Bayesian Model Comparison:

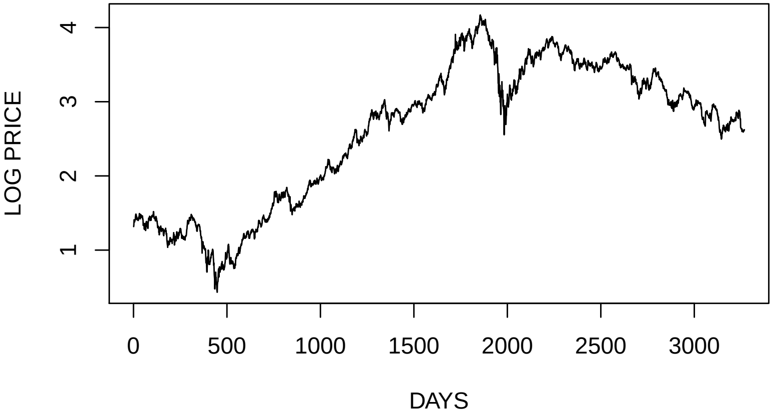
Modeling Petrobrás' log-returns

Hedibert Freitas Lopes

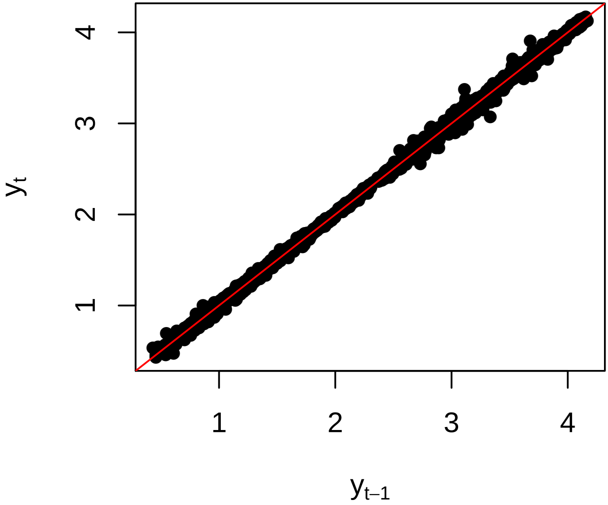
February 2014

Log price: $y_t = \log p_t$

Time span: 12/29/2000 - 12/31/2013 ($n = 3268$ days)

