TUTORIAL: AN INTRODUCTION TO PARTICLE FILTERS

Hedibert Freitas Lopes

Associate Professor of Econometrics and Statistics
The University of Chicago Booth School of Business
5807 South Woodlawn Avenue, Chicago, IL 60637
http://faculty.chicagobooth.edu/hedibert.lopes/research
hlopes@chicagobooth.edu

http://faculty.chicagobooth.edu/hedibert.lopes/teaching/ESTE2009/smc-tutorial.html

.

XIII Brazilian School of Time Series and Econometrics
ICMC-SP July 21-24 2009
São Carlos, Brazil

Tutorial outline

- Part 1: Dynamic linear models (DLM)
- Part 2: Stochastic volatility models (SVM)
- Part 3: Sequential Monte Carlo (SMC) methods
- Part 4: SMC with parameter learning
- Part 5: SMC in SVM

R code

- Part 1: `dlm.R` - `dlm-ffbs.R`
- Part 2: `sv-ar1.R`
- Part 3: `dlm-smc.R` - `boostrapfilter-stepbystep.R` - `dlm-smc-smoothing.R`
- Part 4: `dlm-smc-learningsig2-LW.R` - `dlm-smc-learningsig2-PL.R` - `nonlinearmodel-LW.R`
- Part 5: `sv-LW.R`

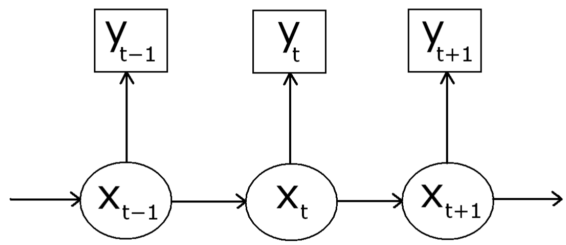
DYNAMIC LINEAR MODELS

Example i. local level model

Let us start this tutorial reviewing the most basic normal dynamic linear model, i.e. West and Harrison's (1997) local level model:

$$
\begin{aligned}
y_{t+1}|x_{t+1},\theta &\sim N(x_{t+1},\sigma^2)\\
x_{t+1}|x_t,\theta &\sim N(x_t,\tau^2)
\end{aligned}
$$

where $x_0 \sim N(m_0, C_0)$ and $\theta = (\sigma^2, \tau^2)$ fixed (for now).



Example i. Evolution, prediction and updating

Let $y^t = (y_1, \ldots, y_t)$.

$$p(x_t|y^t) \Longrightarrow p(x_{t+1}|y^t) \Longrightarrow p(y_{t+1}|x_t) \Longrightarrow p(x_{t+1}|y^{t+1})$$

- Posterior at t: $(x_t|y^t) \sim N(m_t, C_t)$
- Prior at $t+1$: $(x_{t+1}|y^t) \sim N(m_t, R_{t+1})$
- Marginal likelihood: $(y_{t+1}|y^t) \sim N(m_t, Q_{t+1})$
- Posterior at $t+1$: $(x_{t+1}|y^{t+1}) \sim N(m_{t+1}, C_{t+1})$

where $R_{t+1} = C_t + \tau^2$, $Q_{t+1} = R_{t+1} + \sigma^2$, $A_{t+1} = R_{t+1}/Q_{t+1}$, $C_{t+1} = A_{t+1}\sigma^2$, and $m_{t+1} = (1 - A_{t+1})m_t + A_{t+1}y_{t+1}$.

Example i. Backward smoothing

For $t = n$, $x_n|y^n \sim N(m_n^n, C_n^n)$, where

$$
\begin{aligned}
m_n^n &= m_n\\
C_n^n &= C_n
\end{aligned}
$$

For $t < n$, $x_t|y^n \sim N(m_t^n, C_t^n)$, where

$$
\begin{aligned}
m_t^n &= (1 - B_t)m_t + B_t m_{t+1}^n\\
C_t^n &= (1 - B_t)C_t + B_t^2 C_{t+1}^n
\end{aligned}
$$

and

$$B_t = \frac{C_t}{C_t + \tau^2}$$

Example i. Backward sampling

For $t = n$, $x_n|y^n \sim N(a_n^n, R_n^n)$, where

$$\begin{aligned} a_n^n &= m_n \\ R_n^n &= C_n \end{aligned}$$

For $t < n$, $x_t|x_{t+1}, y^n \sim N(a_t^n, R_t^n)$, where

$$\begin{aligned} a_t^n &= (1 - B_t)m_t + B_t x_{t+1} \\ R_t^n &= B_t \tau^2 \end{aligned}$$

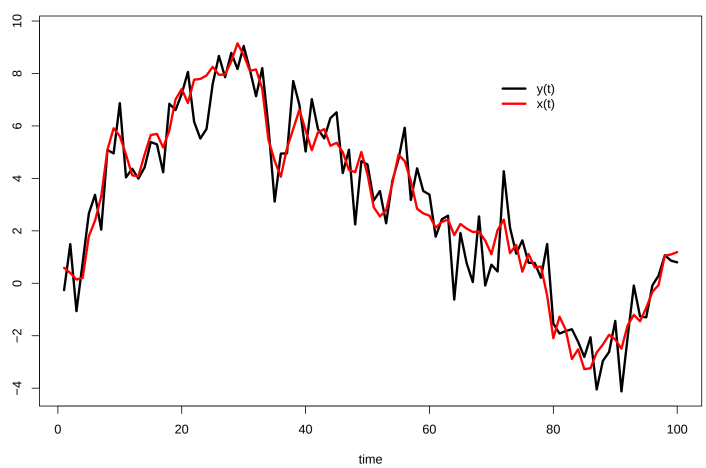
and

$$B_t = \frac{C_t}{C_t + \tau^2}$$

This is basically the Forward filtering, backward sampling algorithm used to sample from $p(x^n|y^n)$ (Carter and Kohn, 1994 and Frühwirth-Schnatter, 1994).

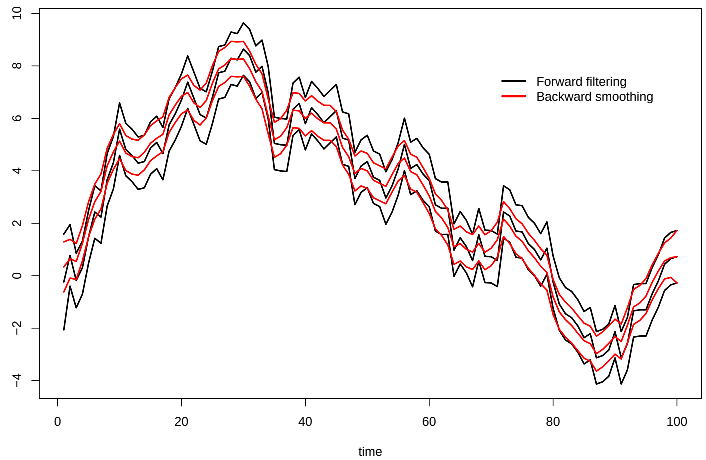
Example i. Simulated data

$n = 100$, $\sigma^2 = 1.0$, $\tau^2 = 0.5$ and $x_0 = 0$.



Example i. $p(x_t|y^t, \theta)$ **versus** $p(x_t|y^n, \theta)$

$m_0 = 0.0$ and $C_0 = 10.0$



Example i. $p(\theta|y^n) \propto p(\theta)p(y^n|\theta)$

We showed earlier that the sequential predictive density is

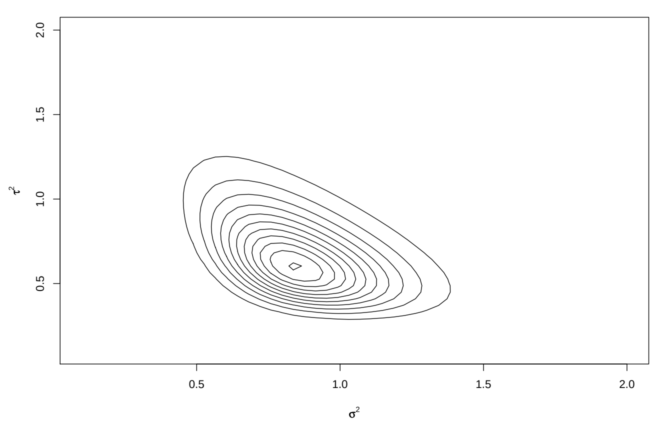
$$(y_t|y^{t-1}) \sim N(m_{t-1}, Q_t)$$

where both m_{t-1} and Q_t were presented before and are functions of $\theta = (\sigma^2, \tau^2)$, y^{t-1}, m_0 and C_0.

Therefore, by Bayes' rule,

$$\begin{aligned} p(\theta|y^n) &\propto p(\theta)p(y^n|\theta) \\ &= p(\theta)\prod_{t=1}^{n} f_N(y_t|m_{t-1}, Q_t). \end{aligned}$$

Example i. $p(y|\sigma^2, \tau^2)$



Example i. MCMC scheme

- Sample θ from $p(\theta|y^n, x^n)$

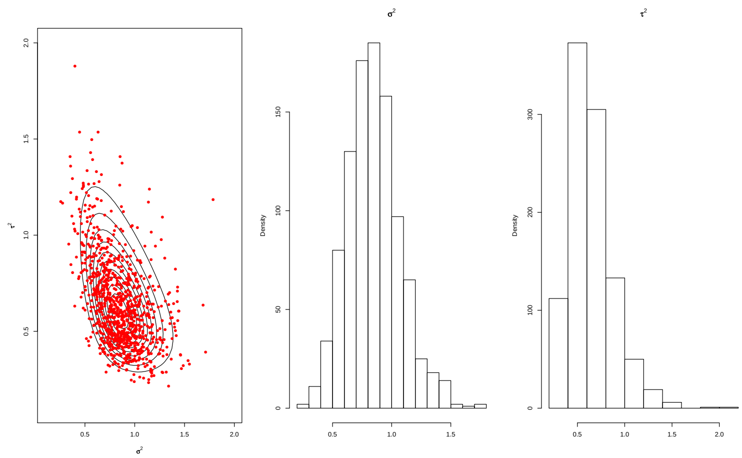
$$p(\theta|y^n, x^n) \propto p(\theta) \prod_{t=1}^{n} p(y_t|x_t, \theta) p(x_t|x_{t-1}, \theta).$$

- Sample x^n from $p(x^n|y^n, \theta)$

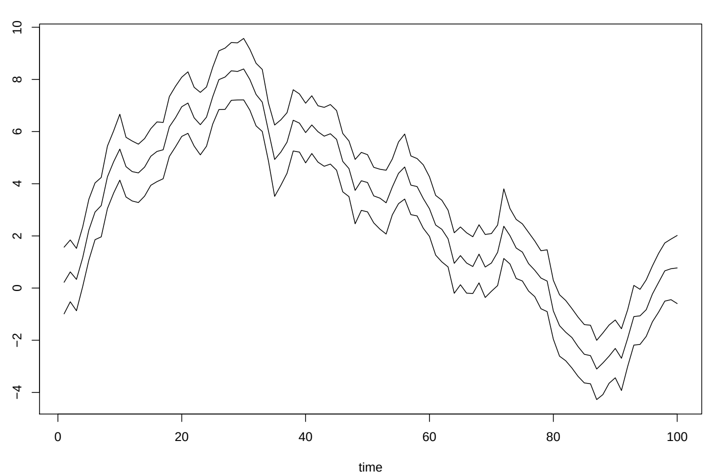
$$p(x^n|y^n, \theta) = \prod_{t=1}^{n} f_N(x_t|a_t^n, R_t^n)$$

where, $a_n^n = a_n$ and $C_n^n = C_n$ and, for $t = 1, \ldots, n-1$, $a_t^n = (1 - B_t)m_t + B_t x_{t+1}$, $R_t^n = B_t \tau^2$ and $B_t = C_t/(C_t + \tau^2)$.

Example i. $p(\sigma^2, \tau^2|y^n)$



Example i. $p(x_t|y^n)$



Example ii. Comparing sampling schemes

Based on Gamerman, Reis and Salazar (2006) Comparison of sampling schemes for dynamic linear models. *International Statistical Review*, 74, 203-214.

First order DLM with $\sigma^2 = 1$

$$\begin{aligned} y_t &= x_t + \epsilon_t \,, & \epsilon_t \sim N(0,1) \\ x_t &= x_{t-1} + \omega_t \,, & \omega_t \sim N(0,\tau^2), \end{aligned}$$

with $(n, \tau^2) \in \{(100, .01), (100, .5), (1000, .01), (1000, .5)\}$.

400 runs: 100 replications per combination.

Priors: $x_1 \sim N(0, 10)$ and σ^2 and τ^2 have inverse Gammas with means set at true values and coefficients of variation set at 10.

Posterior inference: based on 20,000 MCMC draws.

Schemes

Scheme I: Sampling $x_1, \ldots, x_n, \sigma^2$ and τ^2 from their conditionals.

Scheme II: Sampling x^n, σ^2 and τ^2 from their conditionals.

Computing times relative to scheme I

Scheme	n=100	n=1000
II	1.7	1.9

Sample averages (over 100 replications) of effective sample size n_{eff} based on σ^2 (see Gamerman and Lopes, 2006, Chapter 4):

		Scheme	
τ^2	n	I	II
0.01	1000	242	8938
0.01	100	3283	13685
0.50	1000	409	3043
0.50	100	1694	3404

Lessons from Examples i. and ii.

Sequential learning in non-normal and nonlinear dynamic models $p(y_{t+1}|x_{t+1})$ and $p(x_{t+1}|x_t)$ in general rather difficult since

$$\begin{aligned} p(x_{t+1}|y^t) &= \int p(x_{t+1}|x_t)p(x_t|y^t)dx_t \\ p(x_{t+1}|y^{t+1}) &\propto p(y_{t+1}|x_{t+1})p(x_{t+1}|y^t) \end{aligned}$$

are usually unavailable in closed form.

Over the last 20 years:

- FFBS for conditionally Gaussian DLMs;

- Gamerman (1998) for generalized DLMs;
- Carlin, Polson and Stoffer (2002) for more general DMs.

STOCHASTIC VOLATILITY MODELS

Stochastic volatility model

The canonical stochastic volatility model (SV-AR(1), hereafter), is

$$\begin{aligned} y_t &= e^{h_t/2}\varepsilon_t \\ h_t &= \mu + \phi h_{t-1} + \tau\eta_t \end{aligned}$$

where ε_t and η_t are $N(0,1)$ shocks with $E(\varepsilon_t\eta_{t+h}) = 0$ for all h and $E(\varepsilon_t\varepsilon_{t+l}) = E(\eta_t\eta_{t+l}) = 0$ for all $l \neq 0$.

τ^2: volatility of the log-volatility.

$|\phi| < 1$ then h_t is a stationary process.

Let $y^n = (y_1,\ldots,y_n)'$, $h^n = (h_1,\ldots,h_n)'$ and $h_{a:b} = (h_a,\ldots,h_b)'$.

Prior information

Uncertainty about the initial log volatility is $h_0 \sim N(m_0, C_0)$.

Let $\theta = (\mu,\phi)'$, then the prior distribution of (θ,τ^2) is normal-inverse gamma, i.e. $(\theta,\tau^2) \sim NIG(\theta_0, V_0, \nu_0, s_0^2)$:

$$\begin{aligned} \theta|\tau^2 &\sim N(\theta_0, \tau^2 V_0) \\ \tau^2 &\sim IG(\nu_0/2, \nu_0 s_0^2/2) \end{aligned}$$

For example, if $\nu_0 = 10$ and $s_0^2 = 0.018$ then

$$\begin{aligned} E(\tau^2) &= \frac{\nu_0 s_0^2/2}{\nu_0/2-1} = 0.0225 \\ Var(\tau^2) &= \frac{(\nu_0 s_0^2/2)^2}{(\nu_0/2-1)^2(\nu_0/2-2)} = (0.013)^2 \end{aligned}$$

Hyperparameters: m_0, C_0, θ_0, V_0, ν_0 and s_0^2.

Posterior inference

The SV-AR(1) is a dynamic model and posterior inference via MCMC for the the latent log-volatility states h_t can be performed in at least two ways.

Let $h_{-t} = (h_{0:(t-1)}, h_{(t+1):n})$, for $t = 1,\ldots,n-1$ and $h_{-n} = h_{1:(n-1)}$.

