A Bayesian Approach to Dealing With Heteroskedasticity in Time Series

Gregor Kastner

Department of Statistics
University of Klagenfurt, Austria

Univariate SV
- Review and Motivation
- Methods for MCMC Sampling

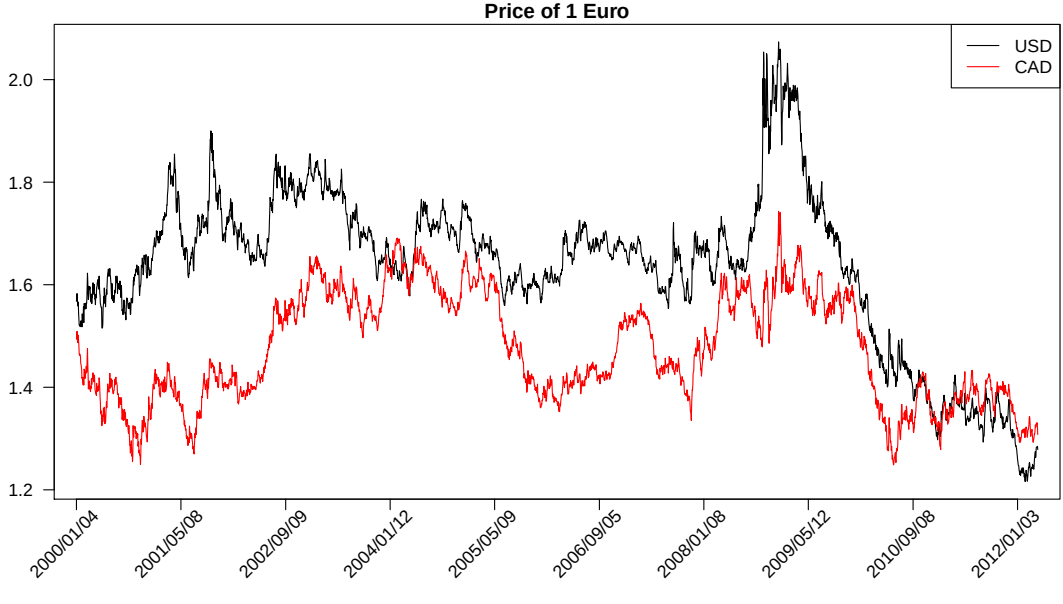
Efficiency Issues
- To Center or Not to Center?
- Interweaving the Parameterizations – ASIS
- Efficiency Comparison

The R Package stochvol
- Motivation
- Running the Sampler
- Assessing the Output

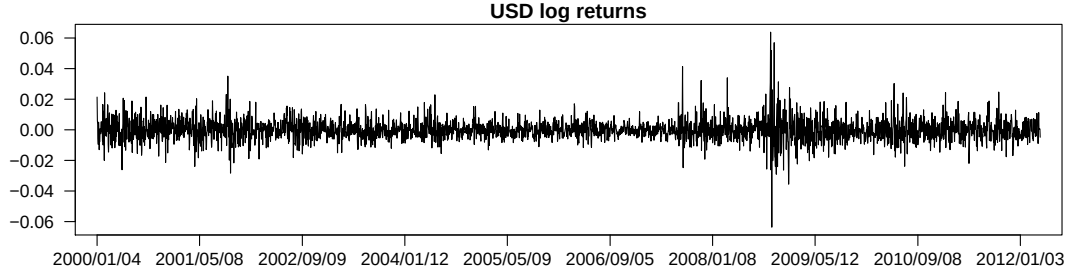
Model extensions
- Heavy Tails
- Leverage/Asymmetry

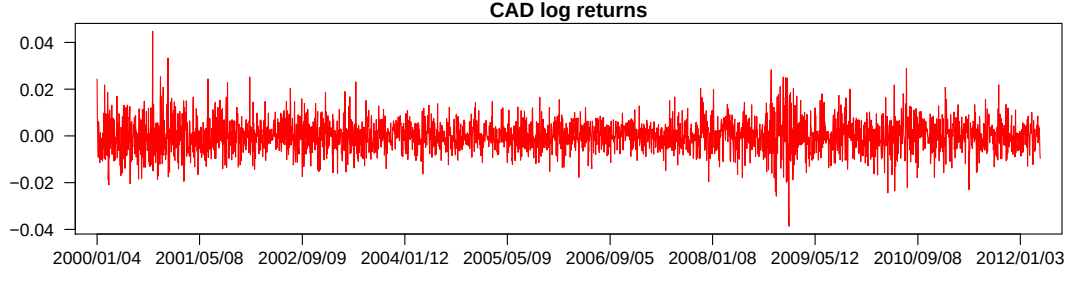
A Bayesian Predictive Framework
- Predictive Density, Predictive Likelihoods and Predictive Bayes Factors
- Exchange Rate Example

Two Exchange Rates (3140 daily observations)

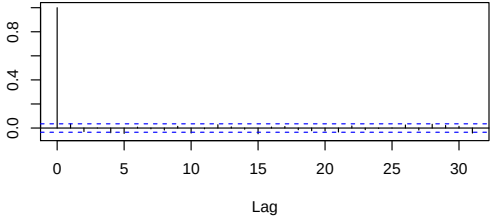


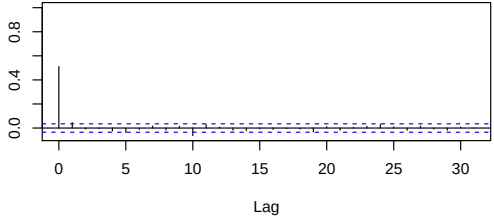
Log-returns


USD log returns
0.06
0.04
0.02
0.00
−0.02
−0.04
−0.06
2000/01/04
2001/05/08
2002/09/09
2004/01/12
2005/05/09
2006/09/05
2008/01/08
2009/05/12
2010/09/08
2012/01/03

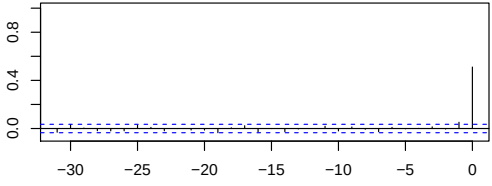

CAD log returns
0.04
0.02
0.00
−0.02
−0.04
2000/01/04
2001/05/08
2002/09/09
2004/01/12
2005/05/09
2006/09/05
2008/01/08
2009/05/12
2010/09/08
2012/01/03

Correlogram (log-returns)

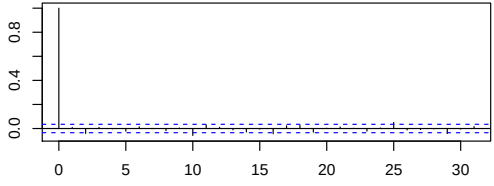




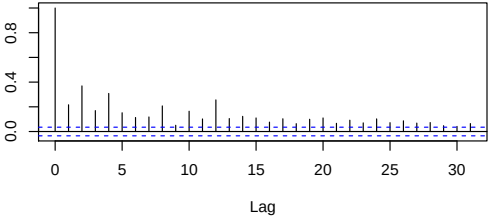
Canadian_dollar & Australian_dollar

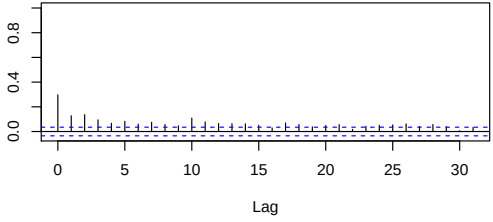


Canadian_dollar

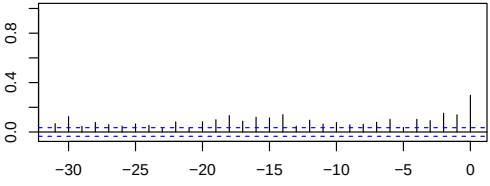


Correlogram (squared log-returns)

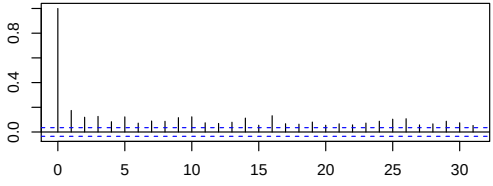




Canadian_dollar & Australian_dollar



Canadian_dollar



GARCH vs. SV

GARCH(1,1)

$$\begin{aligned} \tilde{y}_t &= \sigma_t \varepsilon_t, & \varepsilon_t \sim N(0,1)\,, \\ \sigma_t^2 &= \alpha_0 + \alpha_1 \tilde{y}_{t-1}^2 + \beta \sigma_{t-1}^2, \end{aligned}$$

with σ_0, α_0, α_1 and β unknown.

GARCH vs. SV

GARCH(1,1)

$$\begin{aligned} \tilde{y}_t &= \sigma_t \varepsilon_t, & \varepsilon_t &\sim N(0,1), \\ \sigma_t^2 &= \alpha_0 + \alpha_1 \tilde{y}_{t-1}^2 + \beta \sigma_{t-1}^2, \end{aligned}$$

with σ_0, α_0, α_1 and β unknown.

SV

$$\begin{aligned} \tilde{y}_t &= e^{\frac{h_t}{2}} \varepsilon_t, & \varepsilon_t &\sim N(0,1), \\ h_t &= \mu + \phi(h_{t-1} - \mu) + \sigma \eta_t, & \eta_t &\sim N(0,1), \end{aligned}$$

with $\mathsf{h} = (h_0, \ldots, h_T)$, μ, ϕ and σ unknown.

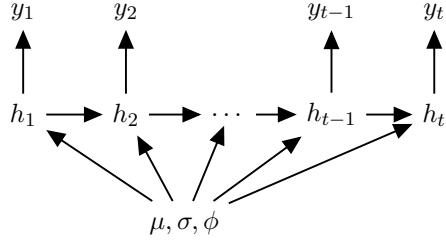
Stochastic Volatility Models ...

- arise as discrete approximations to various diffusion processes in continuous-time asset pricing.
- shares many "nice" properties with (X-)GARCH ...
- ... while still showing important differences, e.g. a more realistic ACF of squared returns.
- shows less/no remaining nonlinear dependencies in residuals (Jacquier et al., 1994).
- have a likelihood that is difficult to evaluate.

Stochastic Volatility Models . . .