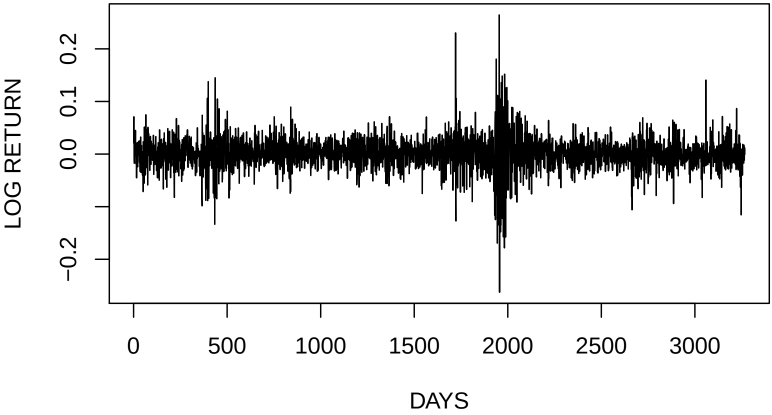


Scatterplot of y_{t-1} versus y_t

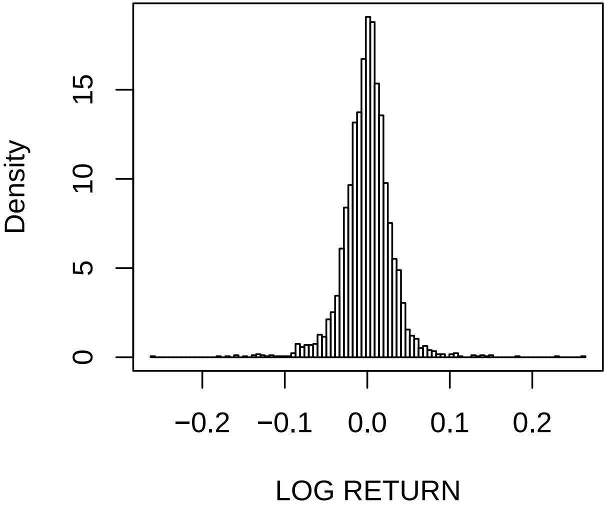


Log return: $r_t = y_t - y_{t-1} = \log(p_t/p_{t-1})$

Time span: 01/02/2001 - 12/31/2013 ($n = 3267$ days)



Histogram of r_t



Models

- Model $\mathcal{M}_0$: $r_1, \ldots, r_n$ iid $N(0, \sigma^2)$
- Model $\mathcal{M}_1$: $r_1, \ldots, r_n$ iid $N(\mu, \sigma^2)$
- Model $\mathcal{M}_2$: For $t = 2, \ldots, n$

$$y_t | y_{t-1} \sim N(\alpha + \beta y_{t-1}, \sigma^2),$$

for $\alpha, \beta \in \Re$ and $\sigma^2 > 0$.

Prior

First $n_0 = 1506$ days (2001-2006): Prior knowledge. Let $\tilde{y}_t$ and $\tilde{r}_t$ be log prices and log returns in the training sample. Let $m_{\tilde{r}}$ and $V_{\tilde{r}}$ be the sample mean and the sample variance of r_ts.

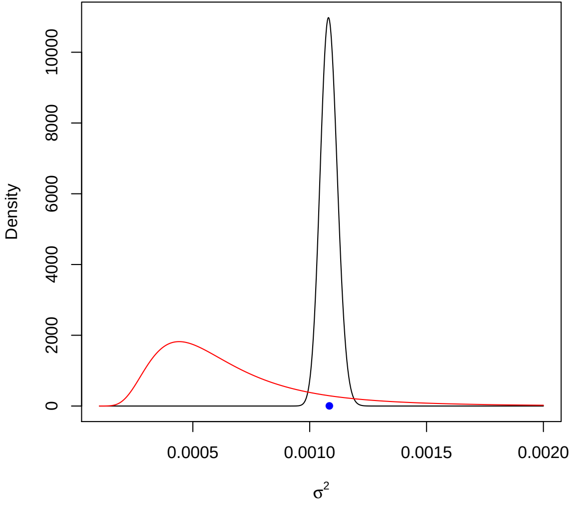
Last $n = 1760$ days (2007-2013): Inference and model comparison.

Model $\mathcal{M}_0$: $\sigma^2 \sim IG\,(5, 4m_{\tilde{r}^2})$

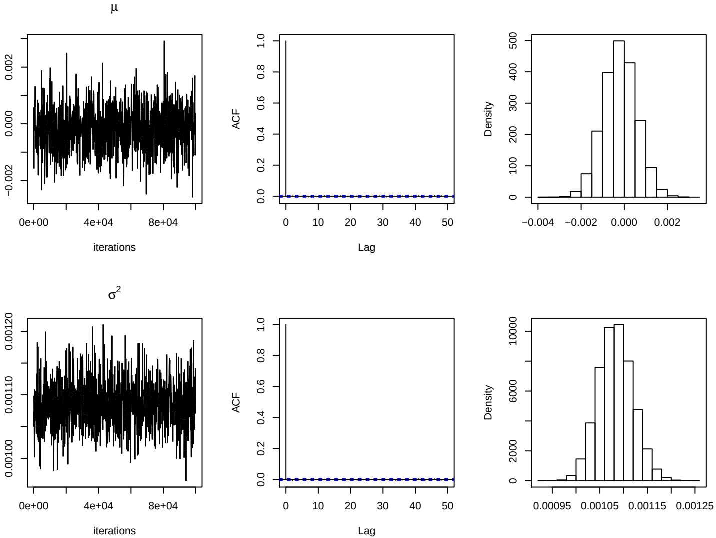
Model $\mathcal{M}_1$: $\mu \sim N(m_{\tilde{r}}, 100V_{\tilde{r}}/(n_0 - 1))$ and $\sigma^2 \sim IG(5, V_{\tilde{r}})$