- Individual moves for h_t
  - $(\theta,\tau^2|h^n,y^n)$
  - $(h_t|h_{-t},\theta,\tau^2,y^n)$, for $t = 1,\ldots,n$

- Block move for h^n
  - $(\theta, \tau^2 | h^n, y^n)$
  - $(h^n | \theta, \tau^2, y^n)$

Sampling $(\theta, \tau^2 | h^n, y^n)$

Conditional on $h_{0:n}$, the posterior distribution of (θ, τ^2) is also normal-inverse gamma:

$$(\theta, \tau^2 | y^n, h_{0:n}) \sim NIG(\theta_1, V_1, \nu_1, s_1^2)$$

where $X = (1_n, h_{0:(n-1)})$, $\nu_1 = \nu_0 + n$

$$\begin{aligned} V_1^{-1} &= V_0^{-1} + X'X \\ V_1^{-1}\theta_1 &= V_0^{-1}\theta_0 + X'h_{1:n} \\ \nu_1 s_1^2 &= \nu_0 s_0^2 + (y - X\theta_1)'(y - X\theta_1) + (\theta_1 - \theta_0)'V_0^{-1}(\theta_1 - \theta_0) \end{aligned}$$

Sampling $(h_0 | \theta, \tau^2, h_1)$

Combining

$$h_0 \sim N(m_0, C_0)$$

and

$$h_1 | h_0 \sim N(\mu + \phi h_0, \tau^2)$$

leads to (by Bayes' theorem)

$$h_0 | h_1 \sim N(m_1, C_1)$$

where

$$\begin{aligned} C_1^{-1} m_1 &= C_0^{-1} m_0 + \phi \tau^{-2} (h_1 - \mu) \\ C_1^{-1} &= C_0^{-1} + \phi^2 \tau^{-2} \end{aligned}$$

Conditional prior distribution of h_t

Given h_{t-1}, θ and τ^2, it can be shown that, for $t = 1, \ldots, n-1$,

$$\begin{pmatrix} h_t \\ h_{t+1} \end{pmatrix} \sim N\left\{ \begin{pmatrix} \mu + \phi h_{t-1} \\ (1+\phi)\mu + \phi^2 h_{t-1} \end{pmatrix}, \tau^2 \begin{pmatrix} 1 & \phi \\ \phi & (1+\phi^2) \end{pmatrix} \right\}$$

so $E(h_t | h_{t-1}, h_{t+1}, \theta, \tau^2)$ and $V(h_t | h_{t-1}, h_{t+1}, \theta, \tau^2)$ are

$$\begin{aligned} \mu_t &= \left(\frac{1-\phi}{1+\phi^2} \right) \mu + \left(\frac{\phi}{1+\phi^2} \right) (h_{t-1} + h_{t+1}) \\ \nu^2 &= \tau^2 (1+\phi^2)^{-1} \end{aligned}$$

respectively. Therefore,

$$\begin{aligned} (h_t | h_{t-1}, h_{t+1}, \theta, \tau^2) &\sim N(\mu_t, \nu^2) \qquad t = 1, \ldots, n-1 \\ (h_n | h_{n-1}, \theta, \tau^2) &\sim N(\mu_n, \tau^2) \end{aligned}$$

where $\mu_n = \mu + \phi h_{n-1}$.

Sampling h_t via random walk Metropolis

Let $\nu_t^2 = \nu^2$ for $t = 1, \ldots, n-1$ and $\nu_n^2 = \tau^2$, then

$$p(h_t|h_{-t}, y^n, \theta, \tau^2) \quad = \quad f_N(h_t; \mu_t, \nu_t^2) f_N(y_t; 0, e^{h_t})$$

for $t = 1, \ldots, n$.

A simple random walk Metropolis algorithm with tuning variance v_h^2 would work as follows:

For $t = 1, \ldots, n$

1. Current state: $h_t^{(j)}$
2. Sample h_t^* from $N(h_t^{(j)}, v_h^2)$
3. Compute the acceptance probability

$$\alpha = \min \left\{ 1, \frac{f_N(h_t^*; \mu_t, \nu_t^2) f_N(y_t; 0, e^{h_t^*})}{f_N(h_t^{(j)}; \mu_t, \nu_t^2) f_N(y_t; 0, e^{h_t^{(j)}})} \right\}$$

4. New state:

$$h_t^{(j+1)} = \begin{cases} h_t^* & \text{w. p. } \alpha \\ h_t^{(j)} & \text{w. p. } 1-\alpha \end{cases}$$

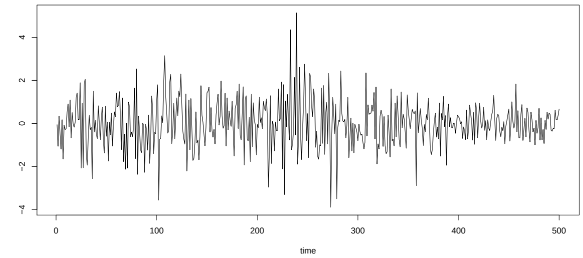
Example i. Simulated data

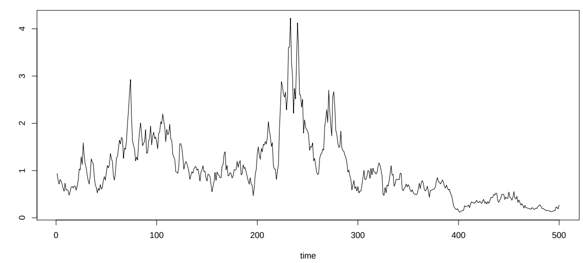
- Simulation setup
  - $n = 500$
  - $h_0 = 0.0$
  - $\mu = -0.00645$
  - $\phi = 0.99$
  - $\tau^2 = 0.15^2$

- Prior distribution
  - $\mu \sim N(0, 100)$
  - $\phi \sim N(0, 100)$
  - $\tau^2 \sim IG(10/2, 0.28125/2)$
  - $h_0 \sim N(0, 100)$

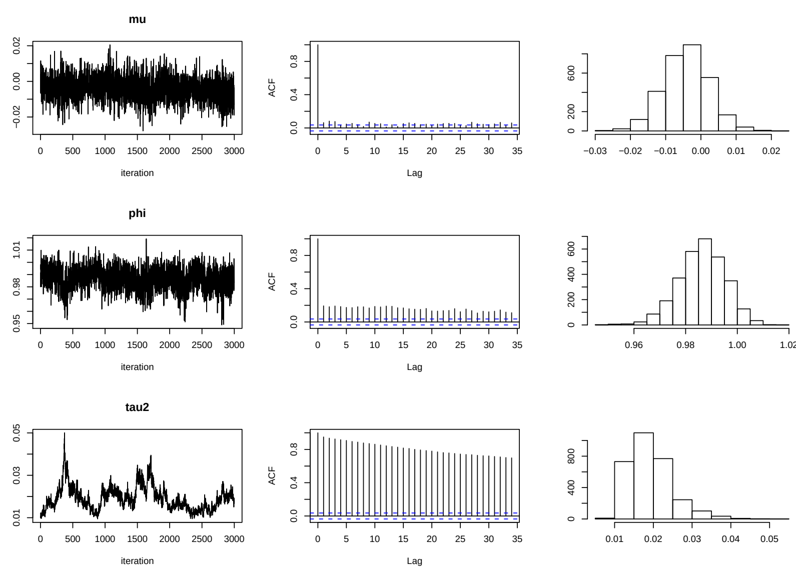
- MCMC setup
  - $M_0 = 1,000$
  - $M = 1,000$

Time series of y_t and $\exp\{h_t\}$

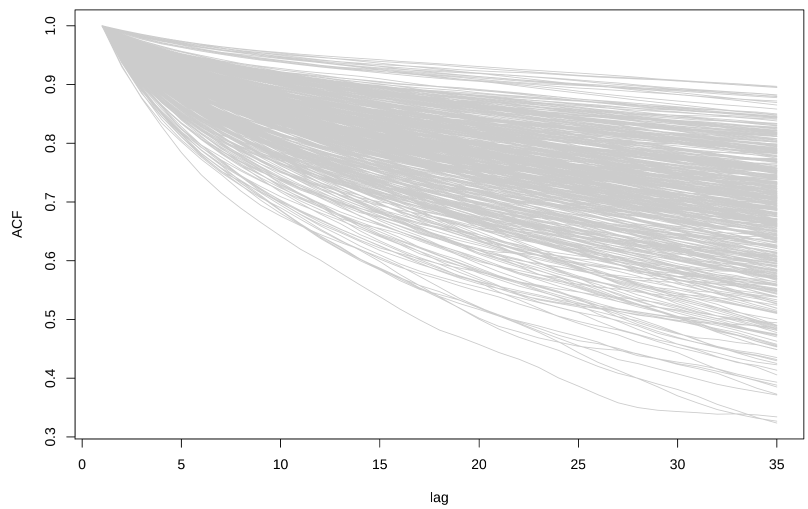




Parameters

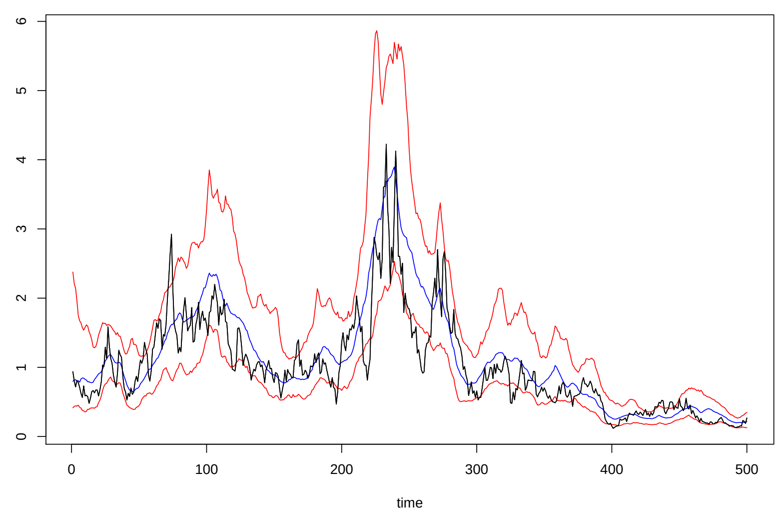


Autocorrelation of h_t



Volatilities

Tuning parameter: $v_h^2 = 0.01$



Sampling h_t via independent Metropolis-Hastings

The full conditional distribution of h_t is given by

$$\begin{aligned} p(h_t|h_{-t}, y^n, \theta, \tau^2) &= p(h_t|h_{t-1}, h_{t+1}, \theta, \tau^2)p(y_t|h_t) \\ &= f_N(h_t; \mu_t, \nu^2) f_N(y_t; 0, e^{h_t}). \end{aligned}$$

Kim, Shephard and Chib (1998) explored the fact that

$$\log p(y_t|h_t) = \text{const} - \frac{1}{2}h_t - \frac{y_t^2}{2}\exp(-h_t)$$

and that a Taylor expansion of $\exp(-h_t)$ around μ_t leads to

$$
\begin{aligned}
\log p(y_t|h_t) &\approx \text{const} - \frac{1}{2}h_t - \frac{y_t^2}{2}\left(e^{-\mu_t} - (h_t-\mu_t)e^{-\mu_t}\right) \\
g(h_t) &= \exp\left\{-\frac{1}{2}h_t(1-y_t^2e^{-\mu_t})\right\}
\end{aligned}
$$

Proposal distribution

Let $\nu_t^2 = \nu^2$ for $t = 1, \ldots, n-1$ and $\nu_n^2 = \tau^2$.

Then, by combining $f_N(h_t; \mu_t, \nu_t^2)$ and $g(h_t)$, for $t = 1, \ldots, n$, leads to the following proposal distribution:

$$
q(h_t|h_{-t}, y^n, \theta, \tau^2) \equiv N\left(h_t; \tilde{\mu}_t, \nu_t^2\right)
$$

where $\tilde{\mu}_t = \mu_t + 0.5\nu_t^2(y_t^2e^{-\mu_t} - 1)$.

Metropolis-Hastings algorithm

For $t = 1, \ldots, n$

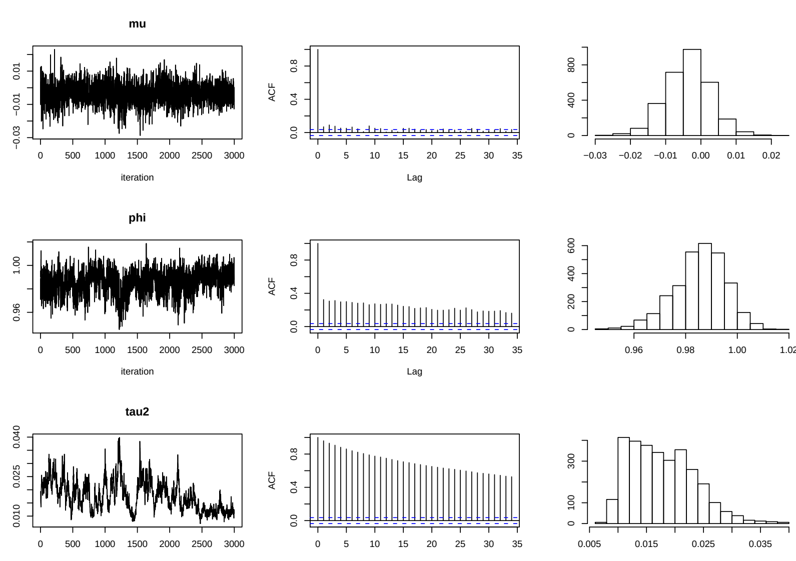
1. Current state: $h_t^{(j)}$
2. Sample h_t^* from $N(\tilde{\mu}_t, \nu_t^2)$
3. Compute the acceptance probability

$$
\alpha = \min\left\{1, \frac{f_N(h_t^*; \mu_t, \nu_t^2) f_N(y_t; 0, e^{h_t^*})}{f_N(h_t^{(j)}; \mu_t, \nu_t^2) f_N(y_t; 0, e^{h_t^{(j)}})} \times \frac{f_N(h_t^{(j)}; \tilde{\mu}_t, \nu_t^2)}{f_N(h_t^*; \tilde{\mu}_t, \nu_t^2)}\right\}
$$

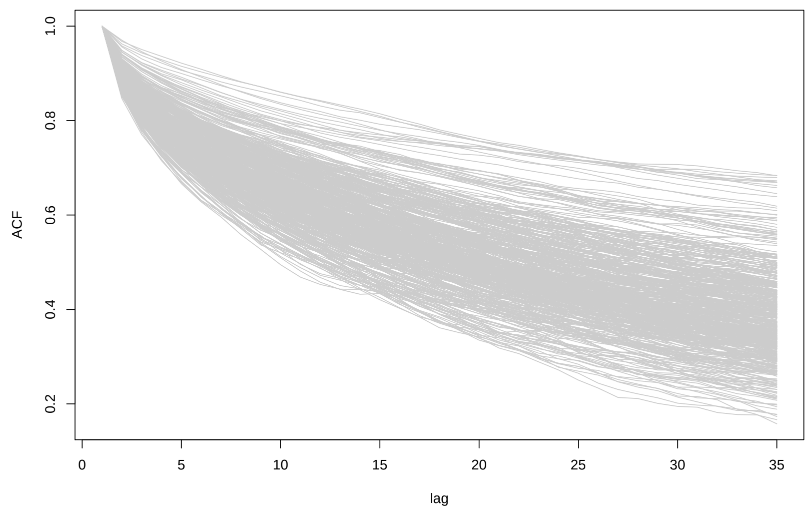
4. New state:

$$
h_t^{(j+1)} = \begin{cases} h_t^* & \text{w. p. } \alpha \\ h_t^{(j)} & \text{w. p. } 1-\alpha \end{cases}
$$

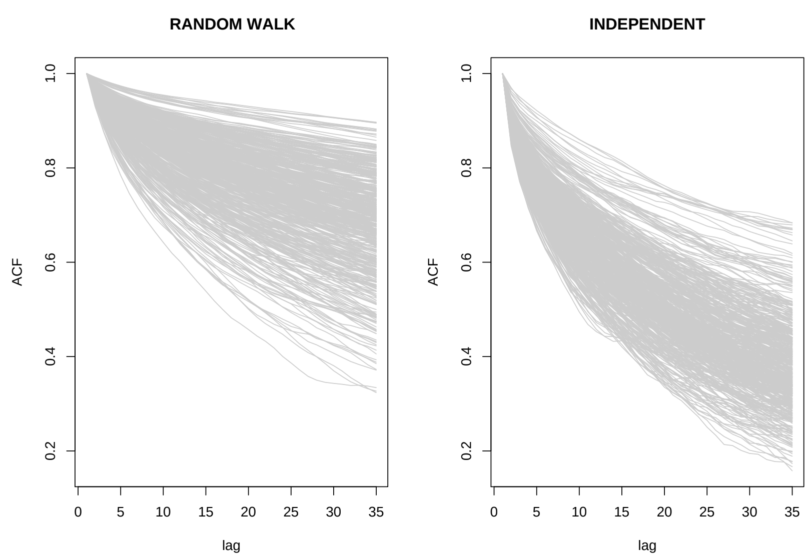
Parameters



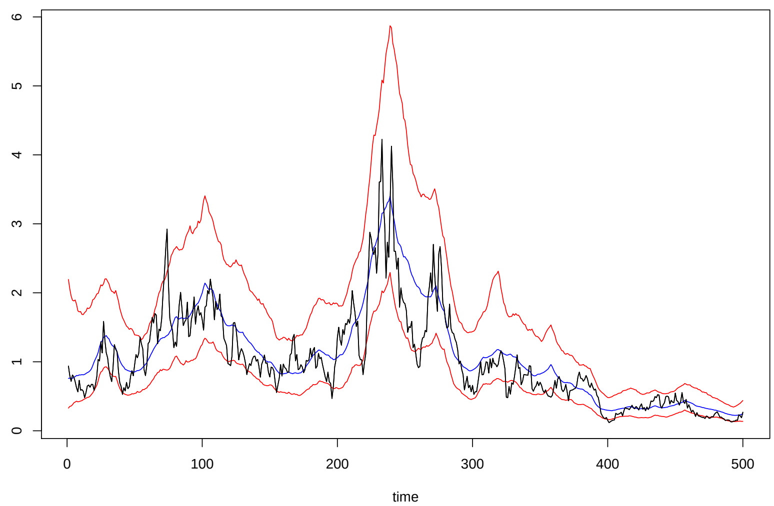
Autocorrelation of h_t



Autocorrelations of h_t for both schemes



Volatilities



Sampling h^n - normal approximation and FFBS

Let $y_t^* = \log y_t^2$ and $\epsilon_t = \log \varepsilon_t^2$.