- arise as discrete approximations to various diffusion processes in continuous-time asset pricing.
- shares many "nice" properties with (X-)GARCH . . .
- . . . while still showing important differences, e.g. a more realistic ACF of squared returns.
- shows less/no remaining nonlinear dependencies in residuals (Jacquier et al., 1994).
- have a likelihood that is difficult to evaluate. Solutions:
  - MM (no score function available, relatively inefficient)
  - approximate linear filtering methods for a QML estimator
  - other methods for approximating the integral used in evaluating the likelihood (revival?)
  - Bayesian methods (mainly MCMC)

The SV model



Priors:

- $\mu \sim N(b_\mu, B_\mu)$
- $(\phi + 1)/2 \sim \mathcal{B}(a_0, b_0)$ as in Kim et al. (1998), implying

$$p(\phi) = \frac{1}{2B(a_0, b_0)} \left(\frac{1+\phi}{2}\right)^{a_0 - 1} \left(\frac{1-\phi}{2}\right)^{b_0 - 1}$$

- $\sigma^2 \sim B_\sigma \cdot \chi_1^2 = \mathcal{G}\left(\frac{1}{2}, \frac{1}{2B_\sigma}\right)$
- $h_0 | \mu, \phi, \sigma \sim N\left(\mu, \sigma^2/(1-\phi^2)\right)$

Sampling Methods

- Single move $(h_t|h_{-t}, \sigma, \phi, \mu, \mathrm{y})$: Jacquier et al. (1994)
- Multi move $(h_t|\sigma, \phi, \mu, \mathrm{y})$: Shephard (1994), Omori et al. (2007)
  - Possible through mixture approximation of $\log(\varepsilon_t^2)$ – see next slide
  - Requires FFBS techniques
- All WithOut a Loop: Rue (2001), McCausland et al. (2011), Kastner and Frühwirth-Schnatter (2014):
  - No FFBS needed
  - No need to invert (band diagonal) precision matrix Ω
  - Fast (band back-substitution)
  - Implemented in C, Matlab, R...

Linearizing and "Gaussianizing" through auxiliary mixture sampling

We like

- Linear models with
- Gaussian increments

Linearizing and "Gaussianizing" through auxiliary mixture sampling

We like

- Linear models with $\times$
- Gaussian increments $\checkmark$

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

Linearizing and "Gaussianizing" through auxiliary mixture sampling

We like

- Linear models with $\times$
- Gaussian increments $\checkmark$

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

Linearization: apply $x \to \log(x^2)$ to the first eq.

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

Linearizing and "Gaussianizing" through auxiliary mixture sampling

We like

- Linear models with ✓
- Gaussian increments ×

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

Approximate $\log(\varepsilon_t^2) \sim \log(\chi_1^2)$
by a mixture of Gaussians!

Linearizing and "Gaussianizing" through auxiliary mixture sampling

New variable: the mixture component indicator, $s_t \in 1, \ldots, 10$.
Prior: s_t is i.i.d. from a fixed discrete distribution.
Given the mixture components, s_t,

$$
\begin{aligned}
y_t^* &= h_t + \xi_t \\
h_{t+1} &= \mu + \phi(h_t - \mu) + \sigma\eta_t \\
\xi_t &\sim \mathcal{N}(m_{s_t}, v_{s_t}^2)
\end{aligned}
$$

Sampler. Repeat:

1. Draw $\boldsymbol{s}$ from $\boldsymbol{s} \mid \boldsymbol{y}^*, \boldsymbol{h}$ via inverse transform sampling
2. Draw $\boldsymbol{h}$ from $\boldsymbol{h} \mid \boldsymbol{y}^*, \boldsymbol{s}, \mu, \sigma, \phi$ "All WithOut a Loop" (AWOL)
3. Draw μ, σ, ϕ from $\mu, \sigma, \phi \mid \boldsymbol{y}^*, \boldsymbol{h}, \boldsymbol{s}$ via Bayesian regression

Univariate SV
Review and Motivation
Methods for MCMC Sampling

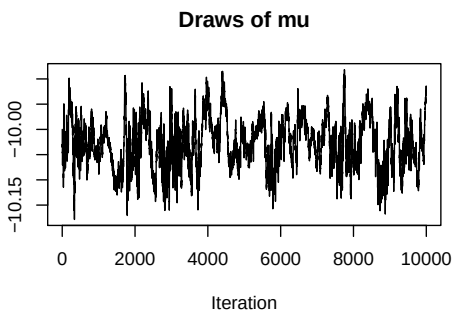
Efficiency Issues
To Center or Not to Center?
Interweaving the Parameterizations – ASIS
Efficiency Comparison

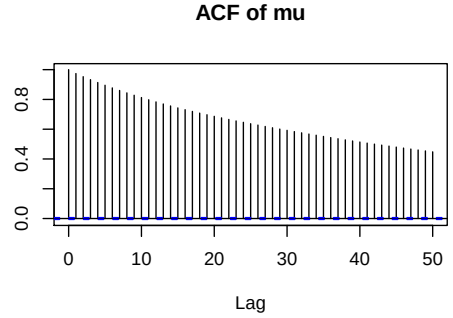
The R Package stochvol
Motivation
Running the Sampler
Assessing the Output

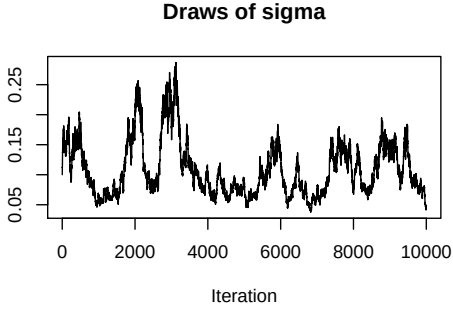
Model extensions
Heavy Tails
Leverage/Asymmetry

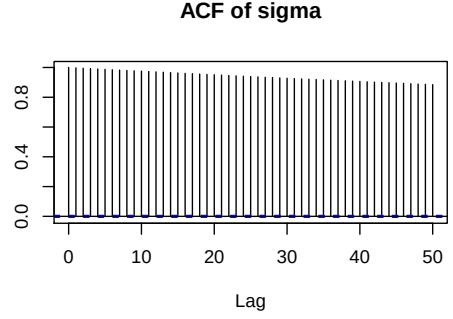
A Bayesian Predictive Framework
Predictive Density, Predictive Likelihoods and Predictive Bayes Factors
Exchange Rate Example

Traceplots and ACFs for Some Data ($\phi = 0.5, \sigma = 0.1$)









The Non-Centered Parameterization

Why bother looking at another version?

- Frühwirth-Schnatter (2004) and Frühwirth-Schnatter and Wagner (2010) show that MCMC sampling improves a lot by considering non-centered versions of a state space model...
- ...if the error variance in the state equation is considerably smaller than the error variance in the observation equation.
- We have $\sigma^2 \lesssim 0.2 \ll v_{s_t}^2$ for typical exchange rate or asset return data

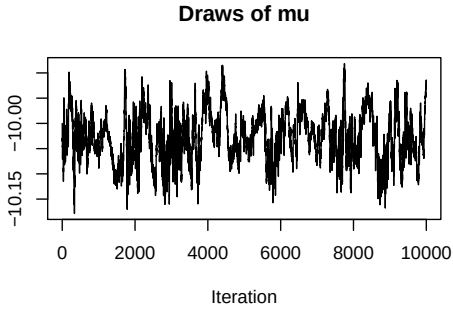
The Non-Centered Parameterization

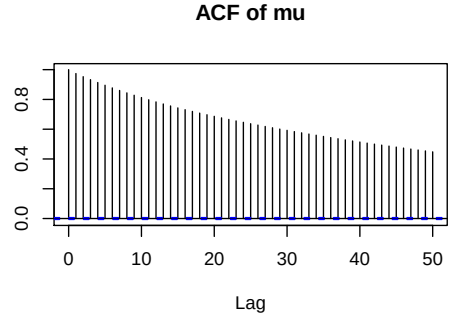
Why bother looking at another version?

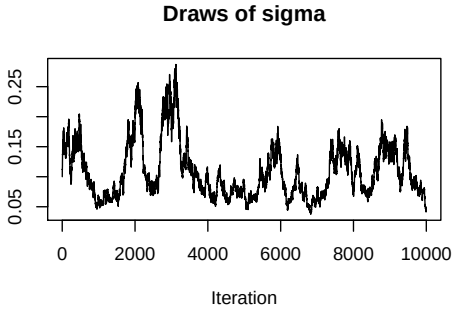
- Frühwirth-Schnatter (2004) and Frühwirth-Schnatter and Wagner (2010) show that MCMC sampling improves a lot by considering non-centered versions of a state space model...
- ...if the error variance in the state equation is considerably smaller than the error variance in the observation equation.
- We have $\sigma^2 \lesssim 0.2 \ll v_{s_t}^2$ for typical exchange rate or asset return data

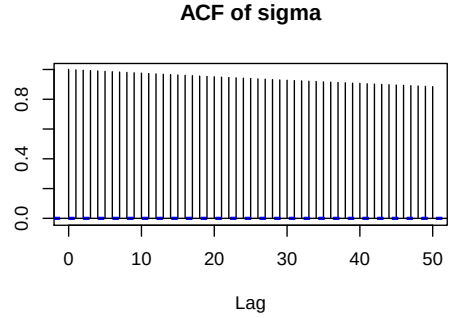
Centered		Non-centered
$y_t^* = h_t + \xi_t$	$\frac{h_t - \mu}{\sigma} = \tilde{h}_t$	$y_t^* = -\mu + \sigma\tilde{h}_t + \xi_t$
$h_{t+1} = \mu + \phi(h_t - \mu) + \sigma\eta_t$	$\Longleftrightarrow$	$\tilde{h}_{t+1} = \phi\tilde{h}_t + \eta_t$

C – Low Volatility of Volatility ($\phi = 0.5, \sigma = 0.1$)