Model $\mathcal{M}_2$: $\alpha \sim N(a, 100V_a)$, $\beta \sim N(b, 100V_b)$, $\sigma^2 \sim IG(5, 4s^2)$. (a, b, s^2) are MLE of $(\alpha, \beta, \sigma^2)$ based on the training sample. V_a and V_b are the MLE variances of estimators a and b, respectively.

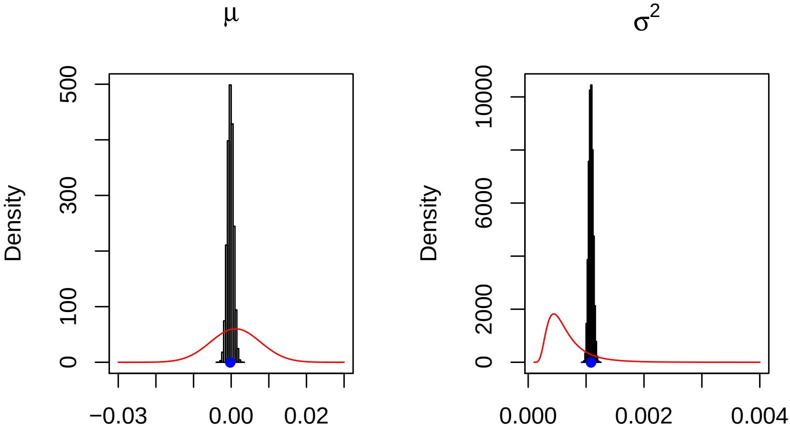
Model $\mathcal{M}_0$: Prior, posterior, MLE



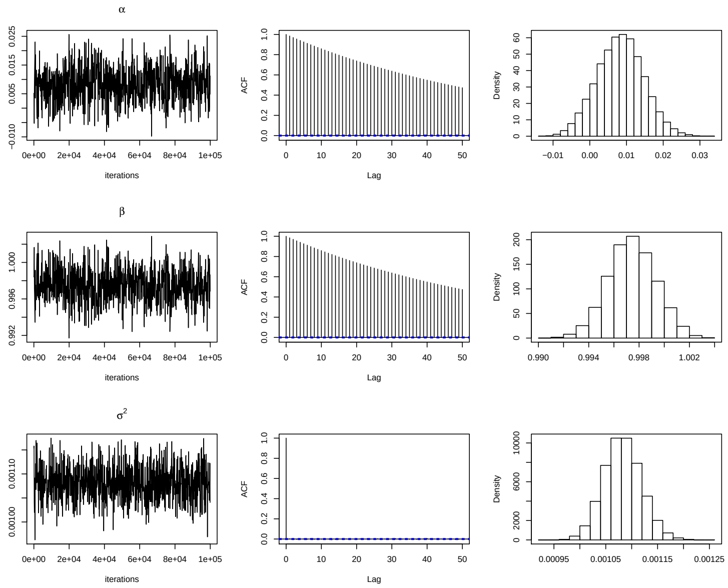
Model $\mathcal{M}_1$: Gibbs sampler output