The SV-AR(1) is a DLM with nonnormal observational errors, i.e.

$$\begin{aligned} y_t^* &= h_t + \epsilon_t \\ h_t &= \mu + \phi h_{t-1} + \tau \eta_t \end{aligned}$$

where $\eta_t \sim N(0,1)$.

The distribution of ϵ_t is $\log \chi_1^2$, where

$$\begin{aligned} E(\epsilon_t) &= -1.27 \\ V(\epsilon_t) &= \frac{\pi^2}{2} = 4.935 \end{aligned}$$

Normal approximation

Let ϵ_t be approximated by $N(\alpha, \sigma^2)$, $z_t = y_t^* - \alpha$, $\alpha = -1.27$ and $\sigma^2 = \pi^2/2$.

Then

$$\begin{aligned} z_t &= h_t + \sigma v_t \\ h_t &= \mu + \phi h_{t-1} + \tau \eta_t \end{aligned}$$

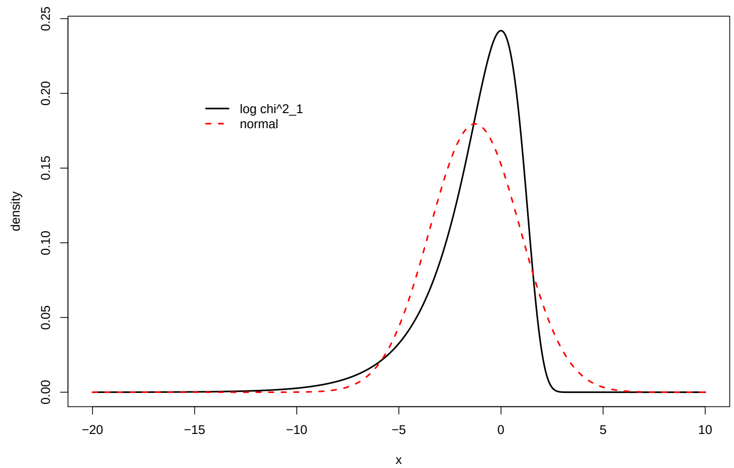
is a simple DLM where v_t and η_t are $N(0,1)$.

Sampling from

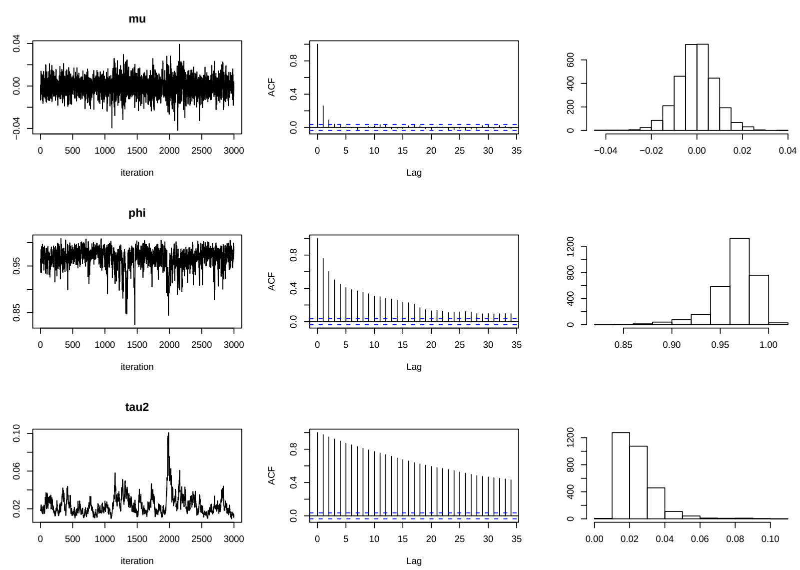
$$p(h^n | \theta, \tau^2, \sigma^2, z^n)$$

can be performed by the FFBS algorithm.

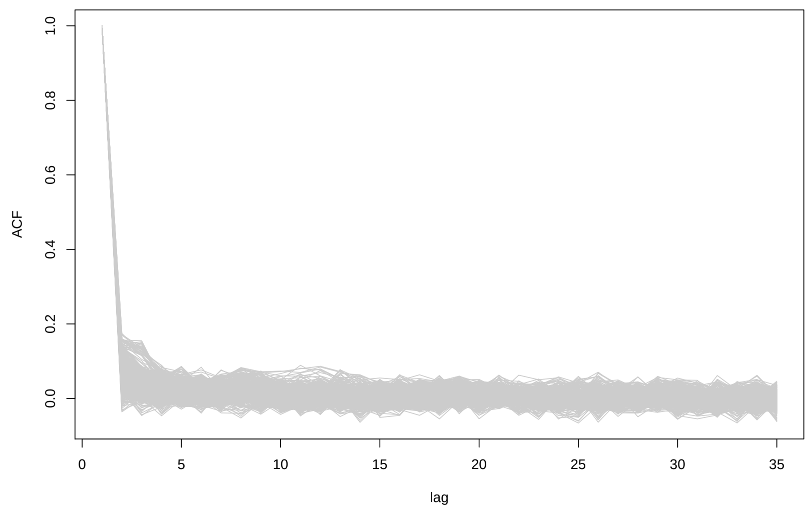
$\log \chi_1^2$ **and** $N(-1.27, \pi^2/2)$



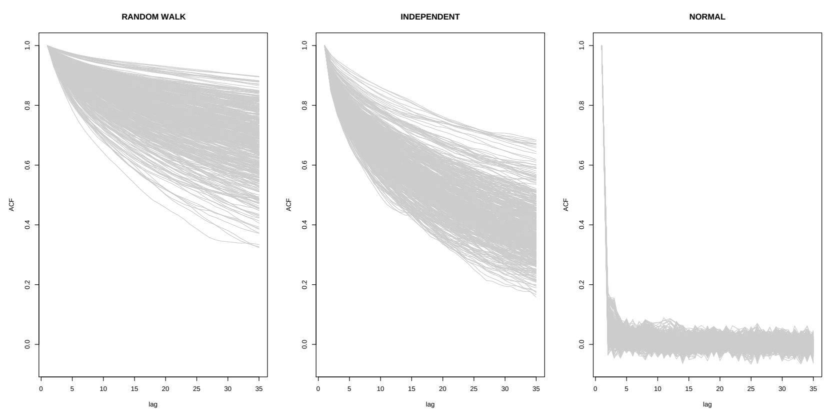
Parameters



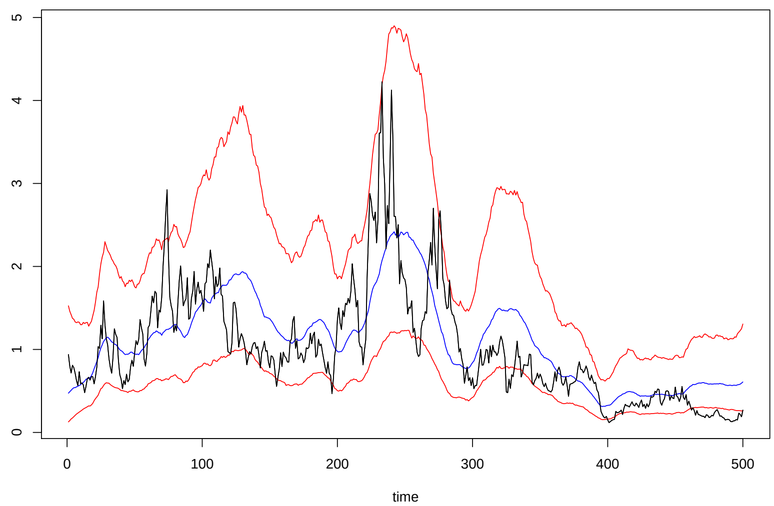
Autocorrelation of h_t



Autocorrelations of h_t for the three schemes



Volatilities



Sampling h^n - mixtures of normals and FFBS

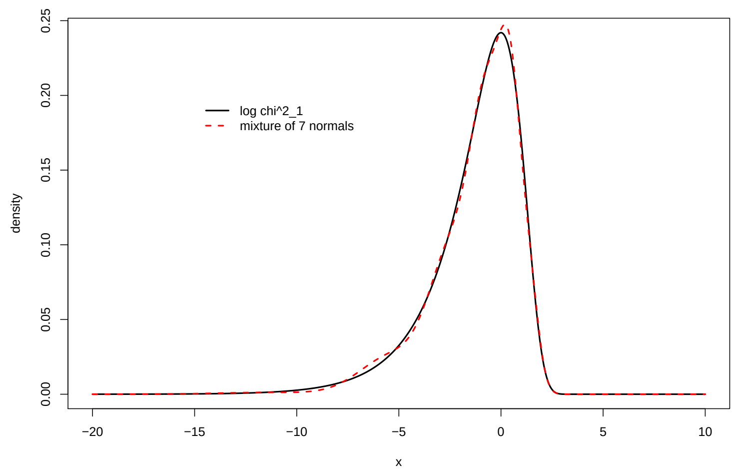
The $\log \chi_1^2$ distribution can be approximated by

$$\sum_{i=1}^{7} \pi_i N(\mu_i, \omega_i^2)$$

where

i	π_i	μ_i	ω_i^2
1	0.00730	-11.40039	5.79596
2	0.10556	-5.24321	2.61369
3	0.00002	-9.83726	5.17950
4	0.04395	1.50746	0.16735
5	0.34001	-0.65098	0.64009
6	0.24566	0.52478	0.34023
7	0.25750	-2.35859	1.26261

$\log \chi_1^2$ **and** $\sum_{i=1}^{7} \pi_i N(\mu_i, \omega_i^2)$



Mixture of normals

Using an argument from the Bayesian analysis of mixture of normal, let $z_1, \ldots, z_n$ be unobservable (latent) indicator variables such that $z_t \in \{1, \ldots, 7\}$ and $Pr(z_t = i) = \pi_i$, for $i = 1, \ldots, 7$.

Therefore, conditional on the z's, y_t is transformed into $\log y_t^2$,

$$\begin{aligned} \log y_t^2 &= h_t + \log \varepsilon_t^2 \\ h_t &= \mu + \phi h_{t-1} + \tau_\eta \eta_t \end{aligned}$$

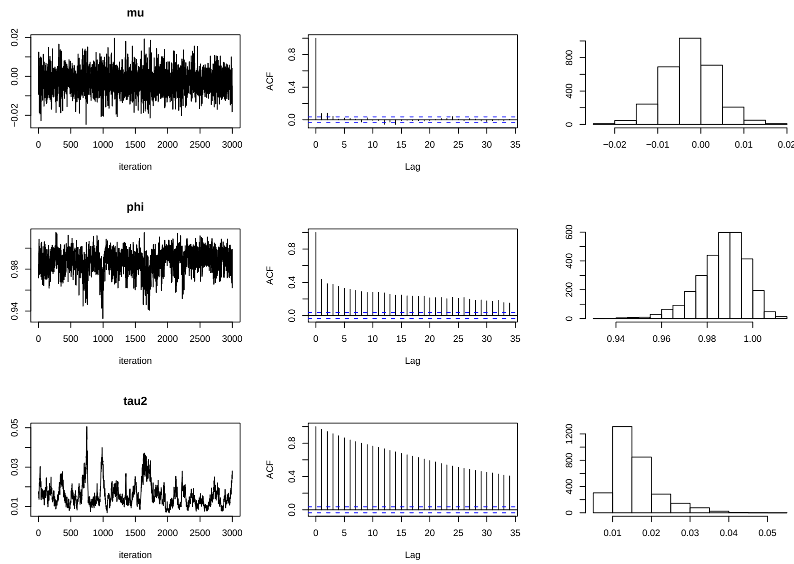
which can be rewritten as a normal DLM:

$$\begin{aligned} \log y_t^2 &= h_t + v_t & v_t \sim N(\mu_{z_t}, \omega_{z_t}^2) \\ h_t &= \mu + \phi h_{t-1} + w_t & w_t \sim N(0, \tau_\eta^2) \end{aligned}$$

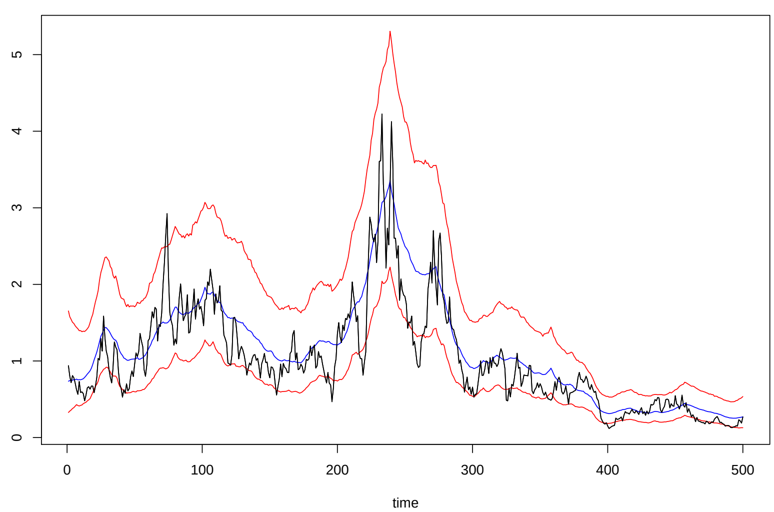
where μ_{z_t} and $\omega_{z_t}^2$ are provided in the previous table.

Then h^n is jointly sampled by using the the FFBS algorithm.