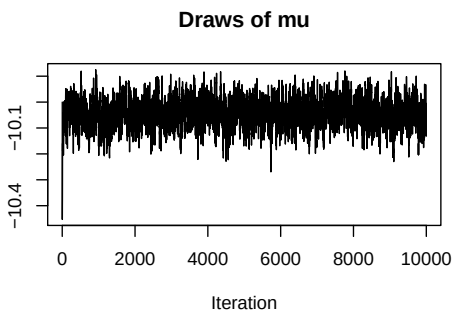


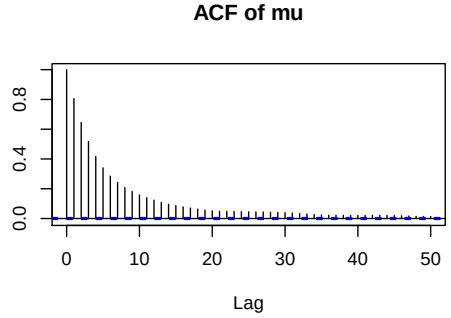


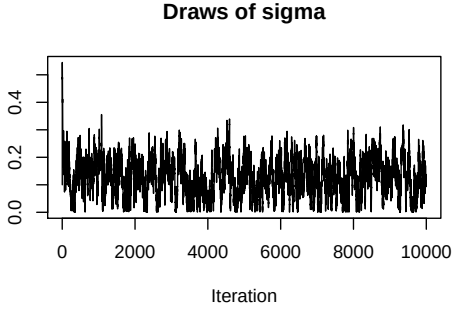


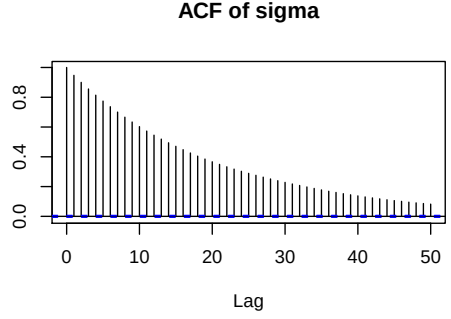


NC – Low Volatility of Volatility ($\phi = 0.5, \sigma = 0.1$)

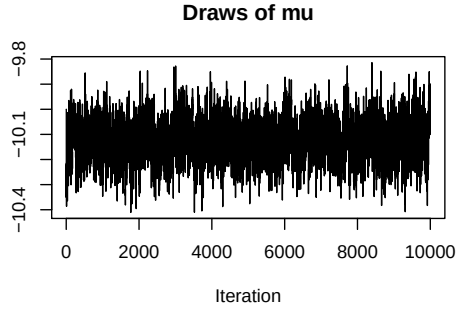


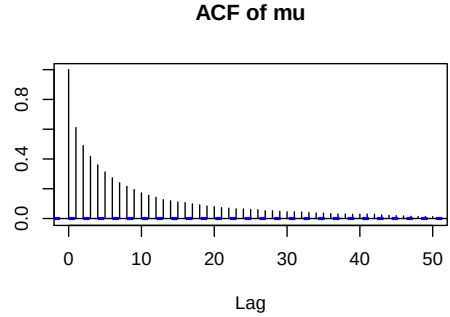


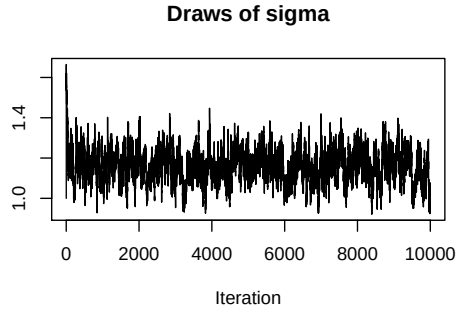


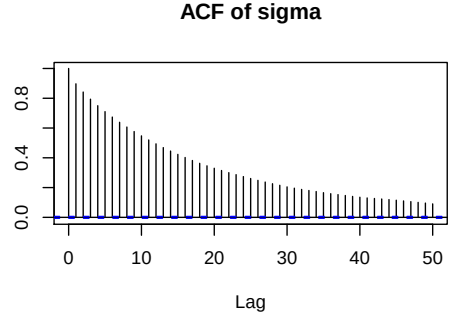


C – High Volatility of Volatility ($\phi = 0.5, \sigma = 1$)

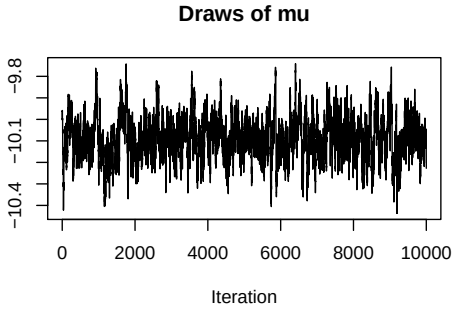


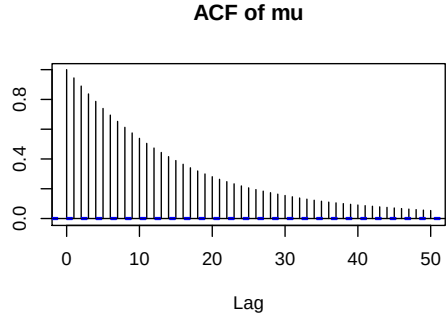


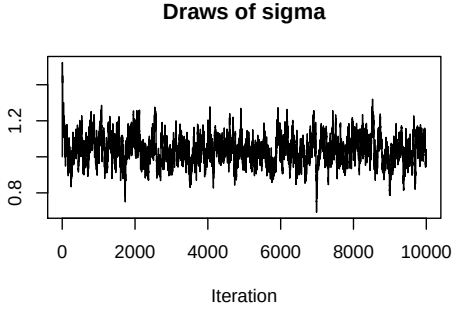


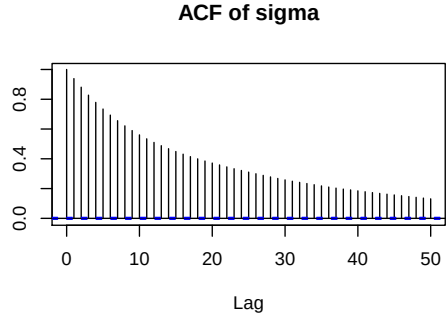


NC – High Volatility of Volatility ($\phi = 0.5, \sigma = 1$)

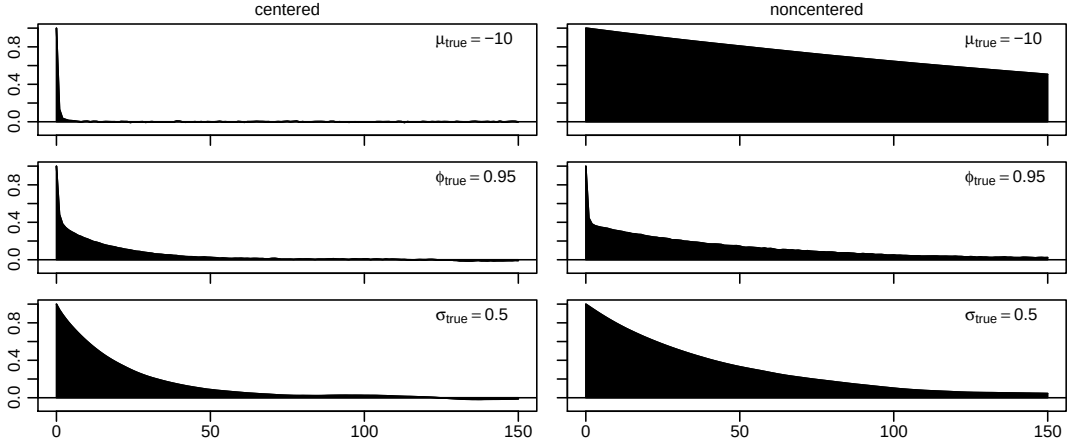




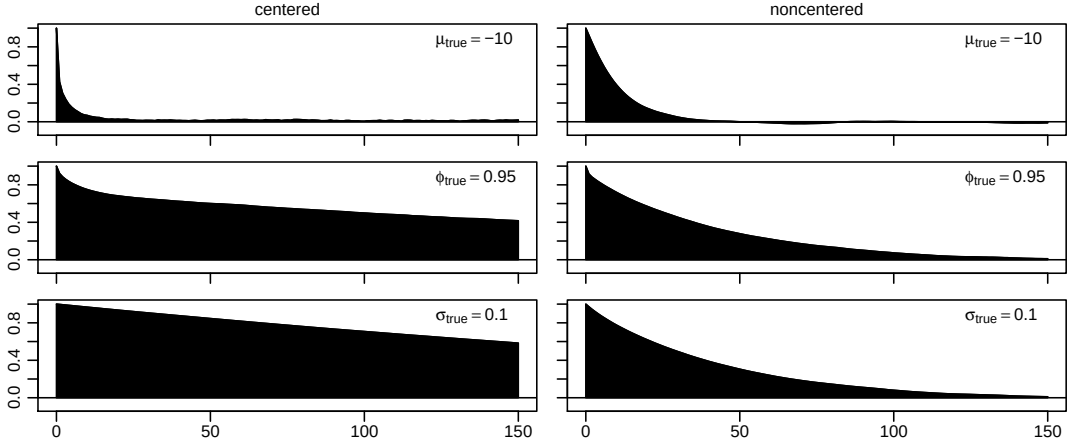




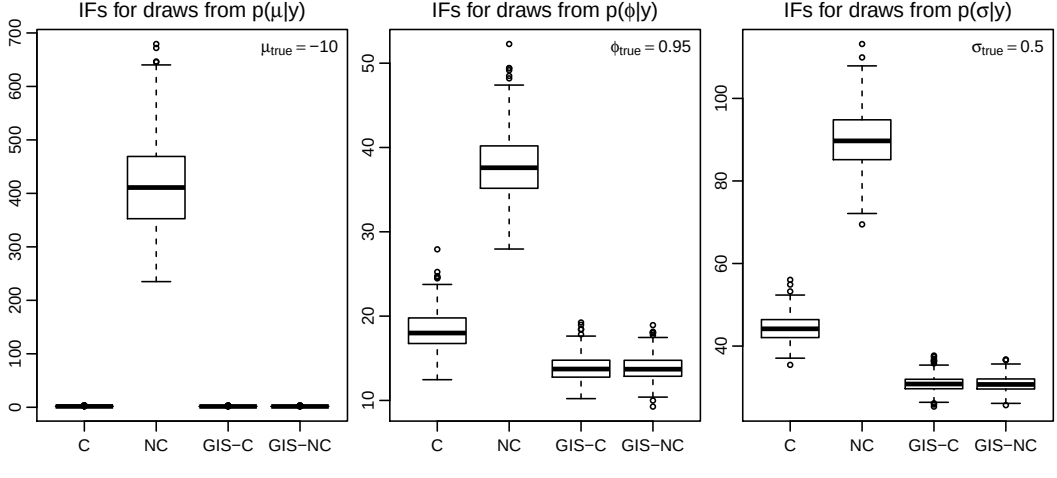
High Volatility of Volatility ($\phi = 0.95, \sigma = 0.5$)