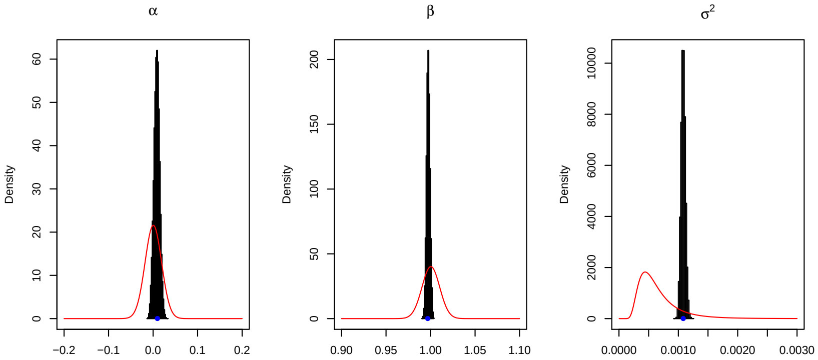
Model $\mathcal{M}_1$: Prior, posterior, MLE



Model $\mathcal{M}_2$: Gibbs sampler output



Model $\mathcal{M}_2$: Prior, posterior, MLE



Comparing the models

Recall that the posterior of θ_i under model $\mathcal{M}_i$ is

$$p(\theta_i|\texttt{data}, \mathcal{M}_i) \propto p(\theta_i|\mathcal{M}_i)p(\texttt{data}|\theta_i, \mathcal{M}_i),$$

while the marginal likelihood under model $\mathcal{M}_i$ is

$$p(\texttt{data}|\mathcal{M}_i) = \int p(\texttt{data}|\theta_i, \mathcal{M}_i)p(\theta_i|\mathcal{M}_i)d\theta_i.$$

Therefore, the posterior model probability of $\mathcal{M}_i$ is

$$Pr(\mathcal{M}_i|\texttt{data}) \propto Pr(\mathcal{M}_i)p(\texttt{data}|\mathcal{M}_i),$$

and the Bayes factor of model $\mathcal{M}_i$ against model $\mathcal{M}_j$ is

$$B_{ij} = \frac{Pr(\mathcal{M}_i)p(\texttt{data}|\mathcal{M}_i)}{Pr(\mathcal{M}_j)p(\texttt{data}|\mathcal{M}_j)}$$

Monte Carlo integration

A major (computational) difficulty is calculating the marginal likelihood (or predictive) under model $\mathcal{M}$, ie. $p(\texttt{data}|\mathcal{M})$.

The simplest *estimator* of $p(\texttt{data}|\mathcal{M})$ is obtained by Monte Carlo integration.

More specifically, when $\{\theta^{(1)}, \ldots, \theta^{(N)}\}$ represents a sample from the prior $p(\theta|\mathcal{M})$, then (for large N)

$$p^N(\texttt{data}|\mathcal{M}) = \frac{1}{N} \sum_{j=1}^{N} p(\texttt{data}|\theta^{(j)}, \mathcal{M}),$$

is the *simple Monte Carlo estimator* of $p(\texttt{data}|\mathcal{M})$.

Harmonic mean identity (and estimator)

An alternative way of calculating $p(\texttt{data}|\mathcal{M})$ is via the *harmonic mean* identity:

$$\frac{1}{p(\texttt{data}|\mathcal{M})} = \int \frac{1}{p(\texttt{data}|\theta, \mathcal{M})} p(\theta|\texttt{data}, \mathcal{M}) d\theta$$

Monte Carlo integration: When $\{\theta^{(1)}, \ldots, \theta^{(N)}\}$ represents a sample from the posterior $p(\theta|\texttt{data}, \mathcal{M})$, then (for large N)

$$p^N(\texttt{data}|\mathcal{M}) = \left(\frac{1}{N} \sum_{j=1}^{N} \frac{1}{p(\texttt{data}|\theta^{(j)}, \mathcal{M})} \right)^{-1}$$

is the *harmonic mean estimator* of $p(\texttt{data}|\mathcal{M})$.