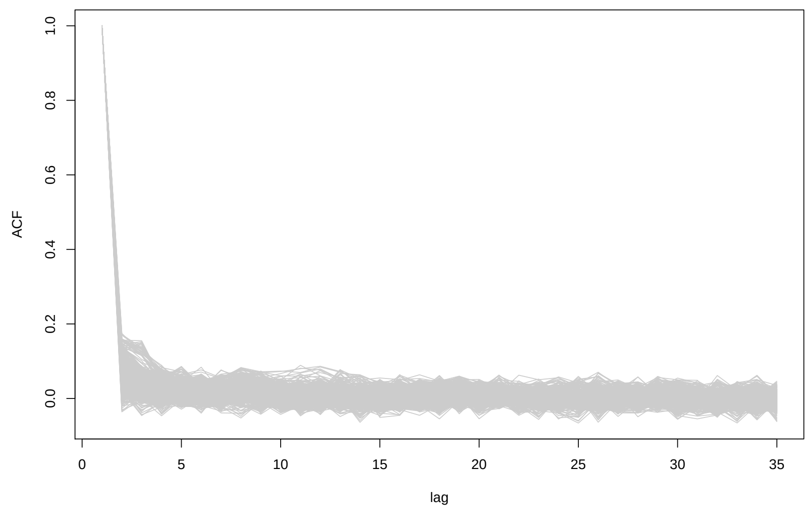
Parameters



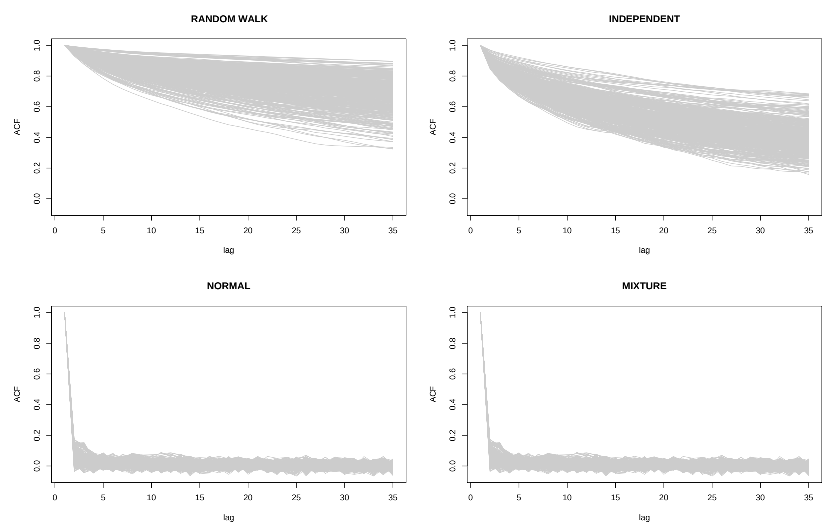
Volatilities



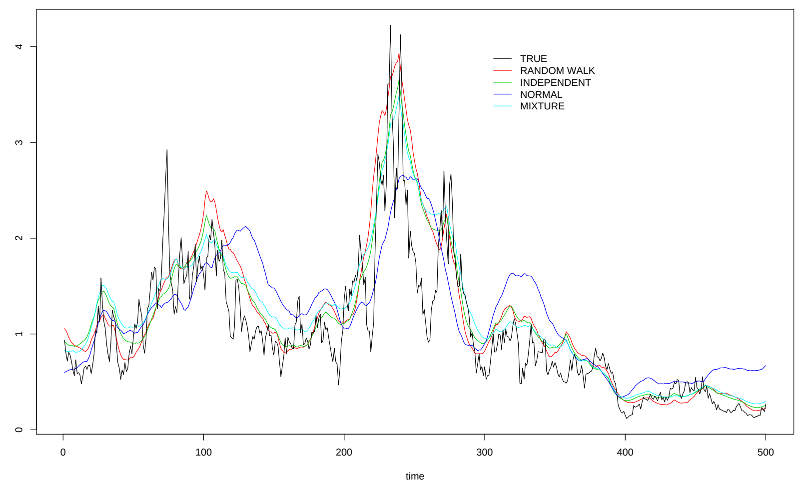
Autocorrelation of h_t



Autocorrelations of h_t for the four schemes



Comparing the four schemes: volatilities


TRUE
RANDOM WALK
INDEPENDENT
NORMAL
MIXTURE
0
1
2
3
4
0
100
200
300
400
500
time

SEQUENTIAL MONTE CARLO METHODS

Nonnormal/nonlinear dynamic models

Most nonnormal and nonlinear dynamic models are defined by

- Observation equation
$$p(y_t|x_t, \psi)$$
- System or evolution equation
$$p(x_t|x_{t-1}, \psi)$$
- Initial distribution
$$p(x_0|\psi)$$

The fixed parameters that drive the state space model, ψ, is kept known and omitted for now.

Evolution and updating

Let the information regarding x_{t-1} at time $t-1$ be summarized by

$$p(x_{t-1}|y^{t-1})$$

Then **Evolution** and **updating** are represented by

$$p(x_t|y^{t-1}) \quad = \quad \int p(x_t|x_{t-1})p(x_{t-1}|y^{t-1})dx_{t-1}$$

and

$$p(x_t|y^t) \quad \propto \quad p(y_t|x_t)p(x_t|y^{t-1})$$

respectively.

These densities are usually **unavailable in closed form**.

The Bayesian boostrap filter

Gordon, Salmond and Smith's (1993) seminal paper uses SIR ideas to obtain draws from $p(x_t|y^t)$ based on draws from $p(x_{t-1}|y^{t-1})$.

SIR: the goal is to draw from $p(x)$ based on draws from $q(x)$.

1. Draw $x_1^*, \ldots, x_N^*$ from q
2. Compute (unnormalized) weights $\omega_i = p(x_i^*)/q(x_i^*)$
3. Draw $x_1, \ldots, x_M$ from $\{x_1^*, \ldots, x_N^*\}$ with weights $\{\omega_1, \ldots, \omega_N\}$

Sampling from the prior: If

$$p(x) \propto \pi(x)l(x)$$

where $\pi(x)$ and $l(x)$ are prior and likelihood, respectively, then a natural (but not necessarily good, actually usually bad!) choice is

$$q(x) = \pi(x).$$

Under this choice, unnormalized weights are likelihoods, i.e.

$$\omega(x) \propto l(x).$$

Example i. Revisiting the 1st order DLM

For illustration, let us reconsider the local level model where closed form solutions are promptly available. The model is

$$\begin{aligned} y_t|x_t &\sim N(x_t, \sigma^2) \\ x_t|x_{t-1} &\sim N(x_{t-1}, \tau^2) \end{aligned}$$

- Posterior at $t = 0$: $(x_0|y_0) \sim N(m_0, C_0)$
- Prior at $t = 1$: $(x_1|y_0) \sim N(m_0, C_0 + \tau^2)$
- Likelihood at time t: $l(x_1; y_1) \propto f_N(x_1; y_1, \sigma^2)$
- Posterior at time t: $(x_1|y_1) \sim N(m_1, C_1)$

where $A_1 = (C_0 + \tau^2)/(C_0 + \tau^2 + \sigma^2)$, $m_1 = (1 - A_1)m_0 + A_1 y_1$ and $C_1 = A_1\sigma^2$.

Example i. One step update

Let $\{(x_0, \omega_0)^{(i)}\}_{i=1}^N$ summarizes $p(x_0|y_0)$. For example,

$$E(g(x_0)|y_0) \approx \frac{1}{N}\sum_{i=1}^{N} \omega_0^{(i)} g(x_0^{(i)}).$$

Then, $\{(x_1, \omega_0)^{(i)}\}_{i=1}^N$ summarizes $p(x_1|y_0)$, where

$$x_1^{(i)} \sim N(x_0^{(i)}, \tau^2) \qquad i = 1, \ldots, N.$$

are draws from the prior $p(x_1|y_0)$.

Then, $\{(x_1, \omega_1)^{(i)}\}_{i=1}^N$ summarizes $p(x_1|y_1)$, where

$$\omega_1^{(i)} = \omega_0^{(i)} f_N(y_1; x_1^{(i)}, \sigma^2) \qquad i = 1, \ldots, N.$$

Example i. Sequential importance sampling (SIS)

Let $\{(x_{t-1}, \omega_{t-1})^{(i)}\}_{i=1}^{N}$ summarizes $p(x_{t-1}|y^{t-1})$.

Then, $\{(x_t, \omega_{t-1})^{(i)}\}_{i=1}^{N}$ summarizes $p(x_t|y^{t-1})$, where

$$\text{Propagation: } x_t^{(i)} \sim N(x_{t-1}^{(i)}, \tau^2) \qquad i = 1, \ldots, N,$$

and $\{(x_t, \omega_t)^{(i)}\}_{i=1}^{N}$ summarizes $p(x^t|y^t)$, where

$$\text{Reweighting: } \omega_t^{(i)} = \omega_{t-1}^{(i)} f_N(y_t; x_t^{(i)}, \sigma^2) \qquad i = 1, \ldots, N.$$

Effective sample size

Liu (1996) proposed using the following measure of degeneracy of an algorithm:

$$N_{\text{eff},t} = \frac{1}{\sum_{i=1}^{N} \left(w_t^{(i)}\right)^2}$$

where w_ts are normalized weights, i.e. $w_t^{(i)} = \omega_t^{(i)} / \sum_{j=1}^{N} \omega_t^{(j)}$.

If $w_t^{(i)} = 1/N$ (equally balanced weights), then

$$N_{\text{eff},t} = N.$$

If $w_t^{(j)} = 1$ for only one j (particle degeneracy) then

$$N_{\text{eff},t} = 1.$$

Example i. SIS with resampling (SISR)

SIS:

- $\{(x_{t-1}, \omega_{t-1})^{(i)}\}_{i=1}^{N}$ summarizes $p(x_{t-1}|y^{t-1})$.
- $\{(\tilde{x}_t, \omega_{t-1})^{(i)}\}_{i=1}^{N}$ summarizes $p(x_t|y^{t-1})$, where $\tilde{x}_t^{(i)} \sim N(x_{t-1}^{(i)}, \tau^2)$, for $i = 1, \ldots, N$.
- $\{(\tilde{x}_t, \tilde{\omega}_t)^{(i)}\}_{i=1}^{N}$ summarizes $p(x^t|y^t)$, where $\tilde{\omega}_t^{(i)} = \omega_{t-1}^{(i)} f_N(y_t; \tilde{x}_t^{(i)}, \sigma^2)$, for $i = 1, \ldots, N$.

Resampling:

Draw $x_t^{(1)}, \ldots, x_t^{(N)}$ from the set $\left\{\tilde{x}_t^{(1)}, \ldots, \tilde{x}_t^{(N)}\right\}$ with weights $\left\{\tilde{\omega}_t^{(1)}, \ldots, \tilde{\omega}_t^{(N)}\right\}$.

Therefore, $\{(x_t, \omega_t)^{(i)}\}_{i=1}^{N}$ summarizes $p(x_t|y^t)$, where $\omega_t = 1/N$.

SIS with Resampling (SISR)

$$\{x_t^1,\ldots,x_t^N\} \sim p(x_t|y^t)$$

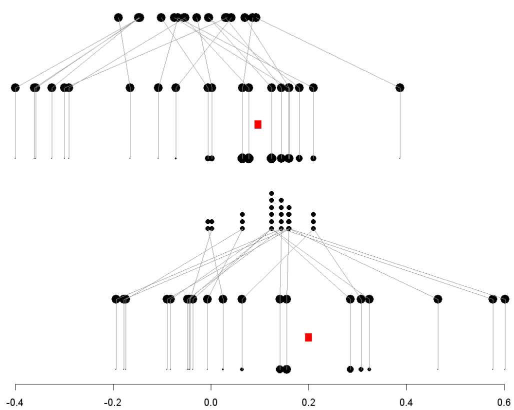
$$\tilde{x}_{t+1}^i \sim p(x_{t+1}|x_t^i)$$

$$\omega_{t+1}^i \propto p(y_{t+1}|\tilde{x}_{t+1}^i)$$

$$\{x_{t+1}^1,\ldots,x_{t+1}^N\} \sim p(x_{t+1}|y^{t+1})$$

$$\tilde{x}_{t+2}^i \sim p(x_{t+2}|x_{t+1}^i)$$

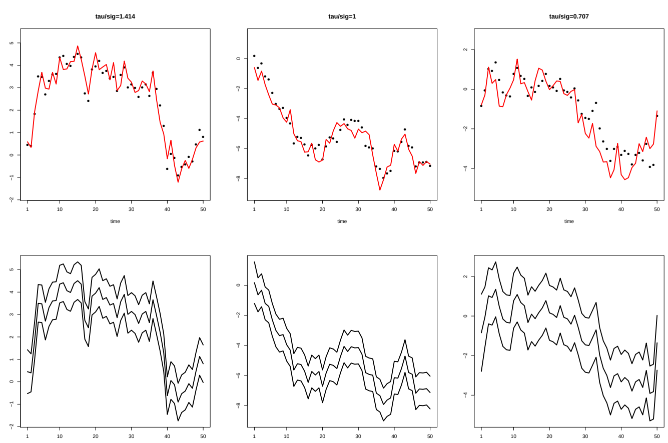
$$\omega_{t+2}^i \propto p(y_{t+2}|\tilde{x}_{t+2}^i)$$



Uniform weights is the goal!

Example i. Simulated data

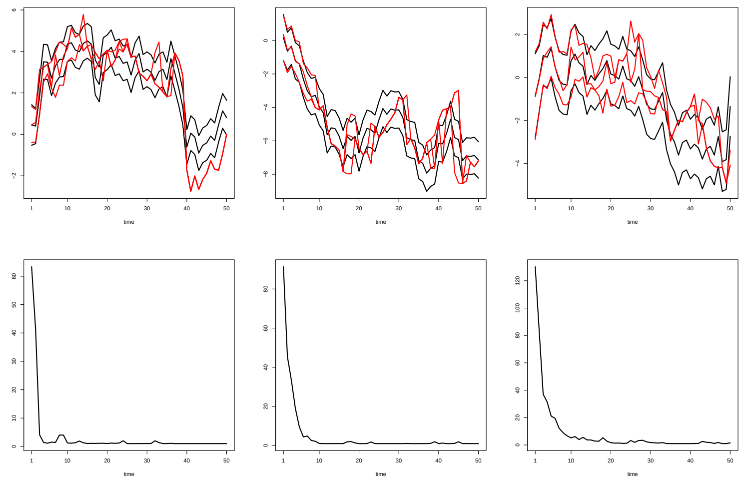
$n = 50$, $x_0 = 0$, $\tau^2 = 0.5$ and $\sigma^2 = (0.25, 0.5, 1.0)$.



Top: y_t and x_t; bottom: m_t and $m_t \pm 2\sqrt{C_t}$.

Left: $\tau/\sigma = 1.414$; center: $\tau/\sigma = 1.000$; right: $\tau/\sigma = 0.707$.

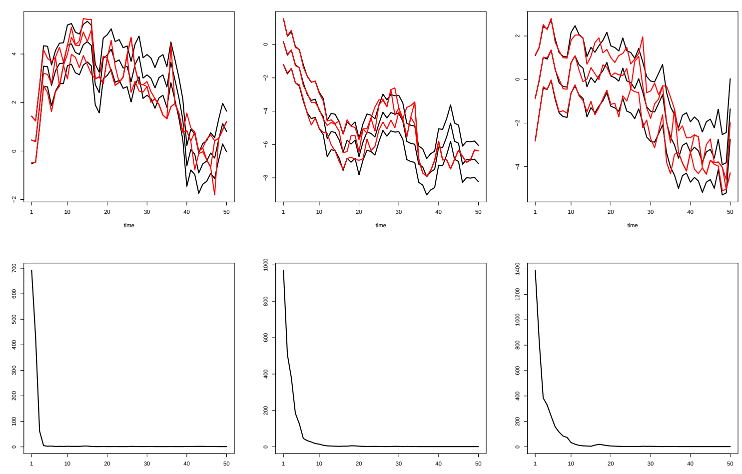
Example i. SIS, $N = 1,000$



Top: States; Bottom: N_{eff}.

Left: $\tau/\sigma = 1.414$; center: $\tau/\sigma = 1.000$; right: $\tau/\sigma = 0.707$.

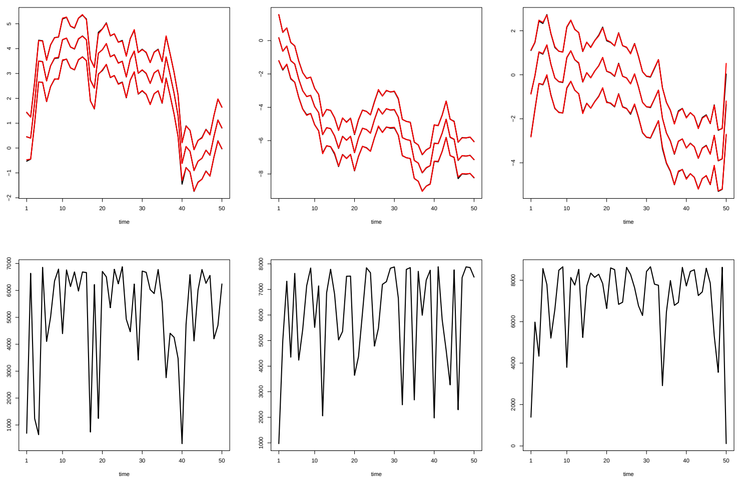
Example i. SIS, $N = 10,000$



Top: States; Bottom: N_{eff}.

Left: $\tau/\sigma = 1.414$; center: $\tau/\sigma = 1.000$; right: $\tau/\sigma = 0.707$.

Example i. SISR, $N = 10,000$



Top: States; Bottom: N_{eff}.