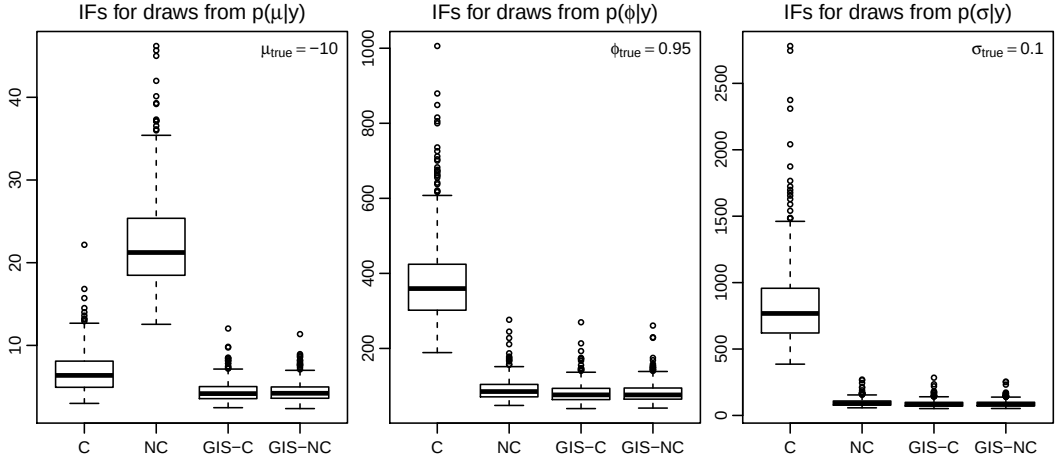
Low Volatility of Volatility ($\phi = 0.95, \sigma = 0.1$)



What about sample variations? (400 repetitions, $\sigma = 0.5$)



What about sample variations? (400 repetitions, $\sigma = 0.1$)



To Center or Not to Center?

$p(\mu\|y)$	σ_{true} \ ϕ_{true}	0	0.5	0.8	0.9	0.95	0.96	0.97	0.98	0.99
C	0.1	626	236	65	19	6	4	3	2	3
	0.2	255	94	22	9	3	2	2	2	3
$T_{\text{CPU}} = 2.33$	0.3	113	47	15	5	2	2	2	2	3
	0.4	72	34	11	3	2	2	2	2	5
	0.5	51	26	8	2	2	2	2	2	8
NC	0.1	9	10	12	12	21	31	51	111	465
	0.2	23	21	15	21	70	107	185	422	1892
$T_{\text{CPU}} = 2.37$	0.3	23	17	17	41	153	231	404	935	3990
	0.4	18	16	23	70	265	413	730	1637	6440
	0.5	17	17	32	107	421	641	1127	2593	10101

Table: Inefficiency factors for 100 000 draws from $p(\mu|y)$ in the various parameterizations. Time series length $T = 5\,000$, the values reported are medians of 400 repetitions, and T_{CPU} denotes the median time to complete 1 000 iterations.

To Center or Not to Center?

$p(\sigma\|y)$	σ_{true} \ ϕ_{true}	0	0.5	0.8	0.9	0.95	0.96	0.97	0.98	0.99
C	0.1	5595	4747	3625	1750	794	581	429	297	197
	0.2	3154	1979	740	374	182	151	124	100	79
$T_{CPU} = 2.33$	0.3	684	474	311	157	89	79	69	60	50
	0.4	248	262	165	91	59	53	48	43	38
	0.5	131	165	106	63	44	41	37	34	31
NC	0.1	59	65	87	121	94	81	72	70	86
	0.2	100	96	120	87	70	70	74	89	141
$T_{CPU} = 2.37$	0.3	62	76	92	70	71	78	89	115	195
	0.4	38	75	75	66	80	89	106	143	257
	0.5	31	68	67	67	91	103	124	172	326

Table: Inefficiency factors for 100 000 draws from $p(\sigma|y)$ in the various parameterizations. Time series length $T = 5\,000$, the values reported are medians of 400 repetitions, and T_{CPU} denotes the median time to complete 1 000 iterations.

Interweaving the Parameterizations

Parameterization:

- has major influence on MCMC sampling efficiency
- which highly depends on values of ϕ and σ.

No big problem for univariate SV models (parameter range typically quite moderate), but crucial for factor models (parameter range typically wide).

Idea: Combine "best of both worlds" by interweaving the parameterizations.

Interweaving the Parameterizations

Interweaving Centered (C) and Fully-Noncentered (NC):

- Yu and Meng (2011) promise a surprisingly simple procedure combining "best of both worlds" – Ancillarity-Sufficiency Interweaving Strategy (ASIS).
- Based on sampling the parameters in question – in our case μ, σ (and ϕ) – twice: once utilizing C and again utilizing NC.
- Not clear whether one should start C and redraw NC ("baseline C", the approach suggested by Yu and Meng, 2011) or vice versa ("baseline NC").

Global interweaving strategy: baseline C (GIS-C)

We *interweave* C and NC by repeating the following steps:

1 Draw h (C)

2 Draw μ, ϕ, σ (C)

2* Calculate $\tilde{\mathrm{h}} = \frac{\mathrm{h}-\mu}{\sigma}$

2** Redraw μ, σ, ϕ (NC)[1]

2*** Calculate $\mathrm{h} = \mu + \sigma\tilde{\mathrm{h}}$ ("bookkeeping")

3 Draw indicators r

[1] Actually, as ϕ is not involved in the reparameterization, one might as well redraw μ and σ only – the difference concerning sampling efficiency is minor.

Global interweaving strategy: baseline NC (GIS-NC)

We *interweave* NC and C by repeating the following steps:

1. Draw $\tilde{\mathsf{h}}$ (NC)
2. Draw μ, ϕ, σ (NC)
2* Calculate $\mathsf{h} = \mu + \sigma\tilde{\mathsf{h}}$
2** Redraw μ, σ, ϕ (C)[2]
2*** Calculate $\tilde{\mathsf{h}} = \frac{\mathsf{h}-\mu}{\sigma}$ ("bookkeeping")
3. Draw indicators r

[2]Here, μ and ϕ are sampled jointly, thus we redraw all three parameters.

Inefficiency Comparison for σ

$p(\sigma\|y)$	σ_{true} \ ϕ_{true}	0	0.5	0.8	0.9	0.95	0.96	0.97	0.98	0.99
C $T_{CPU} = 2.33$	0.1	5595	4747	3625	1750	794	581	429	297	197
	0.2	3154	1979	740	374	182	151	124	100	79
	0.3	684	474	311	157	89	79	69	60	50
	0.4	248	262	165	91	59	53	48	43	38
	0.5	131	165	106	63	44	41	37	34	31
NC $T_{CPU} = 2.37$	0.1	59	65	87	121	94	81	72	70	86
	0.2	100	96	120	87	70	70	74	89	141
	0.3	62	76	92	70	71	78	89	115	195
	0.4	38	75	75	66	80	89	106	143	257
	0.5	31	68	67	67	91	103	124	172	326
GIS-C $T_{CPU} = 2.38$	0.1	58	64	87	118	86	73	63	58	60
	0.2	98	93	113	75	54	50	48	48	51
	0.3	60	72	80	53	42	41	40	40	40
	0.4	35	68	59	42	36	34	34	33	33
	0.5	28	60	48	36	31	30	29	29	28
GIS-NC $T_{CPU} = 2.43$	0.1	58	65	87	118	86	74	63	58	60
	0.2	97	94	111	75	53	50	48	48	51
	0.3	59	72	80	53	42	41	40	40	40
	0.4	35	69	59	42	36	34	34	33	33
	0.5	28	60	47	36	31	30	29	29	28

Percentage Efficiency Gain in ESS

$p(\sigma\|y)$	σ_{true} \ ϕ_{true}	0	0.5	0.8	0.9	0.95	0.96	0.97	0.98	0.99
GIS-C vs. better	0.1	1	0	0	3	10	11	15	21	42
	0.2	3	2	8	16	32	40	53	86	56
	0.3	4	6	16	33	71	90	73	50	26
	0.4	9	9	27	57	66	54	41	30	15
	0.5	12	15	41	77	43	35	27	19	10
GIS-C vs. worse	0.1	9540	7236	4070	1388	826	690	581	411	226
	0.2	3149	2012	565	401	243	202	156	110	180
	0.3	1049	564	290	197	113	93	123	190	391
	0.4	612	280	178	116	124	158	215	331	679
	0.5	372	178	124	89	193	244	324	503	1058

Inefficiency Comparison for μ

$p(\mu\|y)$	σ_{true} \ ϕ_{true}	0	0.5	0.8	0.9	0.95	0.96	0.97	0.98	0.99
C	0.1	626	236	65	19	6	4	3	2	3
	0.2	255	94	22	9	3	2	2	2	3
$T_{CPU} = 2.33$	0.3	113	47	15	5	2	2	2	2	3
	0.4	72	34	11	3	2	2	2	2	5
	0.5	51	26	8	2	2	2	2	2	8
NC	0.1	9	10	12	12	21	31	51	111	465
	0.2	23	21	15	21	70	107	185	422	1892
$T_{CPU} = 2.37$	0.3	23	17	17	41	153	231	404	935	3990
	0.4	18	16	23	70	265	413	730	1637	6440
	0.5	17	17	32	107	421	641	1127	2593	10101
GIS-C	0.1	9	10	11	8	4	3	3	2	3
	0.2	23	20	11	5	2	2	2	2	3
$T_{CPU} = 2.38$	0.3	22	15	9	4	2	2	2	2	3
	0.4	17	13	7	3	2	2	2	2	4
	0.5	15	12	5	2	2	2	2	2	6
GIS-NC	0.1	9	9	11	8	4	3	3	2	3
	0.2	23	20	11	5	2	2	2	2	3
$T_{CPU} = 2.43$	0.3	22	15	9	3	2	2	2	2	3
	0.4	17	13	7	3	2	2	2	2	4
	0.5	15	12	5	2	2	2	2	2	6

Percentage Efficiency Gain in ESS

$p(\mu\|y)$	σ_{true} \ ϕ_{true}	0	0.5	0.8	0.9	0.95	0.96	0.97	0.98	0.99
GIS-C vs. better	0.1	2	3	9	48	47	35	22	13	11
	0.2	3	5	34	62	21	13	9	8	14
	0.3	5	13	79	40	11	8	7	6	22
	0.4	9	23	60	21	7	5	5	7	27
	0.5	12	41	49	15	5	4	7	7	24
GIS-C vs. worse	0.1	6982	2393	479	121	390	848	1880	5132	15822
	0.2	1021	370	106	293	2905	5295	10543	24983	71993
	0.3	412	209	100	1066	8477	13780	25594	55723	139938
	0.4	337	154	253	2555	16344	26907	45668	88141	157190
	0.5	241	114	512	4809	27319	40835	69825	123682	158116

Three Main Insights

- AWOL is easy to implement and fast: no need to calculate covariance matrix, no need for forward filtering / backward smoothing.
- Efficiency of MCMC estimation of univariate SV-models highly depends on the parameter values:
  - If σ^2 or ϕ is small, NC is superior.
  - For highly persistent volatilities, C works better.