Comparing $\mathcal{M}_0, \mathcal{M}_1$ and $\mathcal{M}_2$

Log predictive

$\mathcal{M}_i$	$\log p(\texttt{data}\|\mathcal{M}_i)$		Computation
	Monte Carlo	Harmonic Mean	
$\mathcal{M}_0$	3508.33	3508.33	Closed form
$\mathcal{M}_1$	3509.47	3506.30	Gibbs sampler
$\mathcal{M}_2$	3509.71	3504.69	Gibbs sampler

Posterior model probability

$\mathcal{M}_i$	$Pr(\mathcal{M}_i\|\texttt{data})$		Computation
	Monte Carlo	Harmonic Mean	
$\mathcal{M}_0$	12.4	86.4	Closed form
$\mathcal{M}_1$	38.5	11.3	Gibbs sampler
$\mathcal{M}_2$	49.1	2.3	Gibbs sampler

$\log p_{MC}(\texttt{data}|\mathcal{M}_i) = 3508.34$
$\log p_{HM}(\texttt{data}|\mathcal{M}_i) = 3509.8$

$\mathcal{M}_3$: GARCH(1,1) with t errors - Let us get more serious!

The GARCH(1,1) model with Student-t innovations:

$$\begin{aligned} r_t &\sim t_\nu(0, \rho h_t) \\ h_t &= \alpha_0 + \alpha_1 r_{t-1}^2 + \beta h_{t-1}, \end{aligned}$$

where $\alpha_0 > 0$, $\alpha_1 \geq 0$ and $\beta > 0$.

We set the initial variance to $h_0 = 0$ for convenience.

We let $\rho = (\nu - 2)/\nu$ so that

$$V(r_t | h_t) = \frac{\nu}{\nu - 2} \rho h_t = h_t.$$

Prior

Let $\psi = (\alpha', \beta, \nu)'$ and $\alpha = (\alpha_0, \alpha_1)'$.

The prior distribution of ψ is such that

$$p(\alpha, \beta, \mu) = p(\alpha)p(\beta)p(\nu)$$

where

$$\begin{aligned} \alpha &\sim N_2(\mu_\alpha, \Sigma_\alpha) I_{(\alpha>0)} \\ \beta &\sim N(\mu_\beta, \Sigma_\beta) I_{(\beta>0)} \end{aligned}$$

and

$$p(\nu) = \lambda \exp\{-\lambda(\nu - \delta)\} I_{(\lambda>\delta)}$$

for $\lambda > 0$ and $\delta \geq 2$, such that $E(\nu) = \delta + 1/\lambda$.

Normal case: $\lambda = 100$ and $\delta = 500$.

bayesGARCH

`bayesGARCH:` Bayesian Estimation of the GARCH(1,1) Model with Student-t Innovations

```
bayesGARCH(r,mu.alpha = c(0,0),Sigma.alpha=1000*diag(1,2),
           mu.beta=0,Sigma.beta=1000,
           lambda=0.01,delta=2,control=list())
```

Paper: Ardia and Hoogerheide (2010) Bayesian Estimation of the GARCH(1,1) Model with Student-t Innovations. *The R Journal*, **2**,41-47.

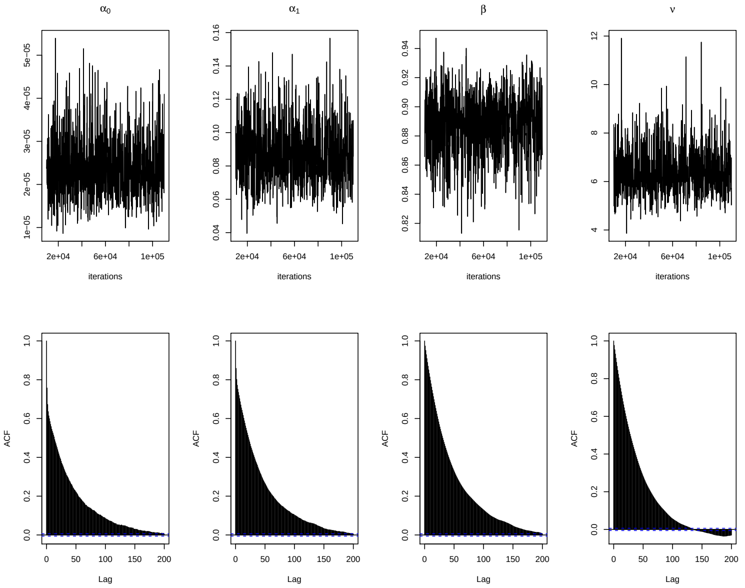
`http://cran.r-project.org/web/packages/bayesGARCH`