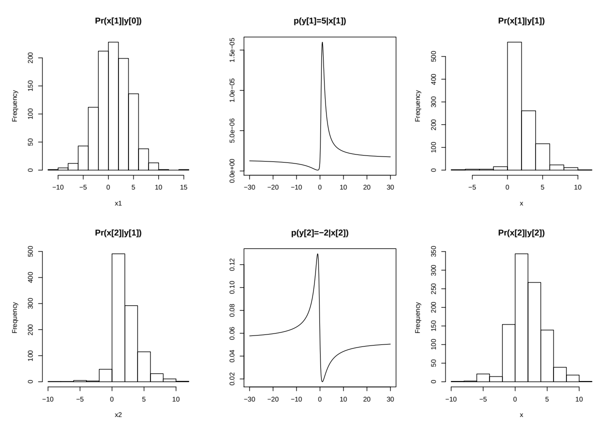
Left: $\tau/\sigma = 1.414$; center: $\tau/\sigma = 1.000$; right: $\tau/\sigma = 0.707$.

Example ii. Boostrap filter step by step

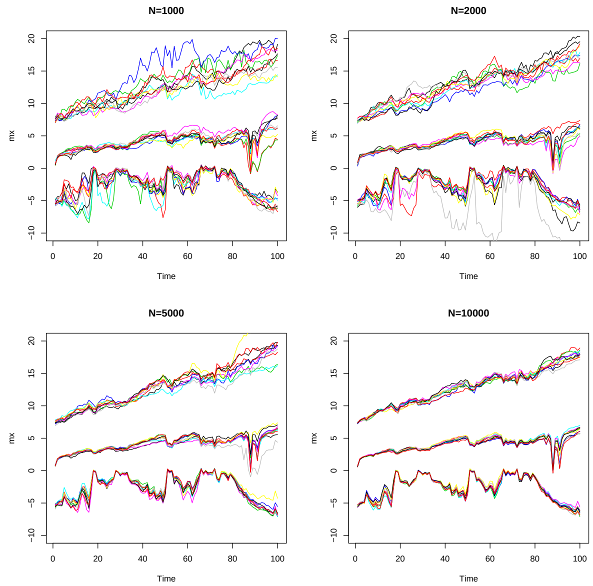
Model

$$
\begin{aligned}
y_t &= x_t/(1+x_t^2) + v_t \qquad v_t \sim N(0,1) \\
x_t &= x_{t-1} + w_t \qquad w_t \sim N(0,0.5)
\end{aligned}
$$

and $x_0 \sim N(1,10)$.



Example ii. Boostrap filter Monte Carlo error



Auxiliary particle filter (APF)

Recall the two main steps in any dynamic model:

$$
\begin{aligned}
p(x_t|y^{t-1}) &= \int p(x_t|x_{t-1})p(x_{t-1}|y^{t-1})dx_{t-1} \\
p(x_t|y^t) &\propto p(y_t|x_t)p(x_t|y^{t-1})
\end{aligned}
$$

- $\left\{(x_{t-1}, \omega_{t-1})^{(i)}\right\}_{i=1}^N$ summarizes $p(x_{t-1}|y^{t-1})$.
- Approximating $p(x_t|y^{-1})$ by

$$p_N(x_t|y^{t-1}) = \sum_{i=1}^N p(x_t|x_{t-1}^{(i)})\omega_{t-1}^{(i)}$$

- Approximating $p(x_t|y^t)$ by

$$p_N(x_t|y^t) = \sum_{i=1}^N p(y_t|x_t)p(x_t|x_{t-1}^{(i)})\omega_{t-1}^{(i)}$$

Pitt and Shephard's (1999) idea

The previous mixture approximation suggests an augmentation scheme where the new target distribution is

$$p_N(x_t, k|y^t) = p(y_t|x_t)p(x_t|x_{t-1}^{(k)})\omega_{t-1}^{(k)}.$$

A natural proposal distribution is

$$q(x_t, k|y^t) = p(y_t|g(x_{t-1}^{(k)}))p(x_t|x_{t-1}^{(k)})\omega_{t-1}^{(k)}$$

where, for instance, $g(x_{t-1}) = E(x_t|x_{t-1})$.

By a simple SIR argument, the weight of the particle x_t is

$$\omega_t \propto \frac{p(y_t|x_t)}{p(y_t|g(x_{t-1}^{(k)}))}$$

APF algorithm

- $\{(x_{t-1}, \omega_{t-1})^{(i)}\}_{i=1}^N$ summarizes $p(x_{t-1}|y^{t-1})$.
- For $j = 1, \ldots, N$
  - Draw k^j from $\{1, \ldots, N\}$ with weights $\{\tilde{\omega}_{t-1}^{(1)}, \ldots, \tilde{\omega}_{t-1}^{(N)}\}$:
  $$\tilde{\omega}_{t-1}^{(i)} = \omega_{t-1}^{(i)} p(y_t|g(x_{t-1}^{(i)}))$$
  - Draw $x_t^{(j)}$ from $p(x_t|x_{t-1}^{(k^j)})$.
  - Compute associated weight
  $$\omega_t^{(j)} \propto \frac{p(y_t|x_t^{(j)})}{p(y_t|g(x_{t-1}^{(k^j)}))}.$$
- $\{(x_t, \omega_t)^{(i)}\}_{i=1}^N$ summarizes $p(x_t|y^t)$.
- Maybe add a SIR step to replenish x_ts.

Smoothing

Godsill, Doucet and West (2004) proposed a smoothing scheme based on particle filter draws.

The key results are

$$p(x^n|y^n) = p(x_n|y^n) \prod_{t=1}^{n-1} p(x_t|x_{t+1}, y^t)$$

and (by Bayes rule and conditional independence)

$$p(x_t|x_{t+1}, y^t) \propto p(x_{t+1}|x_t, y^t) p(x_t|y^t).$$

We can now jointly sample from $p(x^n|y^n)$ by sequentially sampling from filtered particles with weights proportional to $p(x_{t+1}|x_t, y^t)$.

Backward sampling algorithm

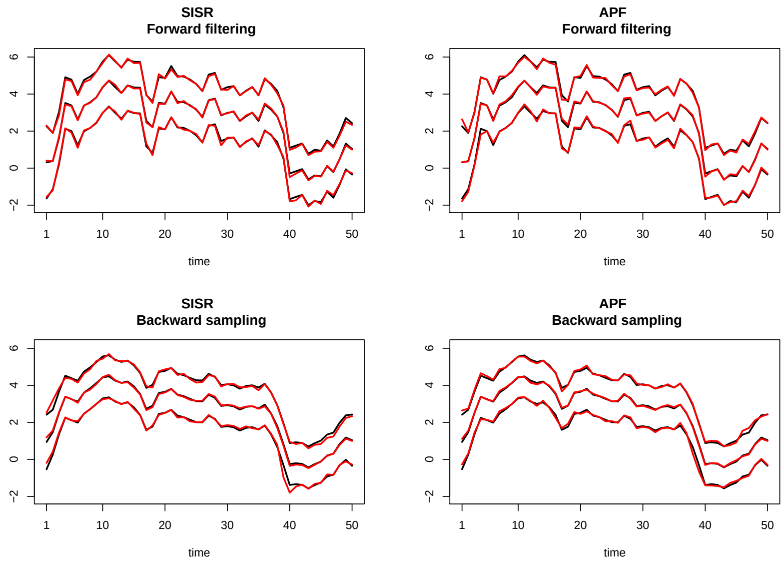
Repeat the following three steps N times.

- Sample $\tilde{x}_n$ from $\{x_n^{(i)}\}_{i=1}^N$ with weights $\{\omega_n^{(i)}\}_{i=1}^N$.
- For $t = n-1, \ldots, 1$

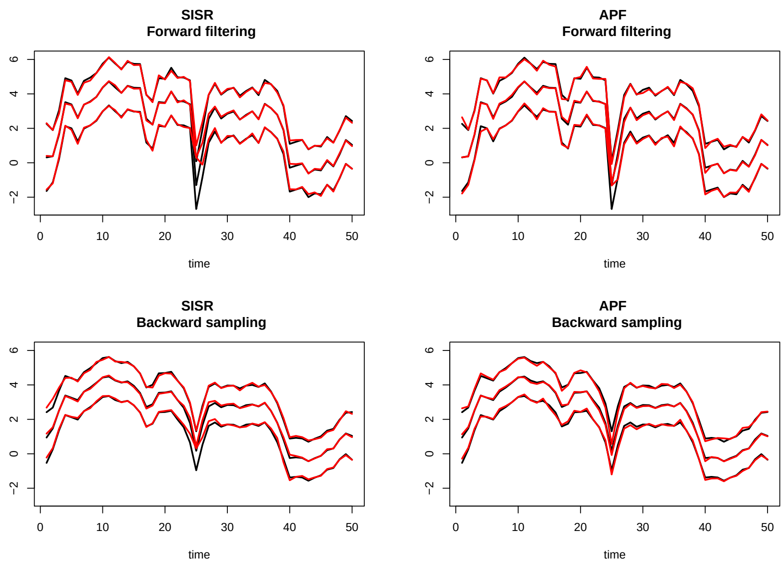
  Sample $\tilde{x}_t$ from $\{x_t^{(i)}\}_{i=1}^N$ with weights $\{\tilde{\omega}_t^{(i)}\}_{i=1}^N$
  $$\tilde{\omega}_t^{(i)} \propto \omega_t^{(i)} p\left(\tilde{x}_{t+1}|x_t^{(i)}\right) \qquad i = 1, \ldots, N$$
- Then $\{\tilde{x}_1^{(j)}, \ldots, \tilde{x}_n^{(j)}\}$ is a draw from $p(x^n|y^n)$.

Example i. smoothing

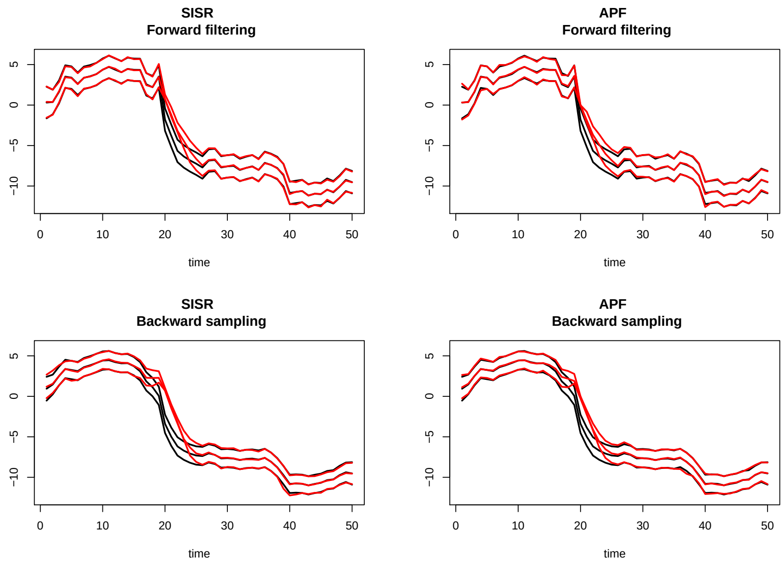
$n = 50$, $\tau^2 = 0.5$, $\sigma^2 = 1$, $x_0 = 0$, $m_0 = 0$, $C_0 = 100$, $N = 1000$.



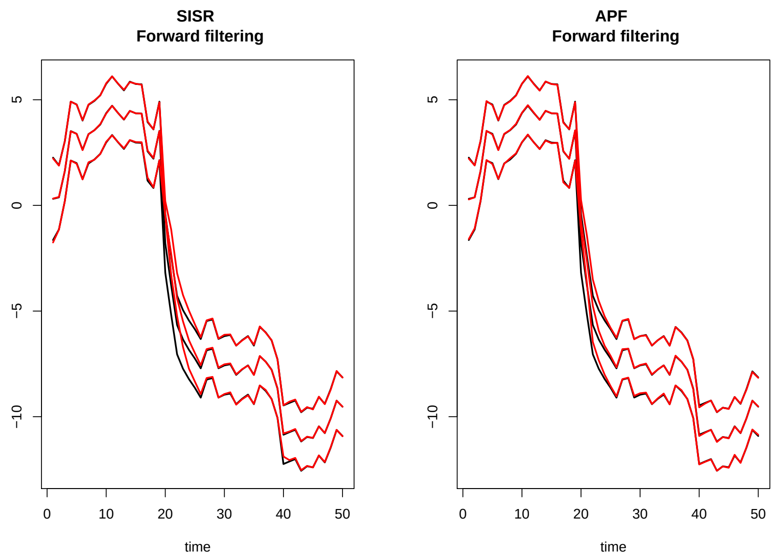
Example i. outlier in y_t



Example i. outlier in x_t



Example i. outlier in x_t (more particles)



SEQUENTIAL MONTE CARLO WITH PARAMETER LEARNING

Revisiting the bootstrap and the AP filters

Consider the following general state space model

$$\begin{aligned} \text{Observation equation} &: \quad p(y_{t+1}|x_{t+1}) \\ \text{State equation} &: \quad p(x_{t+1}|x_t) \end{aligned}$$

For a given time t

$$\{(x_t, \omega_t)^{(i)}\}_{i=1}^N$$

is a particle representation of

$$p(x_t|y^t)$$

where $y^t = (y_1, \dots, y_n)$.

Sample-resample

Goal: $\{(x_{t+1}, \omega_{t+1})^{(i)}\}_{i=1}^N \sim p(x_{t+1}|y^{t+1})$.

Algorithm

- Sample $x_{t+1}^{(i)}$ from $q(x_{t+1}|x_t^{(i)}, y_{t+1})$
- Compute weights

$$\omega_{t+1}^{(i)} = \omega_t^{(i)} \frac{p(y_{t+1}|x_{t+1}^{(i)})p(x_{t+1}^{(i)}|x_t^{(i)})}{q(x_{t+1}^{(i)}|x_t^{(i)}, y_{t+1})}$$

Special case: bootstrap filter

In the bootstrap filter

$$q(x_{t+1}|x_t, y_{t+1}) = p(x_{t+1}|x_t),$$

i.e. the transition equation.

This proposal density has no information about y_{t+1}, so we say that the scheme is *blinded*.

The weights are then proportional to the likelihoods

$$\omega_{t+1}^{(i)} = \omega_t^{(i)} p(y_{t+1}|x_{t+1}^{(i)}).$$

Special case: optimal filter

In the optimal filter

$$q(x_{t+1}|x_t, y_{t+1}) = p(x_{t+1}|x_t, y_{t+1}).$$

The weights are then

$$\omega_{t+1}^{(i)} = \omega_t^{(i)} p(y_{t+1}|y^t) \propto \omega_t^{(i)}$$

so, if $\omega_0 \propto 1$, then $\omega_{t+1} \propto 1$ for all t.

This is a **perfectly adapted** filter.

Resample-sample

Goal: $\{(x_{t+1}, \omega_{t+1})^{(i)}\}_{i=1}^{N} \sim p(x_{t+1}|y^{t+1})$.

Algorithm

- Resample $\tilde{x}_t^{(i)}$ from $\{x_t^{(1)}, \ldots, x_t^{(N)}\}$ with weights
$$q_1(x_t^{(j)}|y_{t+1}) \qquad j = 1, \ldots, N$$
- Sample $x_{t+1}^{(i)}$ from $q_2(x_{t+1}|\tilde{x}_t^{(i)}, y_{t+1})$
- Compute weights
$$\omega_{t+1}^{(i)} = \omega_t^{(i)} \frac{p(y_{t+1}|x_{t+1}^{(i)})p(x_{t+1}^{(i)}|\tilde{x}_t^{(i)})}{q_1(\tilde{x}_t^{(i)}|y_{t+1})q_2(x_{t+1}^{(i)}|\tilde{x}_t^{(i)}, y_{t+1})}$$

Special case: auxiliary particle filter

In the auxiliary particle filter
$$q_1(x_t|y_{t+1}) = p(y_{t+1}|g(x_t))$$
where, for instance, $g(x_t) = E(x_{t+1}|x_t)$.

Also,
$$q_2(x_{t+1}|x_t, y_{t+1}) = p(x_{t+1}|x_t)$$
i.e. the transition equation, so again a *blinded* proposal.

The weights are then equal to
$$\omega_{t+1}^{(i)} = \omega_t^{(i)} \frac{p(y_{t+1}|x_{t+1}^{(i)})}{p(y_{t+1}|g(\tilde{x}_t^{(i)}))}.$$

Special case: optimal filter

In the optimal filter both proposals q_1 and q_2 depend on y_{t+1}, i.e.
$$q_1(x_t|y_{t+1}) = p(y_{t+1}|x_t).$$
and
$$q_2(x_{t+1}|x_t, y_{t+1}) = p(x_{t+1}|x_t, y_{t+1}).$$

The weights are then equal to
$$\omega_{t+1}^{(i)} = \omega_t^{(i)}$$
so, if $\omega_0 \propto 1$, then $\omega_{t+1} \propto 1$ for all t.