  In line with results from Strickland et al. (2008).
- Possible to overcome these issues by interweaving strategies.

For more details, see Kastner and Frühwirth-Schnatter, 2014. Want to try yourself? R-package `stochvol` (Kastner, 2016).

Univariate SV
Review and Motivation
Methods for MCMC Sampling
Efficiency Issues
To Center or Not to Center?
Interweaving the Parameterizations – ASIS
Efficiency Comparison
The R Package stochvol
Motivation
Running the Sampler
Assessing the Output
Model extensions
Heavy Tails
Leverage/Asymmetry
A Bayesian Predictive Framework
Predictive Density, Predictive Likelihoods and Predictive Bayes Factors
Exchange Rate Example

A quote from Bos (2012)

While there are literally thousands of applications of GARCH, for SV, this number is far lower. Two reasons for this relative lack of applied work using SV are apparent. First, there are as of yet no standard packages for estimating SV models, whereas for GARCH, most statistical packages have a wealth of options for incorporating GARCH effects. [...]

Running the sampler (Kastner, 2017)

```
R> library(stochvol)
R> data(exrates)
R> ret <- logret(exrates$USD, demean = TRUE)
R> res <- svsample(ret, priormu    = c(-10, 1),
+                       priorphi   = c(20, 1.1),
+                       priorsigma = .1)
```

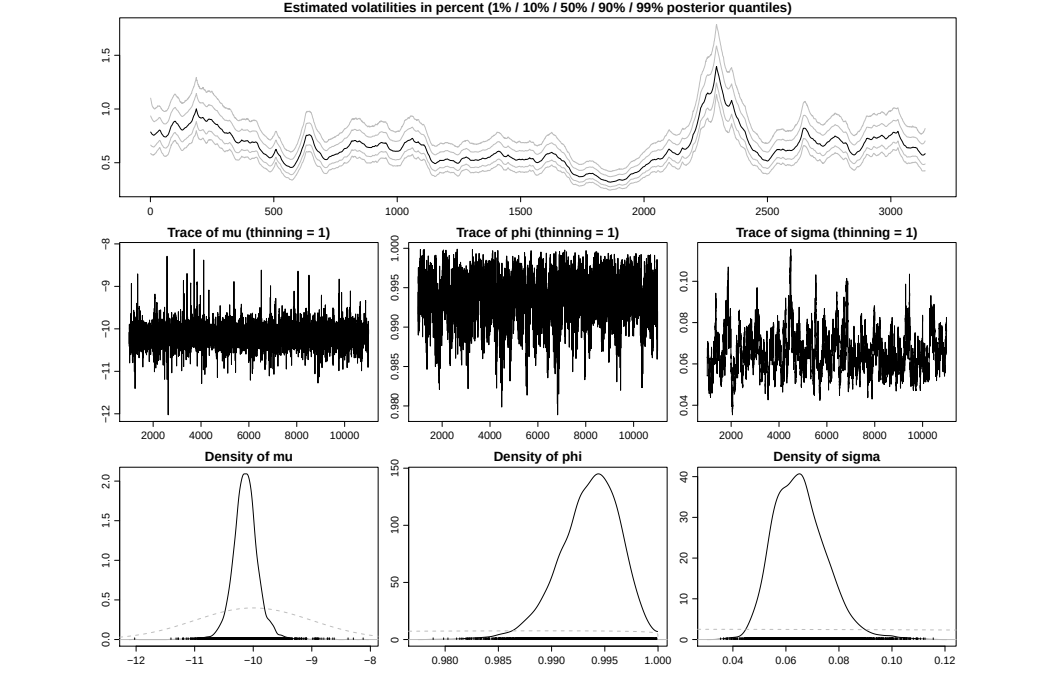
```
Calling GIS_C sampler with 11000 iter. Series length T = 3139.

  0% [++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++] 100%

Timing (elapsed): 12.92 seconds.
613 iterations per second.

Converting results to coda objects... Done!
Summarizing posterior draws... Done!
```

Printing and plotting the resulting object

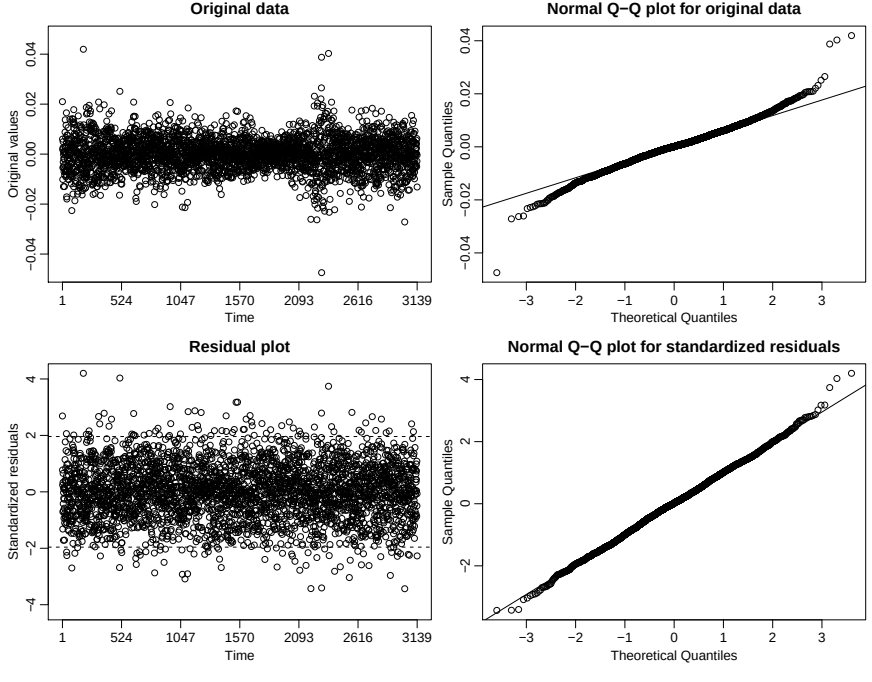
```
R> summary(res, showlatent = FALSE)
```

```
Summary of 10000 MCMC draws after a burn-in of 1000.

Prior distributions:
mu         ~ Normal(mean = -10, sd = 1)
(phi+1)/2  ~ Beta(a0 = 20, b0 = 1.1)
sigma^2    ~ 0.1 * Chisq(df = 1)

Posterior draws of parameters (thinning = 1):
              mean      sd       5%      50%      95%  ESS
mu         -10.1308 0.22806 -10.4661 -10.1345 -9.7746 4537
phi          0.9936 0.00282   0.9886   0.9938  0.9977  386
sigma        0.0654 0.01005   0.0510   0.0646  0.0827  142
exp(mu/2)    0.0064 0.00076   0.0053   0.0063  0.0075 4537
sigma^2      0.0044 0.00139   0.0026   0.0042  0.0068  142
```

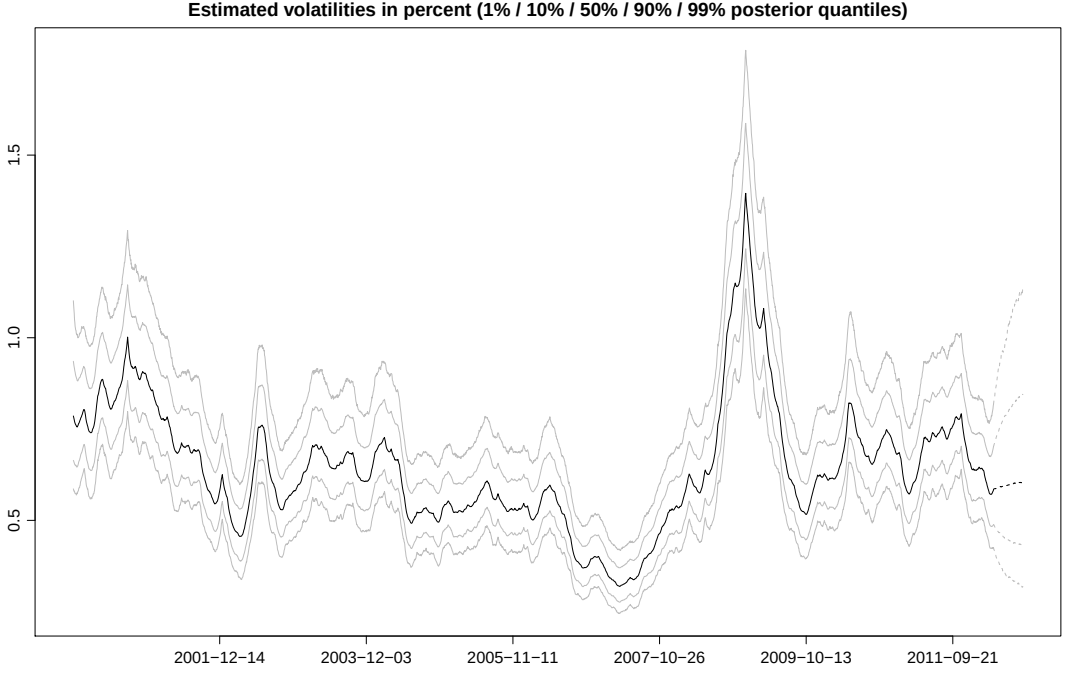
```
R> plot(res)
```


Estimated volatilities in percent (1% / 10% / 50% / 90% / 99% posterior quantiles)
Trace of mu (thinning = 1)
Trace of phi (thinning = 1)
Trace of sigma (thinning = 1)
Density of mu
Density of phi
Density of sigma

