each run

Median effective sampling sizes for T=1000 per DGP

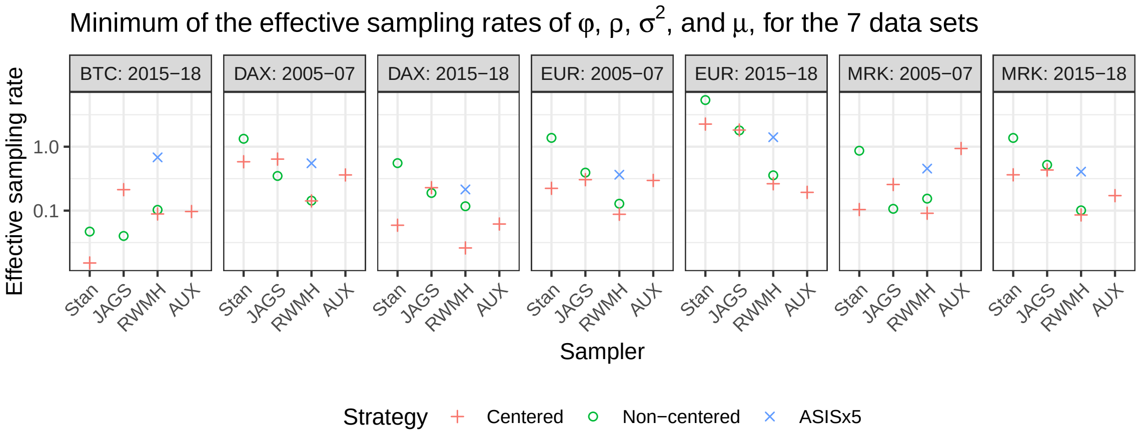


Median effective sampling rates for T=1000 per DGP



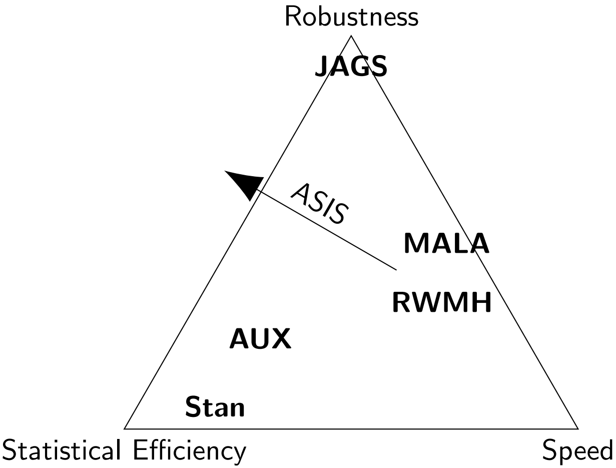
Application

- 7 data sets
- BTC, DAX, EURUSD, Merck KG
- 01/01/2005–31/12/2007 and 01/01/2015–30/06/2018
- Stan, JAGS, AUX, and RWMH with ASISx5



Minimum of the effective sampling rates of φ, ρ, σ², and μ, for the 7 data sets
BTC: 2015−18
DAX: 2005−07
DAX: 2015−18
EUR: 2005−07
EUR: 2015−18
MRK: 2005−07
MRK: 2015−18
Effective sampling rate
1.0
0.1
Stan
JAGS
RWMH
AUX
Sampler
Strategy
Centered
Non−centered
ASISx5

Triangle of Sampler Properties



References

D. Hosszejni and G. Kastner.
Approaches toward the Bayesian estimation of the stochastic volatility model with leverage.
In R. Argiento, D. Durante, and S. Wade, editors, *Bayesian Statistics and New Generations – Selected Contributions from BAYSM 2018*, volume 296 of *Springer Proceedings in Mathematics & Statistics*, pages 75–83. Springer, 2019.

D. Hosszejni and G. Kastner.
Modeling univariate and multivariate stochastic volatility in R with stochvol and factorstochvol.
arXiv preprint:1906.12123, 2019.

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.