Example of R script

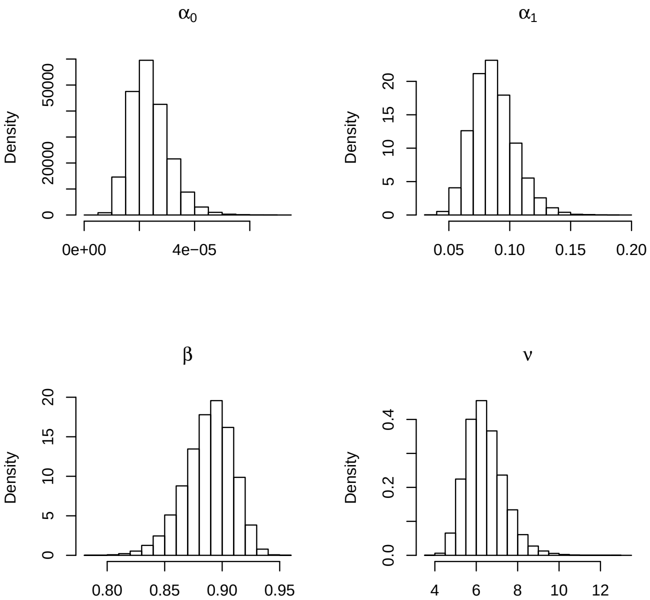
Recall that r_0 are Petrobras' returns for the first part of the data.

```
M0    = 10000       # to be discarded (burn-in)
M     = 10000       # kept for posterior inference
niter = M0+M

MCMC.initial = bayesGARCH(r0,mu.alpha=c(0,0),Sigma.alpha=1000*diag(1,2),
                          mu.beta=0,Sigma.beta=1000,lambda=0.01,delta=2,
                          control=list(n.chain=1,l.chain=niter,refresh=100))

draws = MCMC.initial$chain1

range = (M0+1):niter

par(mfrow=c(2,2))
ts.plot(draws[range,1],xlab="iterations",main=expression(alpha[0]),ylab="")
ts.plot(draws[range,2],xlab="iterations",main=expression(alpha[1]),ylab="")
ts.plot(draws[range,3],xlab="iterations",main=expression(beta),ylab="")
ts.plot(draws[range,4],xlab="iterations",main=expression(nu),ylab="")
```