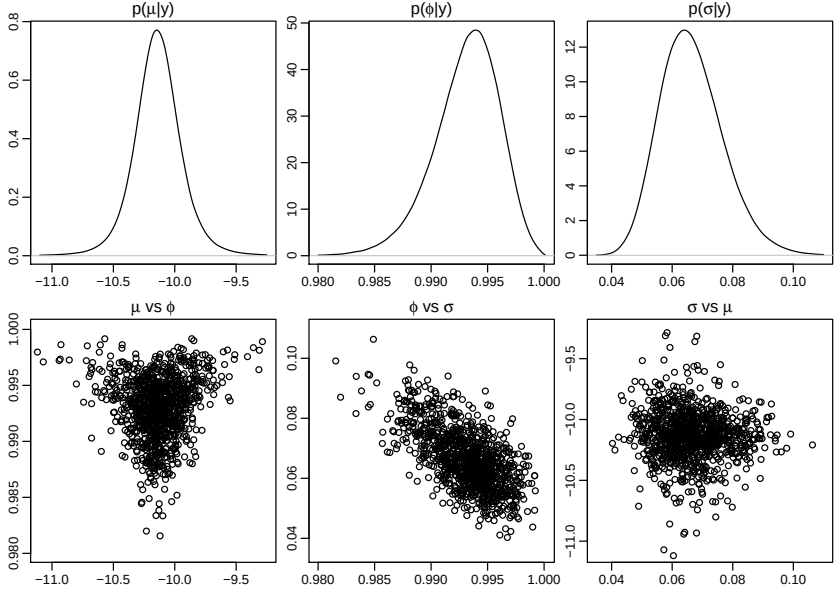
```
R> plot(resid(res), ret)
```



```
R> volplot(res, forecast = 100, dates = exrates$date[-1])
```



A view on the bivariate posterior marginals



Bayesian regression: Concept

Thanks to the "toolbox nature" of Gibbs sampling, **stochvol** can be used as a plug-in tool for other MCMC samplers. Consider a Bayesian normal linear model with T observations and $k = p - 1$ predictors,

$$\mathsf{y}|\boldsymbol{\beta}, \boldsymbol{\Sigma} \sim N(\mathsf{X}\boldsymbol{\beta}, \boldsymbol{\Sigma}) \,.$$

Two specifications of errors:

- Homoskedastic: $\boldsymbol{\Sigma} \equiv \sigma_\epsilon^2 \mathsf{I}$
- Heteroskedastic: $\boldsymbol{\Sigma} \equiv \mathsf{Diag}\left(\mathrm{e}^{h_1}, \ldots, \mathrm{e}^{h_T}\right)$

Bayesian regression: Concept

Thanks to the "toolbox nature" of Gibbs sampling, **stochvol** can be used as a plug-in tool for other MCMC samplers. Consider a Bayesian normal linear model with T observations and $k = p - 1$ predictors,

$$\mathrm{y}|\boldsymbol{\beta}, \boldsymbol{\Sigma} \sim N(\mathrm{X}\boldsymbol{\beta}, \boldsymbol{\Sigma}).$$

Two specifications of errors:

- Homoskedastic: $\boldsymbol{\Sigma} \equiv \sigma_\epsilon^2 \mathrm{I}$
- Heteroskedastic: $\boldsymbol{\Sigma} \equiv \mathrm{Diag}\left(\mathrm{e}^{h_1}, \ldots, \mathrm{e}^{h_T}\right)$

Gibbs sampler for drawing from the posterior:

- Homoskedastic: Using conjugate priors, a simple Gibbs sampler is given through drawing in turn from the full conditional distributions

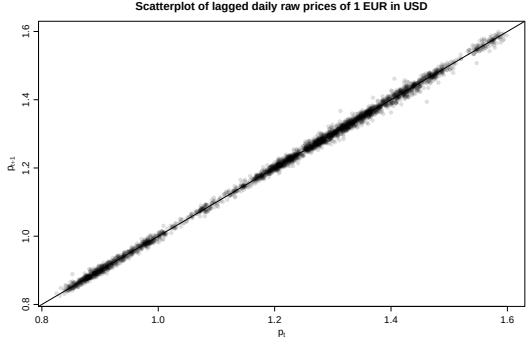
$$\begin{aligned} \boldsymbol{\beta}|\mathrm{y}, \boldsymbol{\Sigma} &\sim N(\mathrm{b}_T, \mathrm{B}_T) && (1) \\ \sigma_\epsilon^2|\mathrm{y}, \boldsymbol{\beta} &\sim \mathcal{G}^{-1}(c_T, C_T) && (2) \end{aligned}$$

- Heteroskedastic: Simply replace (2) by a call to `svsample`!

Bayesian regression: Example

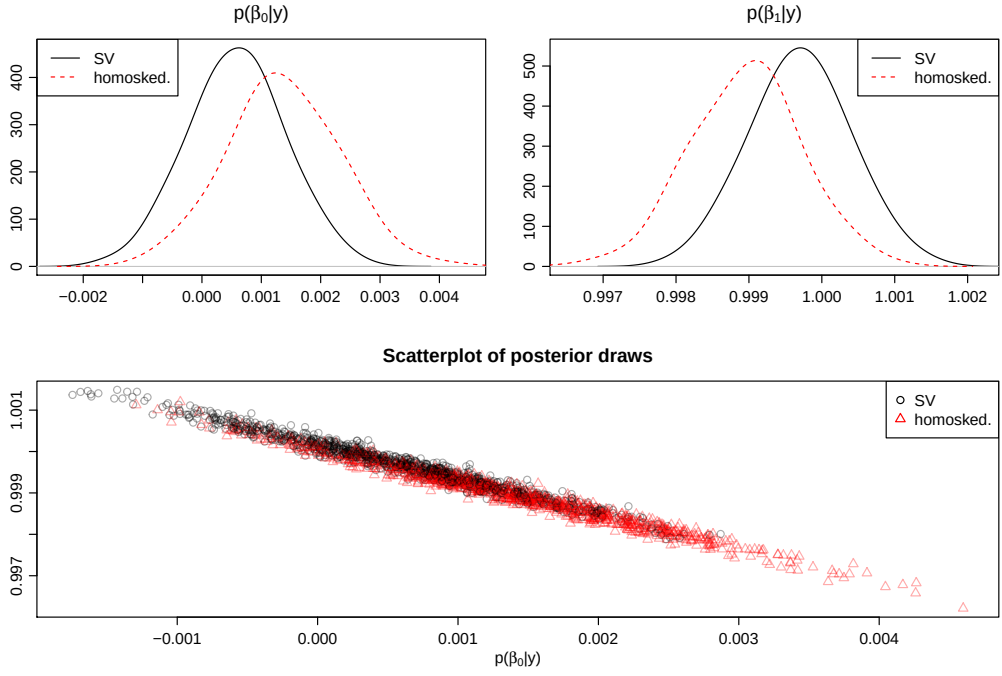
Daily price of 1 EUR in USD from January 3, 2000 until April 4, 2012, denoted by $\mathrm{p} = (p_1, p_2, \ldots, p_T)'$. Let

$$\mathrm{y} = \begin{pmatrix} p_2 \\ p_3 \\ \vdots \\ p_T \end{pmatrix}, \quad \mathrm{X} = \begin{pmatrix} 1 & p_1 \\ 1 & p_2 \\ \vdots & \vdots \\ 1 & p_{T-1} \end{pmatrix}.$$

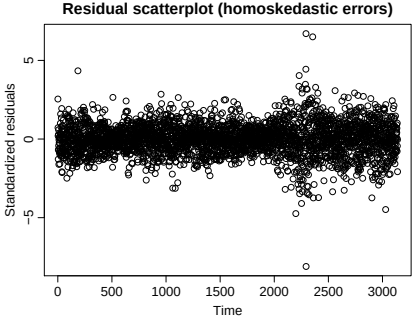


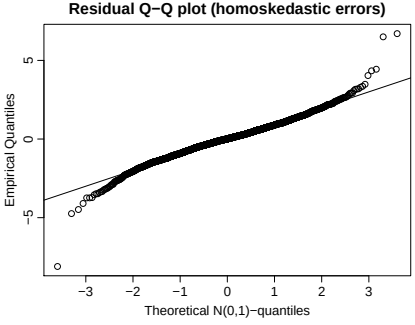
Clearly, we expect the posterior distribution of $\boldsymbol{\beta}$ to spread around $(0, 1)'$ which corresponds to a random walk.

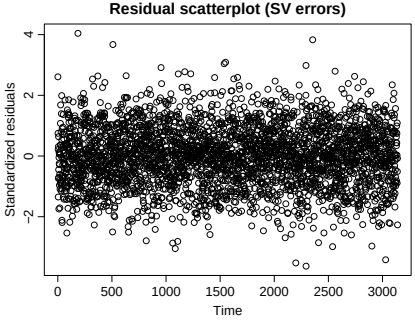
Bayesian regression: Example

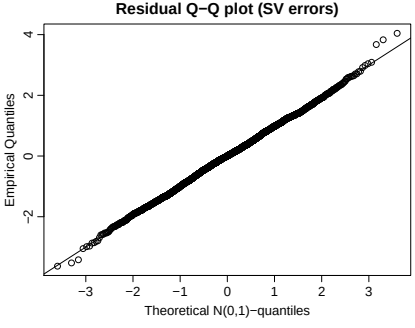


Bayesian regression: Example









Univariate SV
Review and Motivation
Methods for MCMC Sampling

Efficiency Issues
To Center or Not to Center?
Interweaving the Parameterizations – ASIS
Efficiency Comparison

The R Package stochvol
Motivation
Running the Sampler
Assessing the Output

Model extensions
Heavy Tails
Leverage/Asymmetry

A Bayesian Predictive Framework
Predictive Density, Predictive Likelihoods and Predictive Bayes Factors
Exchange Rate Example

What about heavy tails?

Several authors have suggested to use non-normal conditional innovation distributions for stochastic volatility (SV) modeling. Examples:

- Student's t-distribution (Harvey et al., 1994)
- Extended Generalized Inverse Gaussian (Silva et al., 2006)
- (Semi-)parametric innovations (Jensen and Maheu, 2010; Delatola and Griffin, 2011)
- GH skew Student's t-distribution (Nakajima and Omori, 2012)

What about heavy tails?

Let's look at t-distributions:

$$y_t | h_t, \nu \sim t_\nu(0, \exp h_t)$$

with a uniform prior on $(\nu_{\min}, \nu_{\max})$ for ν. Conditionally on h_t, the data is assumed to follow a zero-mean non-standardized Student's t-distribution with ν degrees of freedom and variance $(\nu \exp h_t)/(\nu - 2)$ for $\nu > 2$.