This is a **perfectly adapted** filter.

Resample-sample with learning θ

The objective is to combine $\{(x_t, \theta_t, \omega_t)^{(i)}\}_{i=1}^N \sim p(x_t, \theta|y^t)$ with y_{t+1} to produce $\{(x_{t+1}, \theta_{t+1}, \omega_{t+1})^{(i)}\}_{i=1}^N \sim p(x_{t+1}, \theta|y^{t+1})$.

The index t and $t+1$ in $\theta^{(i)}$ are used to facilitate the identification of the time at which draws are being used.

Algorithm

- Resample $(\tilde{x}_t, \tilde{\theta}_t)^{(i)}$ from $\{(x_t, \theta_t)^{(j)}\}_{j=1}^N$ with weights

$$q_1((x_t, \theta_t)^{(j)}|y_{t+1}) \qquad j = 1, \ldots, N.$$

- Sample $(x_{t+1}, \theta_{t+1})^{(i)}$ from $q_2(x_{t+1}, \theta|(\tilde{x}_t, \tilde{\theta}_t)^{(i)}, y_{t+1})$.

- Compute weights

$$\omega_{t+1}^{(i)} = \omega_t^{(i)} \frac{p(y_{t+1}|(x_{t+1}, \theta_{t+1})^{(i)})}{q_1((\tilde{x}_t, \tilde{\theta}_t)^{(i)})|y_{t+1})} \frac{p((x_{t+1}, \theta_{t+1})^{(i)}|(\tilde{x}_t, \tilde{\theta}_t)^{(i)})}{q_2((x_{t+1}, \theta_{t+1})^{(i)}|(\tilde{x}_t, \tilde{\theta}_t)^{(i)}, y_{t+1})}$$

Questions:

- How to choose q_1 and q_2?

- What is $p(x_{t+1}, \theta_{t+1}|x_t, \theta_t)$?

- Is it okay to decompose it as

$$p(x_{t+1}, \theta_{t+1}|x_t, \theta_t) = p(x_{t+1}|\theta_t, x_t)p(\theta_{t+1}|x_t, \theta_t)?$$

- If so, then what is $p(\theta_{t+1}|x_t, \theta_t)$?

Liu and West (2001)

They approximate $p(\theta|y^t)$ by a N-component mixture of multivariate normal distributions, i.e.

$$p(\theta|y^t) = \sum_{i=1}^N \omega_t^{(i)} f_N(\theta|a\theta_t^{(i)} + (1-a)\bar{\theta}_t, (1-a^2)V_t)$$

where $\bar{\theta}_t = \sum_{i=1}^N \omega_t^{(i)} \theta_t^{(i)}$ and $V_t = \sum_{t=1}^N \omega_t^{(i)} (\theta_t^{(i)} - \bar{\theta}_t)(\theta_t^{(i)} - \bar{\theta}_t)'$.

This leads to

$$p(\theta_{t+1}|x_t^{(i)}, \theta_t^{(i)}) = f_N(\theta_{t+1}|a\theta_t^{(i)} + (1-a)\bar{\theta}_t, (1-a^2)V_t)$$

They use the same decomposition for q_2. So the weights are

$$\omega_{t+1}^{(i)} = \omega_t^{(i)} \frac{p(y_{t+1}|(x_{t+1}, \theta_{t+1})^{(i)})}{q_1((\tilde{x}_t, \tilde{\theta}_t)^{(i)})|y_{t+1})}$$

Resampling step

$$q_1(x_t, \theta_t | y_{t+1}) = p(y_{t+1} | g(x_t), m(\theta_t))$$

where

$$g(x_t) = E(x_{t+1} | x_t, m(\theta_t))$$

for instance, and

$$m(\theta_t) = a\theta_t + (1-a)\bar{\theta}_t$$

The weights are then

$$\omega_{t+1}^{(i)} = \omega_t^{(i)} \frac{p(y_{t+1} | x_{t+1}^{(i)}, \theta_{t+1}^{(i)})}{p(y_{t+1} | g(\tilde{x}_t^{(i)}), m(\tilde{\theta}_t^{(i)}))}$$

Choosing a

Liu and West (2001) use a discount factor argument (see West and Harrison, 1997) to set the parameter a:

$$a = \frac{3\delta - 1}{2\delta}$$

For example,

- $\delta = 0.50$ leads to $a = 0.500$
- $\delta = 0.75$ leads to $a = 0.833$
- $\delta = 0.95$ leads to $a = 0.974$
- $\delta = 1.00$ leads to $a = 1.000$.

In the last case, i.e. $a = 1.0$, the particles of θ will degenerate over time to a single particle.

The LW filter in one page

For particles $\{(x_t, \theta_t, \omega_t)^{(j)}\}_{j=1}^N$ summarizing $p(x_t, \theta | y^t)$, estimates $\bar{\theta}_t = \sum_{i=1}^N \omega_t^{(i)} \theta_t^{(i)}$ and $V_t = \sum_{i=1}^N \omega_t^{(i)} (\theta_t^{(i)} - \bar{\theta}_t)(\theta_t^{(i)} - \bar{\theta}_t)'$, and given shrinkage parameter a, the algorithm runs as follows.

- For $i = 1, \ldots, N$, compute
  - $m(\theta_t^{(i)}) = a\theta_t^{(i)} + (1-a)\bar{\theta}_t$.
  - $g(x_t^{(i)}) = E(x_{t+1} | x_t^{(i)}, m(\theta_t^{(i)}))$.
  - $w_{t+1}^{(i)} = p(y_{t+1} | g(x_t^{(j)}), m(\theta_t^{(j)}))$.
- For $i = 1, \ldots, N$
  - Resample $(\tilde{x}_t, \tilde{\theta}_t)^{(i)}$ from $\{(x_t, \theta_t, w_{t+1})^{(j)}\}_{j=1}^N$.
  - Sample $\theta_{t+1}^{(i)} \sim N(m(\tilde{\theta}_t^{(i)}), (1-a^2)V_t)$.

– Sample $x_{t+1}^{(i)}$ from $p(x_{t+1}|\tilde{x}_t^{(i)},\theta_{t+1}^{(i)})$.
– Compute weight

$$\omega_{t+1}^{(i)} = \omega_t^{(i)} \frac{p(y_{t+1}|x_{t+1}^{(i)},\theta_{t+1}^{(i)})}{p(y_{t+1}|g(\tilde{x}_t^{(i)}),m(\tilde{\theta}_t^{(i)}))}.$$

Example i. first order dynamic linear model

The revisit the first order dynamic linear model

$$\begin{aligned} y_t &= x_t + \nu_t \qquad \nu_t \sim N(0,\sigma^2) \\ x_t &= x_{t-1} + \omega_t \qquad \omega_t \sim N(0,\tau^2) \end{aligned}$$

where $x_0 = 25$, $\sigma^2 = 0.1$, $\tau^2 = (0.2, 0.1, 0.05)$ and $n = 200$.

Prior setup:

$$\begin{aligned} \sigma^2 &\sim IG(a_0, b_0) \\ x_0 &\sim N(m_0, C_0) \end{aligned}$$

where $a_0 = 5$, $b_0 = 0.4$, $m_0 = 25$ and $C_0 = 100$.

Particle filter setup:

$$\begin{aligned} N &= 2000 \\ \delta &= (0.75, 0.95) \end{aligned}$$

Example i. LW + optimal propagation

Liu and West's (2001) filter with optimal resampling proposal, i.e.

$$p(x_{t+1}|x_t,\sigma^2,y_{t+1}) = f_N(x_{t+1}|m_{t+1},C_{t+1})$$

where

$$\begin{aligned} C_{t+1}^{-1} &= \tau^{-2} + \sigma^{-2} \\ m_{t+1} &= C_{t+1}(\sigma^{-2}y_{t+1} + \tau^{-2}x_t) \end{aligned}$$

Example i. LW + optimal propagation + kernel for σ^2

Optimal propagation + with mixture approximating σ^2 directly, i.e.

$$q(\sigma^2|x_t,\sigma_t^2,y_{t+1}) \propto f_N(y_{t+1};x_t,\sigma^2)f_{IG}(\sigma^2|\alpha(\sigma_t^2),\beta(\sigma_t^2))$$

where

$$\begin{aligned} \alpha(\sigma_t^2) &= \frac{\{m(\sigma_t^2)\}^2}{v(\sigma_t^2)} + 2 \\ \beta(\sigma_t^2) &= m(\sigma_t^2)\alpha(\sigma_t^2) \end{aligned}$$

and

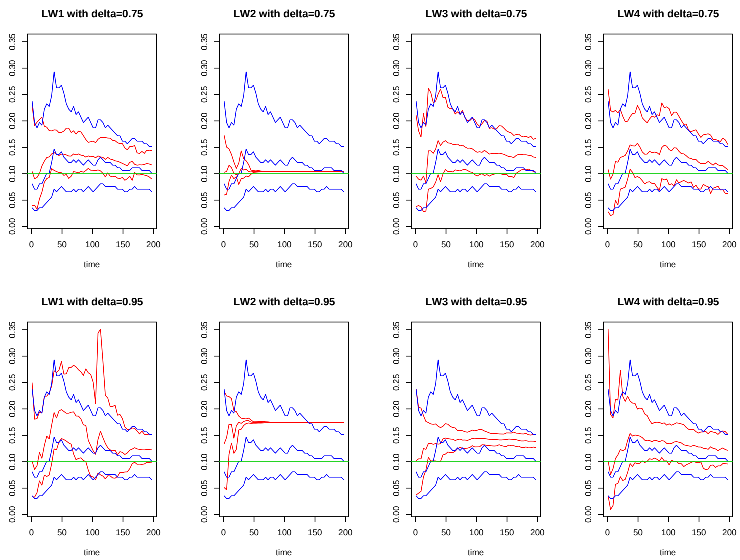
$$\begin{aligned} m(\sigma_t^2) &= a\sigma_t^2 + (1-a)\bar{\sigma}^2 \\ v(\sigma_t^2) &= (1-a^2)S_{\sigma^2}^2 \end{aligned}$$

with $\bar{\sigma}^2$ and $S_{\sigma^2}^2$ the particle approximation to the mean and variance of σ^2 from $p(\sigma^2|y^t)$.

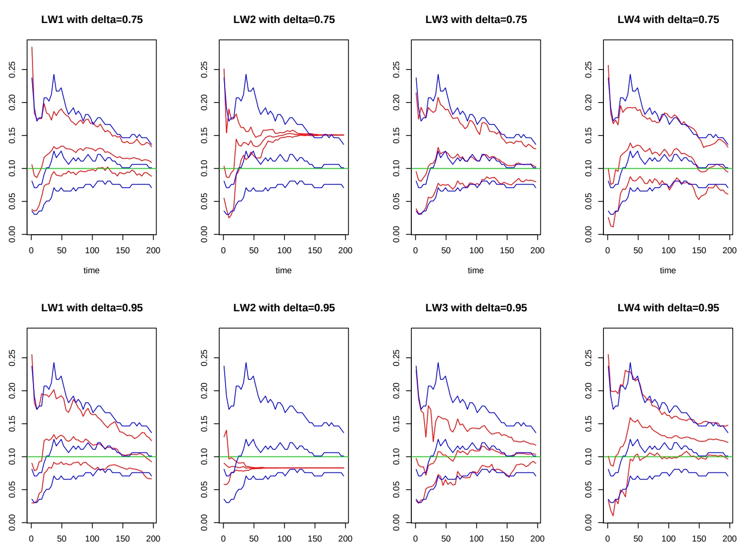
Example i. Comparing various LW filters

- LW1 : LW + $\log \sigma^2$
- LW2 : LW + σ^2
- LW3 : LW + $\log \sigma^2$ + optimal propagation
- LW4 : LW + σ^2 + optimal propagation

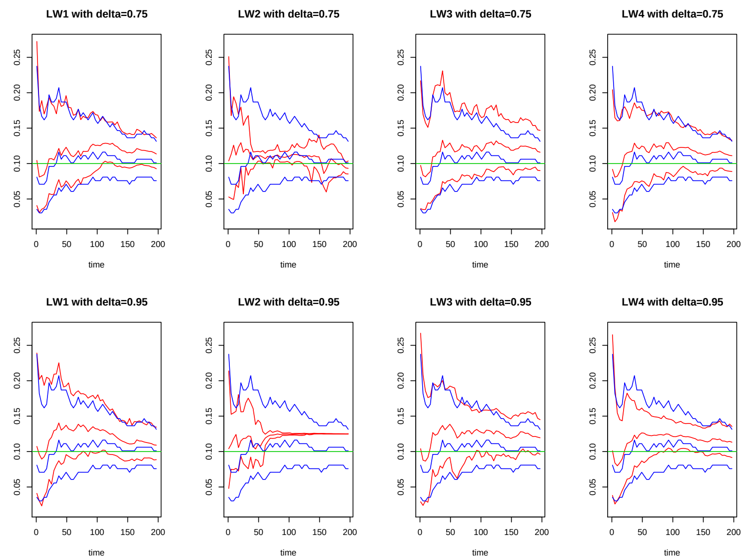
Example i. $\tau/\sigma = 1.4$



Example i. $\tau/\sigma = 1.0$



Example i. $\tau/\sigma = 0.7$



Example ii. nonlinear dynamic model

Let y_t, for $t = 1, \ldots, n$, be modeled as

$$
\begin{aligned}
(y_t|x_t, \psi) &\sim N(x_t^2/20, \sigma^2) \\
(x_t|x_{t-1}, \psi) &\sim N(G'_{x_{t-1}}\xi, \tau^2)
\end{aligned}
$$

where $G'_{x_t} = \big(x_t, x_t/(1+x_t^2), \cos(1.2t)\big)$, $\psi = (\xi', \sigma^2, \tau^2)$ and $\xi = (\alpha, \beta, \gamma)'$.

Prior distributions for θ_0, ξ, σ^2 and τ^2 are

$$
\begin{aligned}
x_0 &\sim N(m_0, V_0) \\
\xi &\sim N(c_0, C_0) \\
\sigma^2 &\sim IG(a_0, A_0) \\
\tau^2 &\sim IG(b_0, B_0)
\end{aligned}
$$

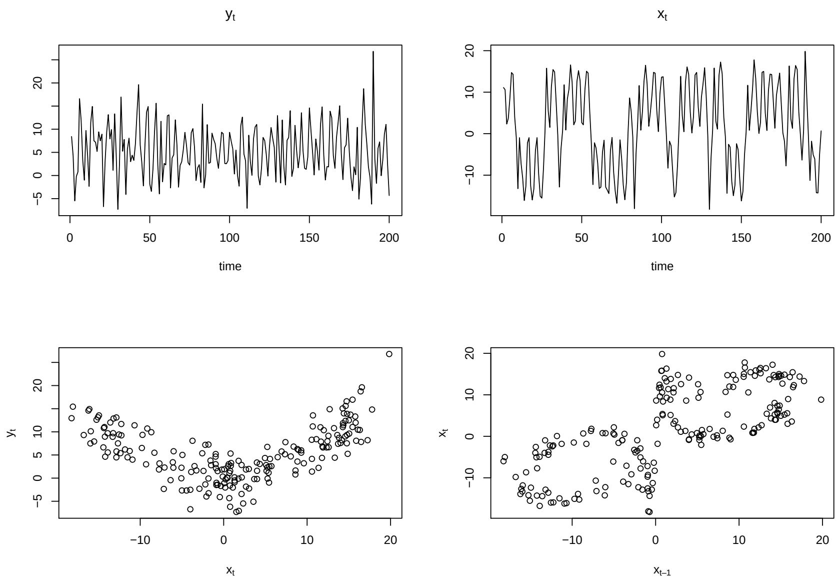
Example ii. Simulation set up

We simulated $n = 200$ observations based on $\xi = (0.5, 25, 8)'$, $\sigma^2 = 10$, $\tau^2 = 1$ and $x_0 = 0.1$.