Model $\mathcal{M}_3$: MCMC output

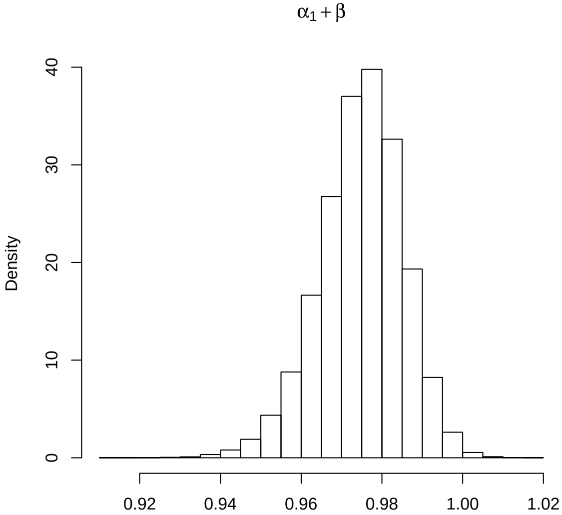


Model $\mathcal{M}_3$: Marginal posterior distributions



Model $\mathcal{M}_3$: $p(\alpha_1 + \beta | \texttt{data})$

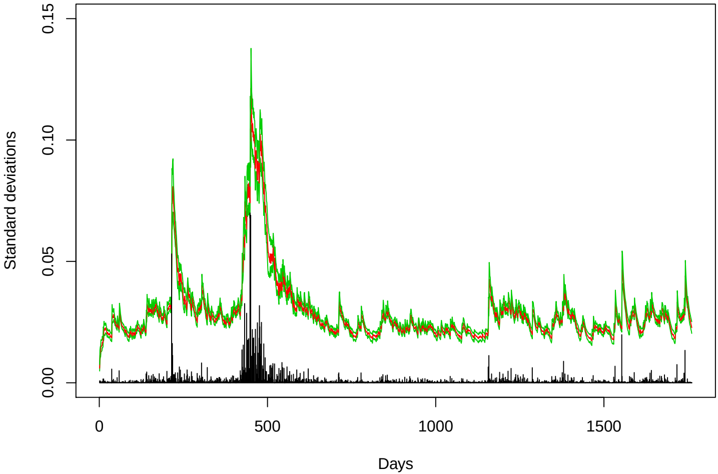
$$Pr(\alpha_1 + \beta > 1 | \texttt{data}) = 0.0034$$



Model $\mathcal{M}_3$: Quantiles from $p(h_t^{1/2}|\texttt{data})$

Percentiles 2.5%, 50% and 97.5% of $p(h_t^{1/2}|\texttt{data})$

Black vertical lines: r_t^2



Comparing all 4 models

Model $\mathcal{M}_i$	$\log p(\texttt{data}\|\mathcal{M}_i)$ Monte Carlo	Harmonic Mean	MCMC Scheme
$\mathcal{M}_0$	3508.33	3508.33	Closed form
$\mathcal{M}_1$	3509.47	3506.30	Gibbs sampler
$\mathcal{M}_2$	3509.71	3504.69	Gibbs sampler
$\mathcal{M}_3$	-	**3901.46**	Metropolis algorithm

$\mathcal{M}_4$: SV-AR(1) - For later in the course.

Model: The basic stochastic volatility model can be written as

$$\begin{aligned} r_t | h_t &\sim N(0, \exp\{h_t\}) \\ h_t | h_{t-1} &\sim N(\mu + \phi(h_{t-1} - \mu), \sigma^2) \\ h_0 &\sim N(\mu, \sigma^2/(1-\phi^2)) \end{aligned}$$

where $\mu \in \Re$, $\beta \in (-1,1)$ and $\sigma^2 > 0$.

Prior: A possible prior distribution is

$$\begin{aligned} \mu &\sim N(b_\mu, B_\mu) \\ \frac{\phi+1}{2} &\sim Beta(a_0, b_0) \qquad E(\phi) = \frac{2a_0}{a_0+b_0} - 1 \\ \sigma^2 &\sim G(1/2, 1/(2B_\sigma)) \qquad E(\sigma^2) = B_\sigma, \end{aligned}$$

and

$$V(\phi) = \frac{4a_0 b_0}{(a_0+b_0)^2(a_0+b_0+1)}$$

The SV-AR(1) model falls into the more general class of dynamic models (aka state-space models).

The usual estimation technique is the Kalman filter, which can not be directly implemented here due to the nonlinearity of the observation equation.