The data augmentation trick:

$$\begin{aligned}
y_t | h_t, \tau_t &\sim N(0, \tau_t \exp h_t) \\
\tau_t | \nu &\sim \mathcal{G}^{-1}(\nu/2, \nu/2) \\
h_t | h_{t-1}, \mu, \phi, \sigma &\sim N\big(\mu + \phi(h_{t-1} - \mu), \sigma_\eta^2\big) \\
h_0 | \mu, \phi, \sigma &\sim N\big(\mu, \sigma_\eta^2/(1 - \phi^2)\big)
\end{aligned}$$

Modifications to the MCMC sampler

- Treating $\boldsymbol{\tau} = (\tau_1, \ldots, \tau_n)^\top$ as latent data and letting $\tilde{y}_t = y_t/\sqrt{\tau_t}$ for $t \in \{1, \ldots, n\}$, we have

$$\tilde{y}_t | h_t \sim N(0, \exp h_t)\,,$$

  and the AWOL sampler can be directly applied to the transformed data.

- It is straightforward to see that the marginal posterior is given through

$$\tau_t | y_t, h_t, \nu \sim \mathcal{G}^{-1}\left(\frac{\nu+1}{2}, \frac{\nu + y_t^2 \exp(-h_t)}{2}\right),$$

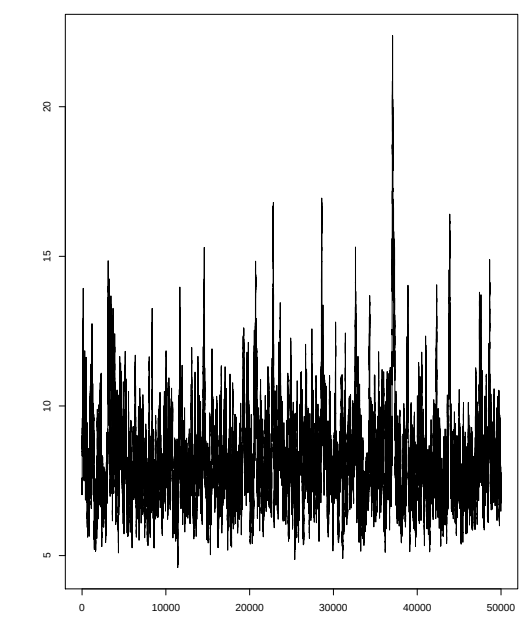
  independently for each $t \in \{1, \ldots, n\}$.

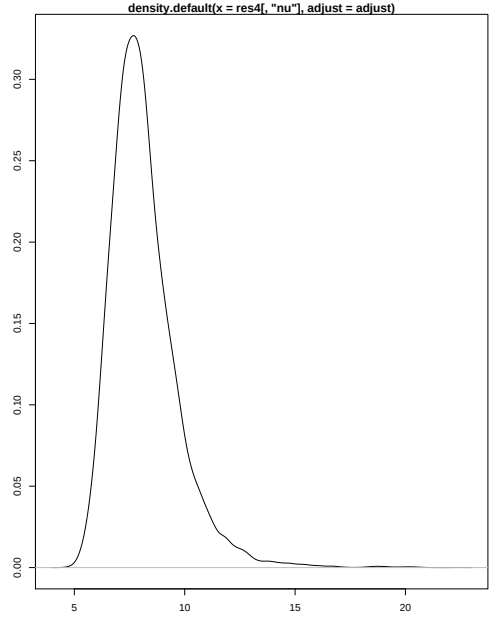
- The full conditional posterior of the degrees of freedom parameter, $\nu|\cdot$, only depends on $\boldsymbol{\tau}$. Its density is given through the product of n univariate truncated inverse gamma densities which may be written as

$$p(\nu|\cdot) = p(\nu|\boldsymbol{\tau}) \propto \left(\frac{\nu}{2}\right)^{n\nu/2} \Gamma\left(\frac{\nu}{2}\right)^{-n} \left(\prod_{t=1}^{n} \tau_t\right)^{-\nu/2} \exp\left\{-\frac{\nu}{2}\sum_{t=1}^{n}\frac{1}{\tau_t}\right\}$$

  for $\nu \in (a, b)$ and zero elsewhere $\Rightarrow$ MH-update with independence proposal.

AR(1)-SVt model for CZK-EUR exchange rates: Draws from the marginal posterior of ν and smoothed kernel density estimate



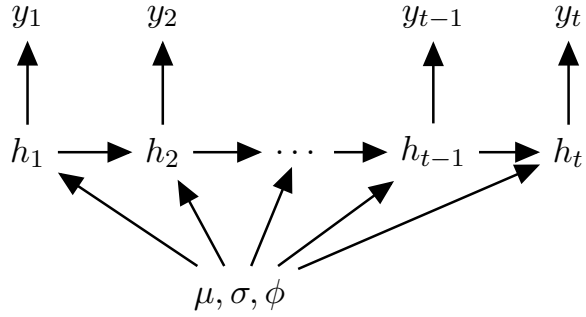


What about asymmetry?

Stochastic volatility model without leverage:

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

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

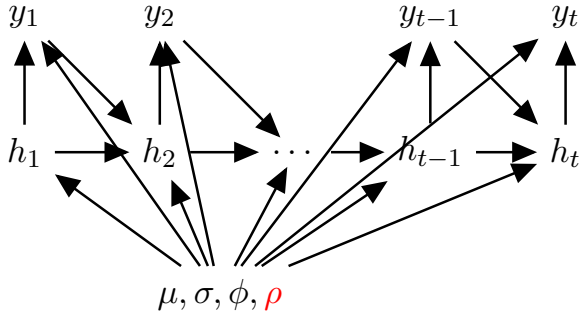


What about asymmetry?

Stochastic volatility model with leverage:

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

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



Estimation: SV vs. SV with leverage (Omori et al., 2007)

Similarities

1. Approximate linear Gaussian state space, 10 components
2. Efficiency problems under extreme "true" parameters

Differences

1. ρ is in the model
   $\implies$ ε_t is not independent of the model parameters
   $\implies$ **bivariate mixture approximation** is needed to get rid of the $\log(\chi^2_1)$ distribution
2. When `stochvol` samples the parameters from $\mu, \sigma, \phi \mid \boldsymbol{y}^*, \boldsymbol{h}, \boldsymbol{s}$ via Bayesian regression, it exploits some conditional independencies that don't hold here

Estimation: SV vs. SV with leverage (Omori et al., 2007)

Centered parameterization. $M_{s_t}, V_{s_t}, m_{s_t}, v_{s_t}$ are constants of the mixture.

$$y_t^* = m_{s_t} + h_t + v_{s_t} Z_t^1$$

$$h_{t+1} = \mu + \sigma\rho\text{sgn}(y_t) M_{s_t} + \phi(h_t - \mu) + \sigma\left(\rho\text{sgn}(y_t) V_{s_t} Z_t^1 + \sqrt{1-\rho^2} Z_t^2\right)$$

$$Z_t^i \sim \text{i.i.d. } \mathcal{N}(0,1)$$

Sampler. Repeat:

1. Draw $\boldsymbol{s}$ from $\boldsymbol{s} \mid \boldsymbol{y^*}, \boldsymbol{h}, \mu, \sigma, \phi, \rho, \text{sgn}(\boldsymbol{y})$ via inverse transform sampling
2. Draw σ, ϕ, ρ from $\sigma, \phi, \rho \mid \boldsymbol{y^*}, \boldsymbol{s}, ~~\boldsymbol{h}~~, \text{sgn}(\boldsymbol{y})$ via indep. MH
3. Draw $\boldsymbol{h}, \mu$ from $\boldsymbol{h}, \mu \mid \boldsymbol{y^*}, \boldsymbol{s}, \sigma, \phi, \rho, \text{sgn}(\boldsymbol{y})$ via the simulation smoother

Now available in stochvol 3+ (Hosszejni and Kastner, 2021)

Univariate SV
Review and Motivation
Methods for MCMC Sampling
Efficiency Issues
To Center or Not to Center?
Interweaving the Parameterizations – ASIS
Efficiency Comparison
The R Package stochvol
Motivation
Running the Sampler
Assessing the Output
Model extensions
Heavy Tails
Leverage/Asymmetry
A Bayesian Predictive Framework
Predictive Density, Predictive Likelihoods and Predictive Bayes Factors
Exchange Rate Example

A Bayesian predictive framework

Within a Bayesian framework, a natural way of assessing the predictive performance of a given model is through its *predictive density* (sometimes also referred to as *posterior predictive distribution*). It is given through

$$p(y_{t+1}|\mathbf{y}^o_{[1:t]}) = \int_{\boldsymbol{\kappa}} p(y_{t+1}|\mathbf{y}^o_{[1:t]}, \boldsymbol{\kappa}) \times p(\boldsymbol{\kappa}|\mathbf{y}^o_{[1:t]})\, \mathrm{d}\boldsymbol{\kappa},$$

where $\boldsymbol{\kappa}$ denotes the vector of all unobservables, i.e., parameters and possible latent variables.

A Bayesian predictive framework

In the SV errors case, Equation 3 is a $(n + p + 3)$-dimensional integral which cannot be solved analytically. Nevertheless, it may be evaluated at an arbitrary point x through Monte Carlo integration,

$$p(x|\mathrm{y}^o_{[1:t]}) \approx \frac{1}{M} \sum_{m=1}^{M} p(x|\mathrm{y}^o_{[1:t]}, \boldsymbol{\kappa}^{(m)}_{[1:t]}), \tag{3}$$

where $\boldsymbol{\kappa}^{(m)}_{[1:t]}$ stands for the mth draw from the respective posterior distribution up to time t. If Equation 3 is evaluated at $x = y^o_{t+1}$, we refer to it as the (one-step-ahead) *predictive likelihood* at time $t + 1$, denoted PL_{t+1}.

Moreover, draws from the posterior predictive distribution can be obtained by simulating values $y^{(m)}_{t+1}$ from the distribution given through the density $p(y_{t+1}|\mathrm{y}^o_{[1:t]}, \boldsymbol{\kappa}^{(m)}_{[1:t]})$, the summands of Equation 3.