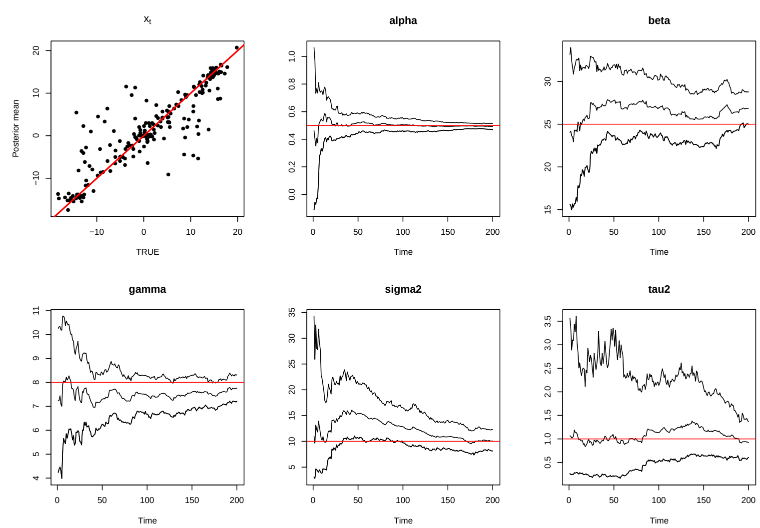
Prior hyperparameters:

$$
\begin{aligned}
m_0 &= 0.0 \quad \text{and} \quad V_0 = 5 \\
c_0 &= (0.5, 25, 8)' \quad \text{and} \quad C_0 = \text{diag}(0.1, 16, 2) \\
a_0 &= 3 \quad \text{and} \quad A_0 = 20 \\
b_0 &= 3 \quad \text{and} \quad B_0 = 2
\end{aligned}
$$

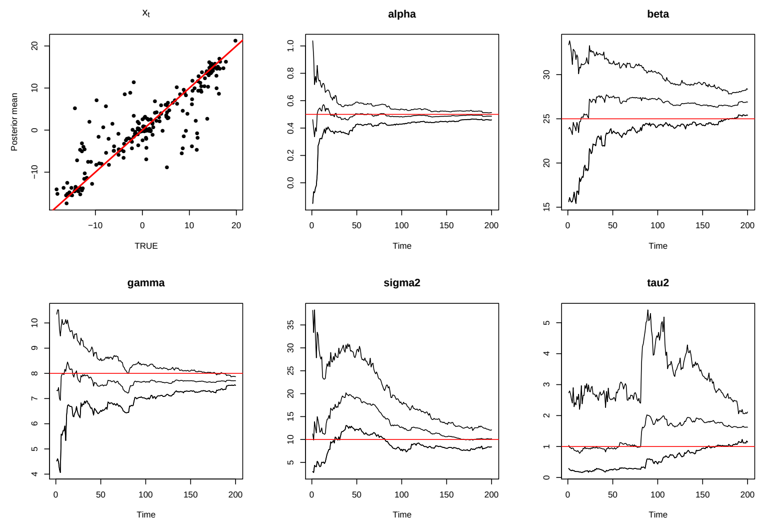
Example ii. Simulated data



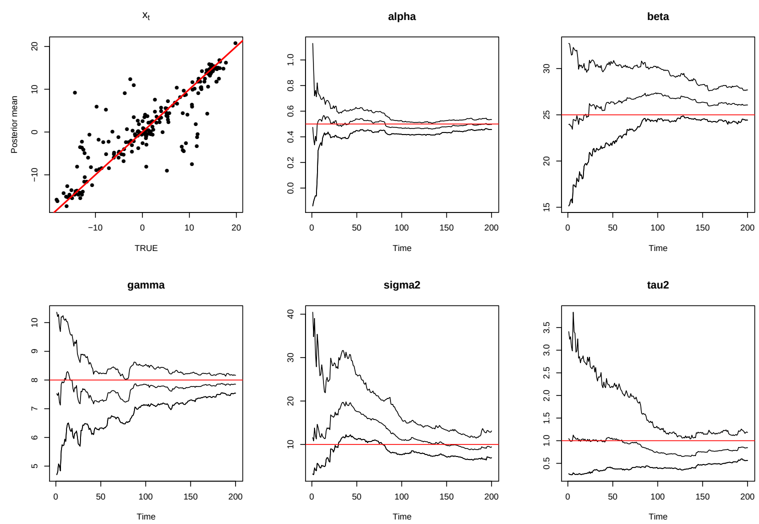
Example ii. $(N, \delta, a) = (2000, 0.75, 0.83)$



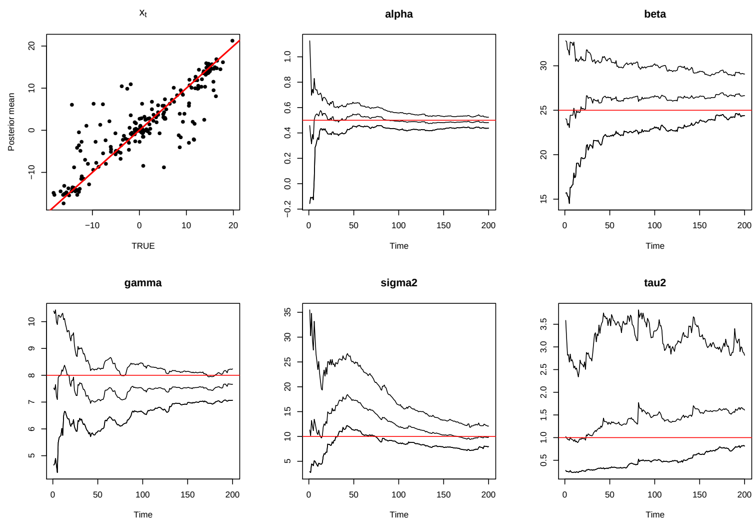
Example ii. $(N, \delta, a) = (2000, 0.90, 0.94)$



Example ii. $(N, \delta, a) = (5000, 0.90, 0.94)$

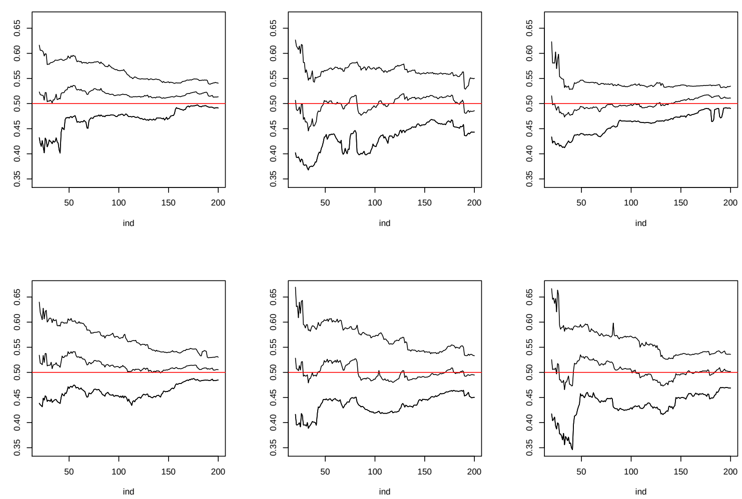


Example ii. $(N, \delta, a) = (10000, 0.90, 0.94)$

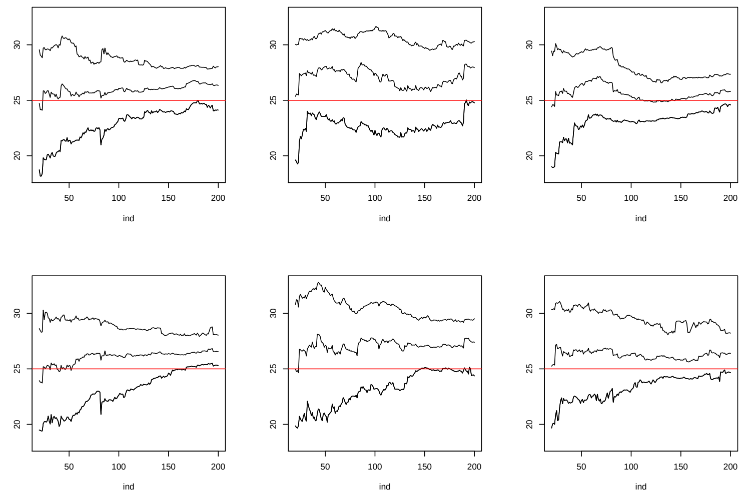


Example ii. Assessing MC error - α

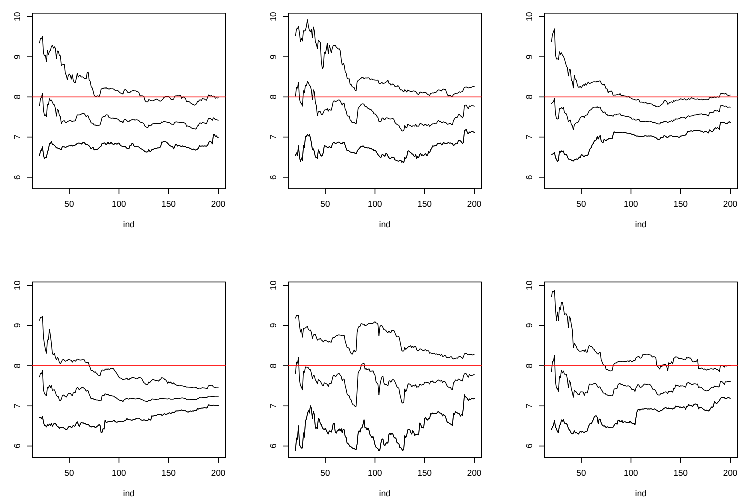
$N = 5000$



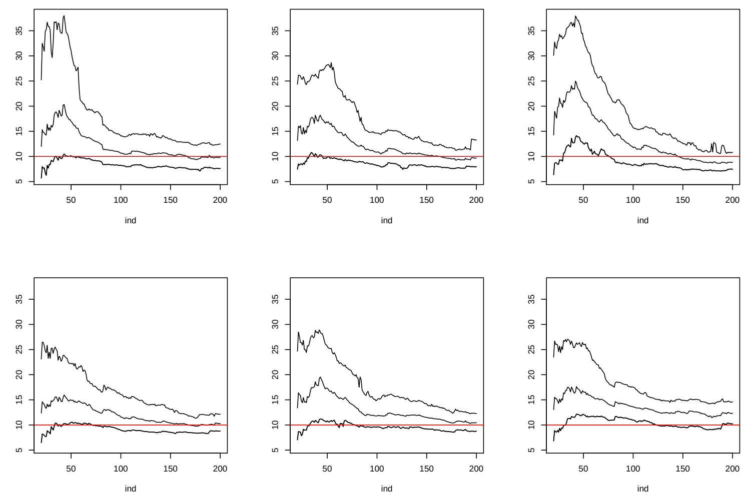
Example ii. Assessing MC error - β



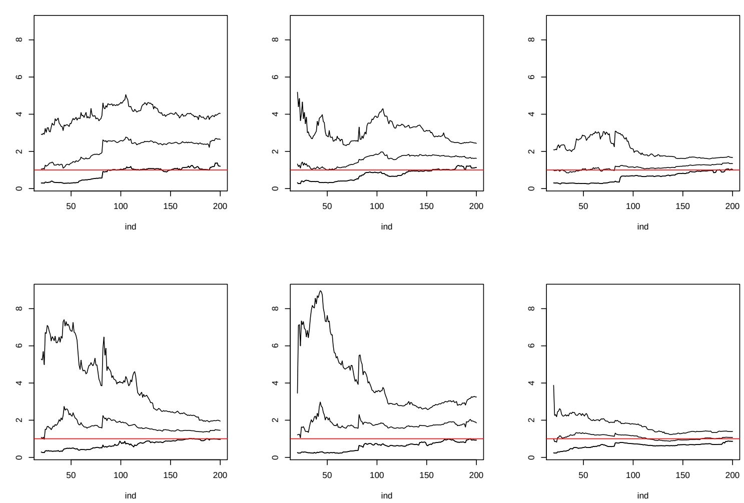
Example ii. Assessing MC error - γ



Example ii. Assessing MC error - σ^2

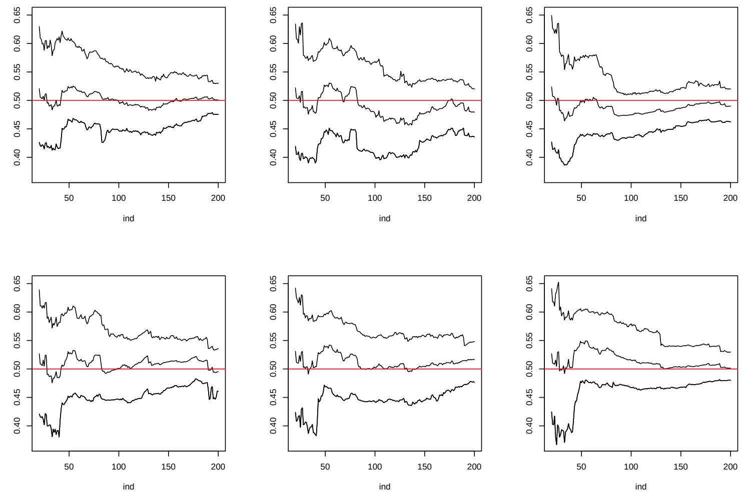


Example ii. Assessing MC error - τ^2

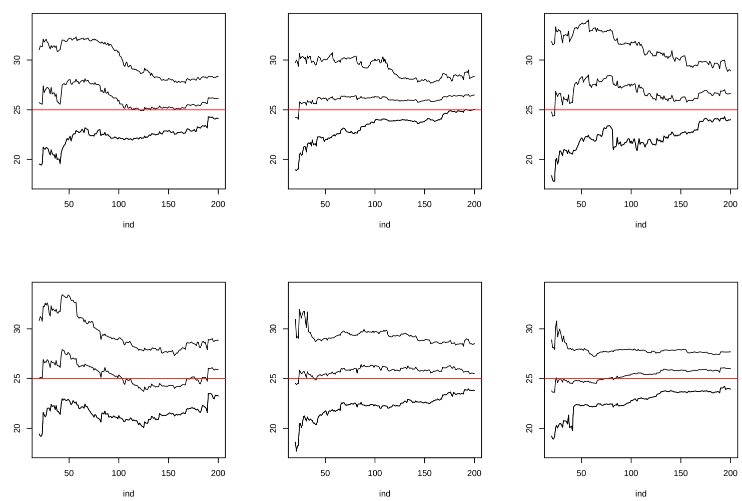


Example ii. Assessing MC error - α

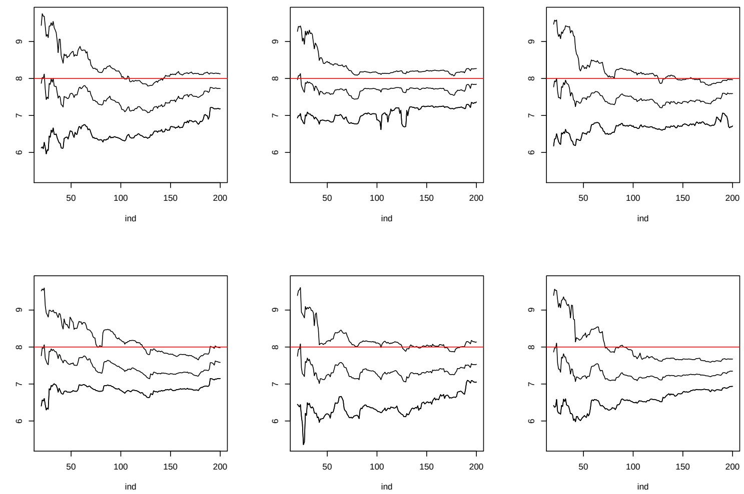
$N = 10000$



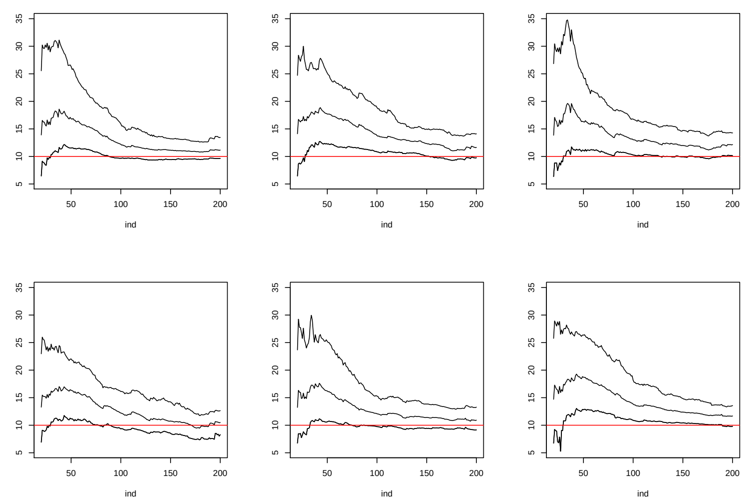
Example ii. Assessing MC error - β



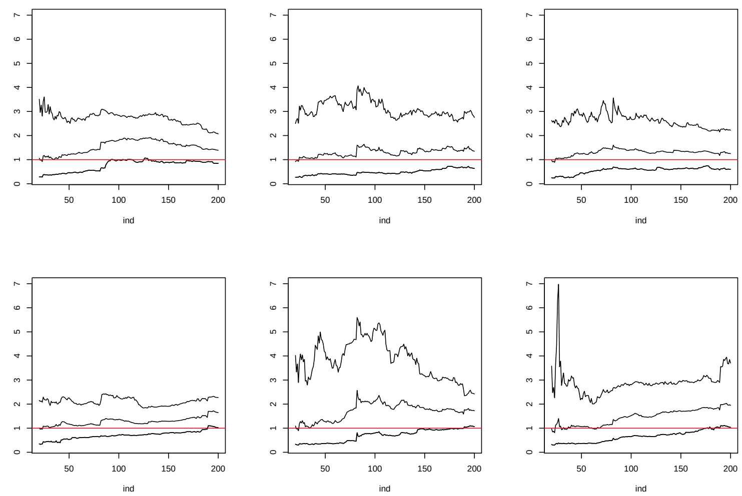
Example ii. Assessing MC error - γ



Example ii. Assessing MC error - σ^2



Example ii. Assessing MC error - τ^2



Particle Learning (PL)

Carvalho, Johannes, Lopes and Polson (2009) introduce Particle Learning (PL) as the following **resample-sample** scheme, where s_t is the vector of sufficient statistics for θ.