We will introduce (much later) the well-known Forward filtering backward sampling (FFBS) scheme that made posterior inference in the SV-AR(1) straightforward.

stochvol

stochvol: Efficient Bayesian Inference for SV Models

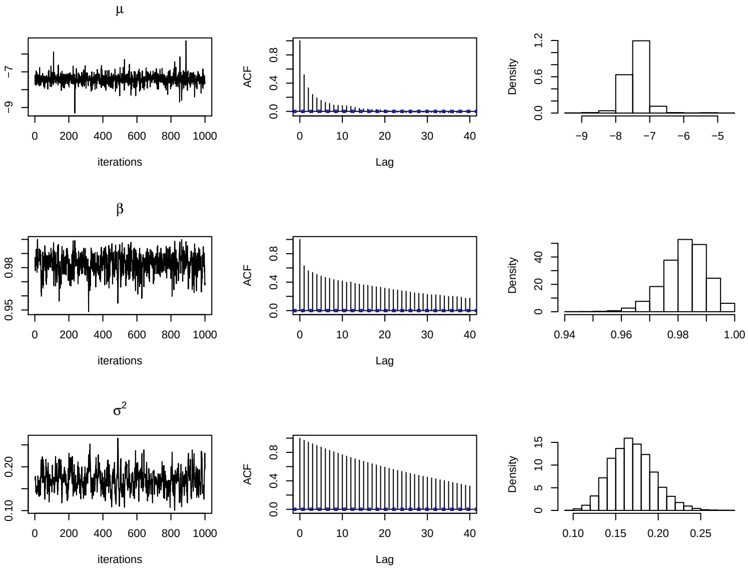
```
svsample(y,draws=10000,burnin=1000,priormu=c(-10,3),
         priorphi=c(5,1.5),priorsigma=1,thinpara=1,
         thinlatent=1,thintime=1,quiet=FALSE,
         startpara,startlatent,expert,...)
```

Example:

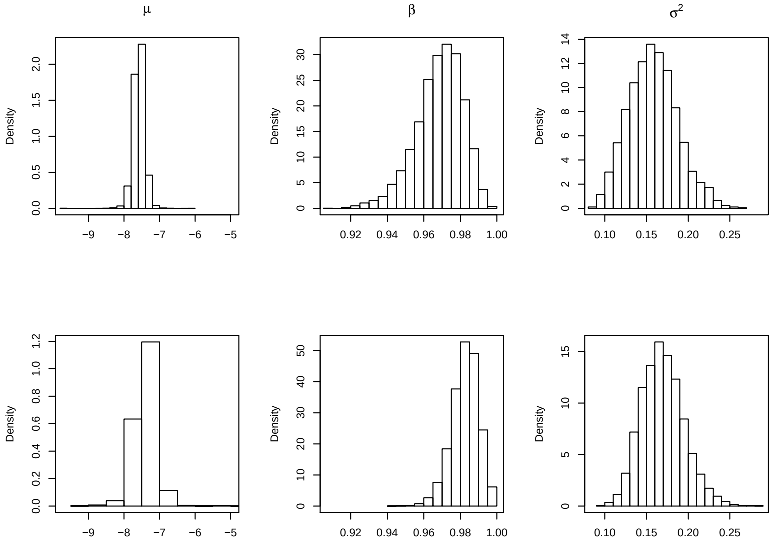
```
# Simulate a highly persistent SV process
# Obtain 5000 draws from the sampler
sim   = svsim(500,mu=-10,phi=0.99,sigma=0.2)
draws = svsample(sim$y,draws=5000,burnin=100,
                 priormu=c(-10,1),
                 priorphi=c(20,1.2),
                 priorsigma=0.2)
```

Paper: Kastner and Frühwirth-Schnatter (2013) Ancillarity-sufficiency interweaving strategy for boosting MCMC estimation of stochastic volatility models. *CSDA*, http://dx.doi.org/10.1016/j.csda.2013.01.002.

Model $\mathcal{M}_4$: MCMC output



Model $\mathcal{M}_4$: Marginal prior and posterior distributions



Model $\mathcal{M}_4$: Quantiles from $p(\exp\{h_t/2\}|\texttt{data})$

Percentiles 2.5%, 50% and 97.5% of $p(\exp\{h_t/2\}|\texttt{data})$
Black vertical lines: r_t^2