A Bayesian predictive framework

It is worth pointing out that log predictive likelihoods also carry an intrinsic connection to the log *marginal likelihood*, defined through

$$\log ML = \log p(\mathrm{y}^o) = \log \int_{\boldsymbol{\kappa}} p(\mathrm{y}^o|\boldsymbol{\kappa}) \times p(\boldsymbol{\kappa}) \, \mathrm{d}\boldsymbol{\kappa}.$$

This real number corresponds to the logarithm of the normalizing constant which appears in the denominator of Bayes' law and is often used for evaluating model evidence. It can straightforwardly be decomposed into the sum of the logarithms of the one-step-ahead predictive likelihoods:

$$\log ML = \log p(\mathrm{y}^o) = \log \prod_{t=1}^{n} p(y_t^o|\mathrm{y}_{[1:t-1]}^o) = \sum_{t=1}^{n} \log PL_t.$$

A Bayesian predictive framework

Cumulative sums of log PL_t also allow for model comparison through cumulative log predictive Bayes factors. Letting $PL_t(A)$ denote the predictive likelihood of model A at time t, and $PL_t(B)$ the corresponding value of model B, the cumulative log predictive Bayes factor at time u (and starting point s) in favor of model A over model B is straightforwardly given through

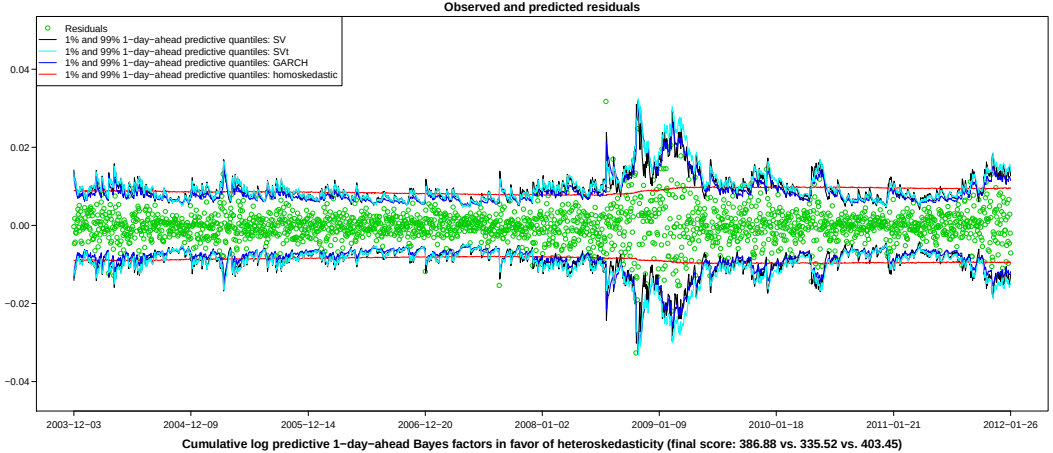
$$\log\left[\frac{p_A(\mathrm{y}^o_{[s+1:u]}|\mathrm{y}^o_{[1:s]})}{p_B(\mathrm{y}^o_{[s+1:u]}|\mathrm{y}^o_{[1:s]})}\right] = \sum_{t=s+1}^{u} \log\left[\frac{PL_t(A)}{PL_t(B)}\right] = \sum_{t=s+1}^{u} [\log PL_t(A) - \log PL_t(B)].$$

When the cumulative log predictive Bayes factor is positive at a given point in time, there is evidence in favor of model A, and vice versa. In this context, information up to time s is regarded as prior information, while out-of-sample predictive evaluation starts at time $s+1$. Note that the usual (overall) log Bayes factor is a special case of this equation for $s = 0$ and $u = n$. More details in (Geweke and Amisano, 2010; Kastner, 2016).

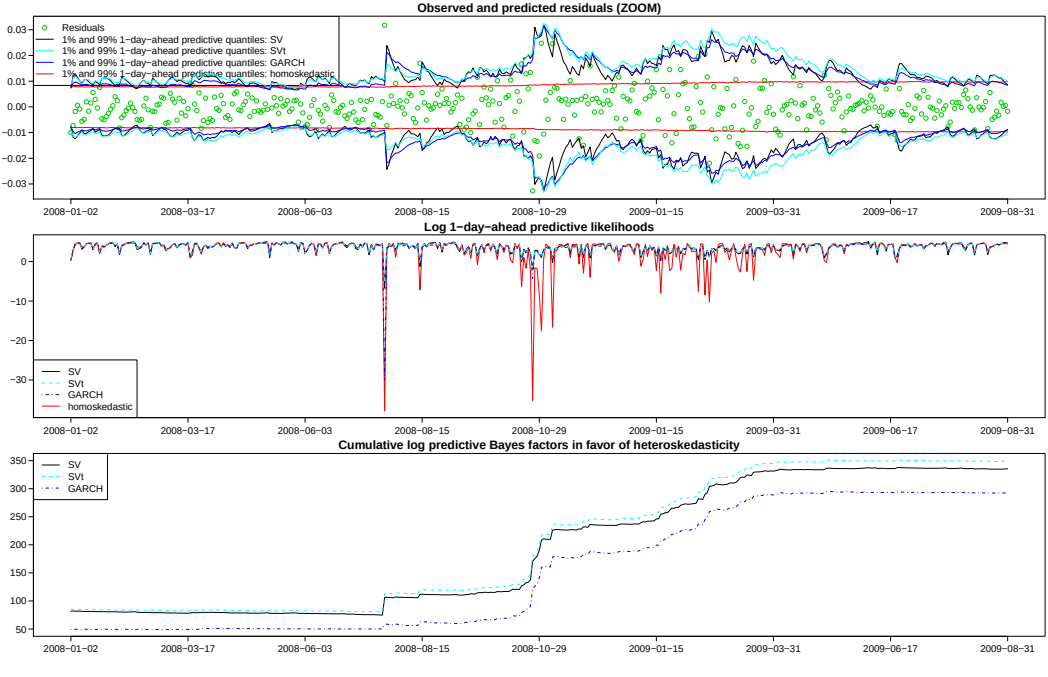
Predictive Comparison – CZK



Predictive Comparison – CZK



Predictive Comparison – CZK



Bos, Charles S. (2012). "Relating stochastic volatility estimation methods". In: *Handbook of Volatility Models and Their Applications*. Ed. by Luc Bauwens et al. Wiley, pp. 147–174. DOI: 10.1002/9781118272039.ch6.

Delatola, Eleni-Ioanna and Jim E. Griffin (2011). "Bayesian Nonparametric Modelling of the Return Distribution with Stochastic Volatility". *Bayesian Analysis* 6, pp. 901–926. DOI: 10.1214/11-BA632.

Frühwirth-Schnatter, Sylvia (2004). "Efficient Bayesian parameter estimation". In: *State Space and Unobserved Component Models: Theory and Applications*. Ed. by Andrew Harvey et al. Cambridge: Cambridge University Press, pp. 123–151.

Frühwirth-Schnatter, Sylvia and Helga Wagner (2010). "Stochastic Model Specification Search for Gaussian and Partial Non-Gaussian State Space Models". *Journal of Econometrics* 154, pp. 85–100. DOI: 10.1016/j.jeconom.2009.07.003.

Geweke, John and Gianni Amisano (2010). "Comparing and Evaluating Bayesian Predictive Distributions of Asset Returns". *International Journal of Forecasting* 26, pp. 216–230. DOI: 10.1016/j.ijforecast.2009.10.007.

Harvey, A. C. et al. (1994). "Multivariate Stochastic Variance Models". *Review of Economic Studies* 61, pp. 247–264. DOI: 10.2307/2297980.

Hosszejni, Darjus and Gregor Kastner (2021). "Modeling Univariate and Multivariate Stochastic Volatility in R with stochvol and factorstochvol". *Journal of Statistical Software* 100, 1–34. DOI: 10.18637/jss.v100.i12.

Jacquier, Eric et al. (1994). "Bayesian analysis of stochastic volatility models". *Journal of Business & Economic Statistics* 12, pp. 371–389. DOI: 10.1080/07350015.1994.10524553.

Jensen, Mark J and John M Maheu (2010). "Bayesian semiparametric stochastic volatility modeling". *Journal of Econometrics* 157, pp. 306–316. DOI: doi:10.1016/j.jeconom.2010.01.014.

Kastner, Gregor (2016). "Dealing with Stochastic Volatility in Time Series Using the R Package stochvol". *Journal of Statistical Software* 69, pp. 1–30. DOI: 10.18637/jss.v069.i05.

– (2017). *stochvol: Efficient Bayesian inference for stochastic volatility (SV) models*. R package version 1.3.3. URL: https://CRAN.R-project.org/package=stochvol.

Kastner, Gregor and Sylvia Frühwirth-Schnatter (2014). "Ancillarity-Sufficiency Interweaving Strategy (ASIS) for Boosting MCMC Estimation of Stochastic Volatility Models". *Computational Statistics and Data Analysis* 76, pp. 408–423. DOI: 10.1016/j.csda.2013.01.002.

Kim, S. et al. (1998). "Stochastic Volatility: Likelihood Inference and Comparison with ARCH Models". *Review of Economic Studies* 65, pp. 361–393. DOI: 10.1111/1467-937X.00050.

McCausland, W. J. et al. (2011). "Simulation smoothing for state-space models: A computational efficiency analysis". *Computational Statistics and Data Analysis* 55, pp. 199–212. DOI: 10.1016/j.csda.2010.07.009.

Nakajima, Jouchi and Yasuhiro Omori (2012). "Stochastic volatility model with leverage and asymmetrically heavy-tailed error using GH skew Student's t-distribution". *Computational Statistics and Data Analysis* 56, pp. 3690–3704. ISSN: 0167-9473. DOI: 10.1016/j.csda.2010.07.012.

Omori, Yashuhiro et al. (2007). "Stochastic volatility with leverage: Fast and efficient likelihood inference". *Journal of Econometrics* 140, pp. 425–449. DOI: 10.1016/j.jeconom.2006.07.008.

Rue, H. (2001). "Fast sampling of Gaussian Markov random fields". *Journal of the Royal Statistical Society, Ser. B* 63, pp. 325–338. DOI: 10.1111/1467-9868.00288.

Shephard, Neil (1994). "Partial non-Gaussian state space". *Biometrika* 81, pp. 115–131. DOI: 10.1093/biomet/81.1.115.

Silva, Ralph S. et al. (2006). "The Extended Generalized Inverse Gaussian Distribution for Log-Linear and Stochastic Volatility Models". *Brazilian Journal of Probability and Statistics* 20, pp. 67–91.

Strickland, Chris M. et al. (2008). "Parameterisation and efficient MCMC estimation of non-Gaussian state space models". *Computational Statistics and Data Analysis* 52, pp. 2911–2930. DOI: 10.1016/j.csda.2007.10.010.

Yu, Yaming and Xiao-Li Meng (2011). "To Center or Not to Center: That Is Not the Question—An Ancillarity-Suffiency Interweaving Strategy (ASIS) for Boosting MCMC Efficiency". *Journal of Computational and Graphical Statistics* 20, pp. 531–570. DOI: 10.1198/jcgs.2011.203main.