- Posterior at t: $\{(x_t, s_t, \theta)^{(i)}\}_{i=1}^N \sim p(x_t, s_t, \theta | y^t)$.
- Resampling weights: $w_{t+1}^{(j)} \propto p(y_{t+1}|x_t^{(j)}, \theta^{(j)})$, $j = 1, \ldots, N$.
- For $i = 1, \ldots, N$
  - Resample: Draw $\{(\tilde{x}_t, \tilde{s}_t, \tilde{\theta})^{(i)}\}_{i=1}^N$ from $\{(x_t, s_t, \theta)^{(j)}, w_{t+1}\}_{j=1}^N$.
  - Sample: Draw $x_{t+1}^{(i)} \sim p(x_{t+1}|(\tilde{x}_t, \tilde{\theta})^{(i)}, y_{t+1})$.
  - Recursive sufficient statistics: $s_{t+1}^{(i)} = \mathcal{S}(\tilde{s}_t^{(i)}, x_{t+1}^{(i)}, y_{t+1})$.
  - Offline sampling of fixed parameters: $\theta^{(i)} \sim p(\theta|s_{t+1}^{(i)})$.

PL ingredients

Resampling distribution

$$p(y_{t+1}|x_t, \theta)$$

Propagating distribution

$$p(x_{t+1}|x_t, \theta, y_{t+1})$$

Recursive sufficient statistics

$$s_{t+1} = \mathcal{S}(s_t, x_{t+1}, y_{t+1})$$

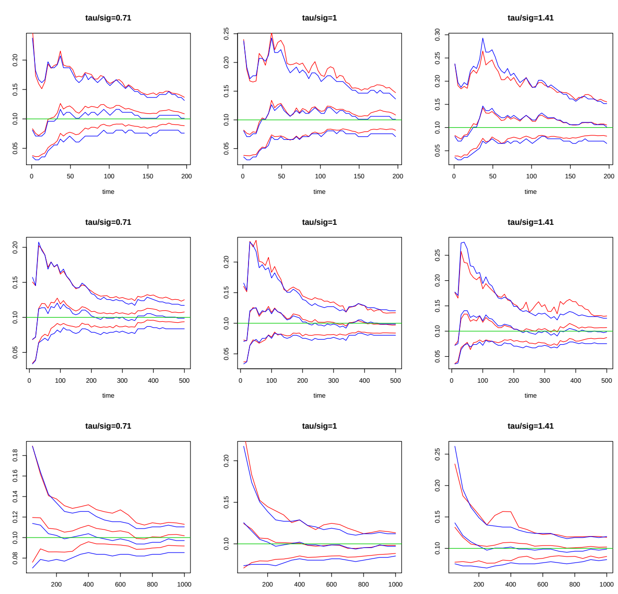
Example i. 1st order dynamic linear model via PL

- $p(y_{t+1}|x_t, \sigma^2)$ is
$$N(y_{t+1}; x_t, \sigma^2 + \tau^2)$$

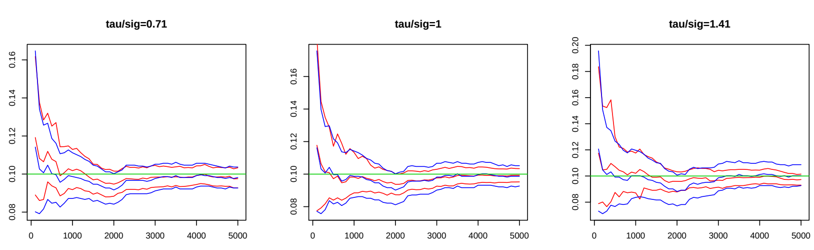
- $p(x_{t+1}|x_t, \sigma^2, y_{t+1})$ is
$$N(x_{t+1}; Ay_{t+1} + (1 - A)x_t, A\sigma^2)$$
where $A = \tau^2/(\tau^2 + \sigma^2)$.

- $p(\sigma^2|s_{t+1})$ is
$$IG(a_{t+1}, b_{t+1})$$
where $s_{t+1} = (a_{t+1}, b_{t+1})$ is recursively updated
$$\begin{aligned} a_{t+1} &= a_t + \frac{1}{2} \\ b_{t+1} &= b_t + \frac{(y_t - x_t)^2}{2} \end{aligned}$$

Example i. learning τ^2 **-** $n = 200, 500, 1000$



Example i. learning τ^2 **-** $n = 5000$ **and** $N = 2000$



Example iii. MCMC and PL comparison

- **Simulation set up:** For $t = 1, \ldots, n = 300$,

$$p(y_t|x_t, \theta) \equiv f_N(x_t; \sigma^2) \quad (1)$$
$$p(x_t|x_{t-1}, \theta) \equiv f_N(\rho x_{t-1}; \tau^2) \quad (2)$$

  where $\theta = (\rho, \sigma^2, \tau^2) = (1.0, 1.0, 0.25)$ and $x_0 = 0$.

- **Model set up:** Equations (1) and (2) above plus

$$\begin{aligned} p(\theta, x_0) &\equiv f_N(\rho; r_0, W_0) f_{IG}(\sigma^2; a_0, b_0) \\ &\times f_{IG}(\tau^2; c_0, d_0) f_N(x_0; m_0, V_0) \end{aligned}$$

  where $r_0 = 0$, $W_0 = 3$, $a_0 = 3$, $b_0 = 2$, $c_0 = 3$, $d_0 = 0.5$, $m_0 = 0$ and $V_0 = 3$.

- **MCMC set up:** $M_0 = 100K$, $L = 100$ and $M = 20K$. A total of $2100K$ draws.

- **SMC set up:** $M = 20K$ particles.

Example iii. MCMC algorithms

- Gibbs sampler (GIBBS)
  - Sample x from $p(x|y, \theta)$ - FFBS
  - Sample σ^2 from $p(\sigma^2|x, y)$
  - Sample ρ from $p(\rho|x, \tau^2)$
  - Sample τ^2 from $p(\tau^2|x, \rho)$
- Random-walk Metropolis (RW)
  - Sample x from $p(x|y, \theta)$ - FFBS
  - Sample θ^* from

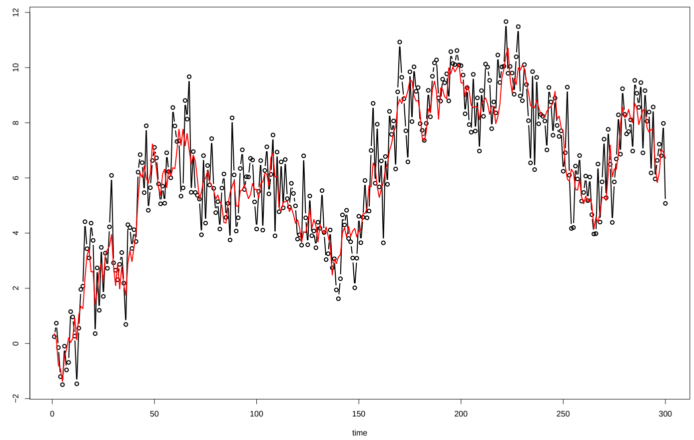
$$q(\theta^*|\theta) = q_\rho(\rho, V_\rho) q_{\sigma^2}(\sigma, V_{\sigma^2}) q_{\tau^2}(\tau^2, V_{\tau^2}),$$

    with $V_\rho = 0.01$, $V_{\sigma^2} = 0.01$ and $V_{\tau^2} = 0.01$, and accept with probability

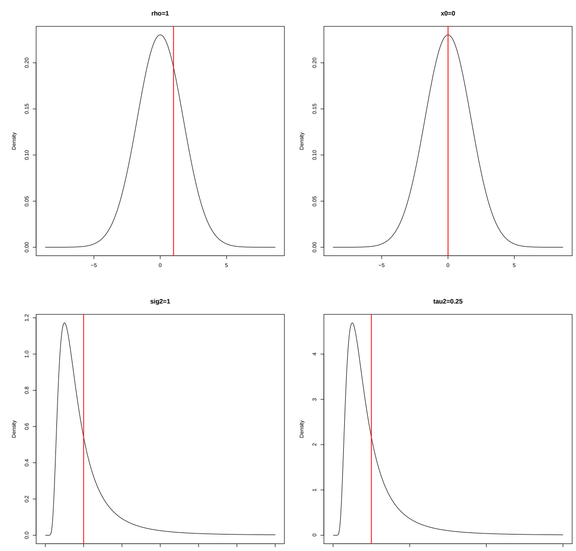
$$\alpha = \min\left\{1, \frac{p(y|\theta^*)p(\theta^*)q(\theta^*|\theta)}{p(y|\theta)p(\theta)q(\theta|\theta^*)}\right\}.$$

Note : Since $p(y|\theta) = \int p(y|x, \theta)p(x|\theta)dx$ can be analytically derived, x and θ are jointly sampled.

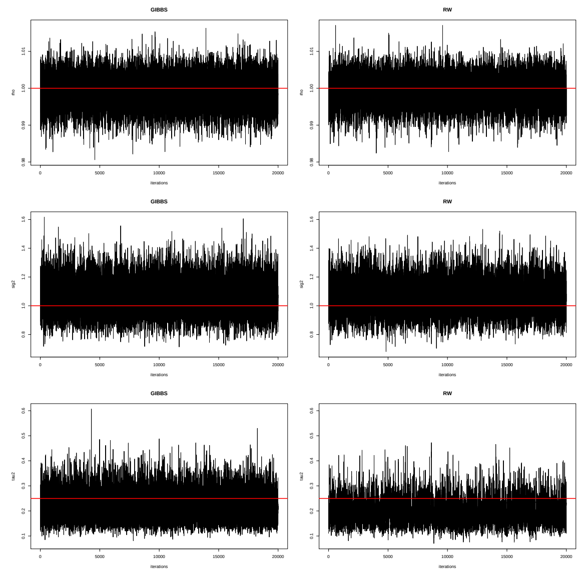
Example iii. Simulated y_t and x_t



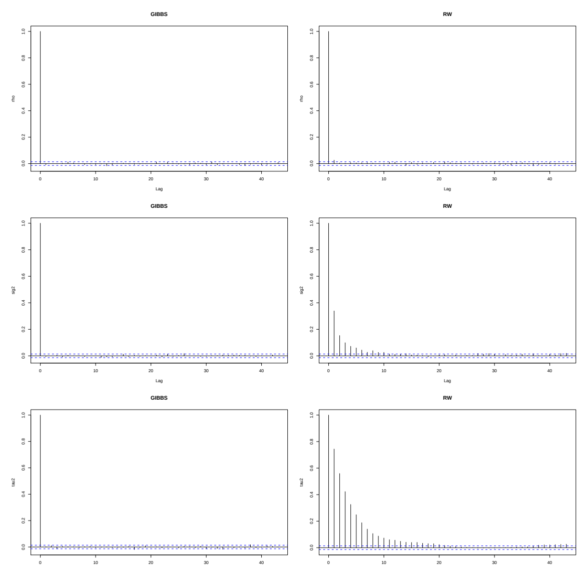
Example iii. Prior of (θ, x_0)



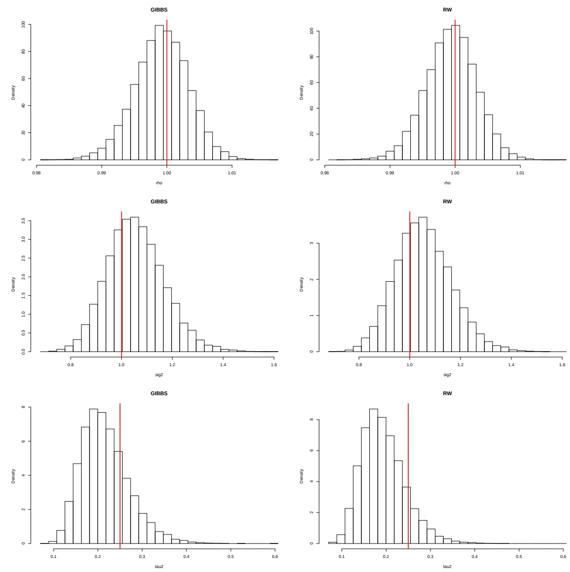
Example iii. MCMC trace plots



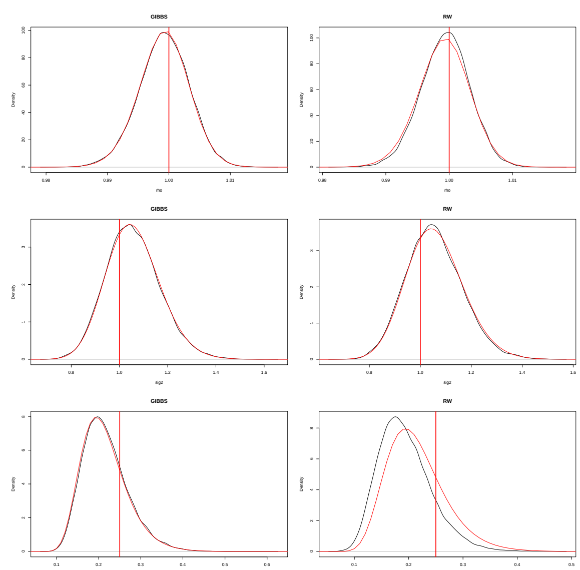
Example iii. MCMC autocorrelation plots



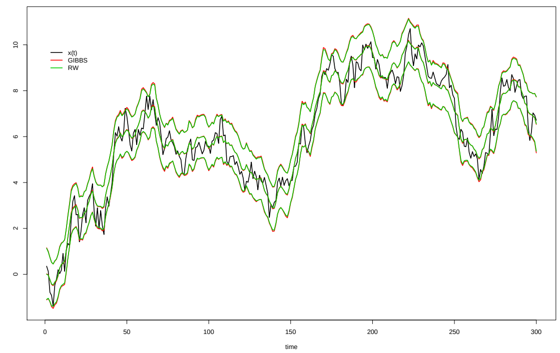
Example iii. MCMC marginal posteriors



Example iii. MCMC and true marginal posteriors

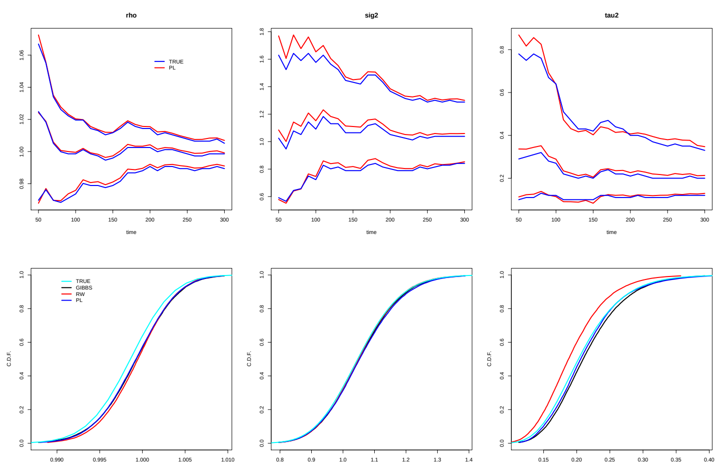


Example iii. $p(x_t|y^n)$ via MCMC



Posterior medians and 95% credibility intervals.

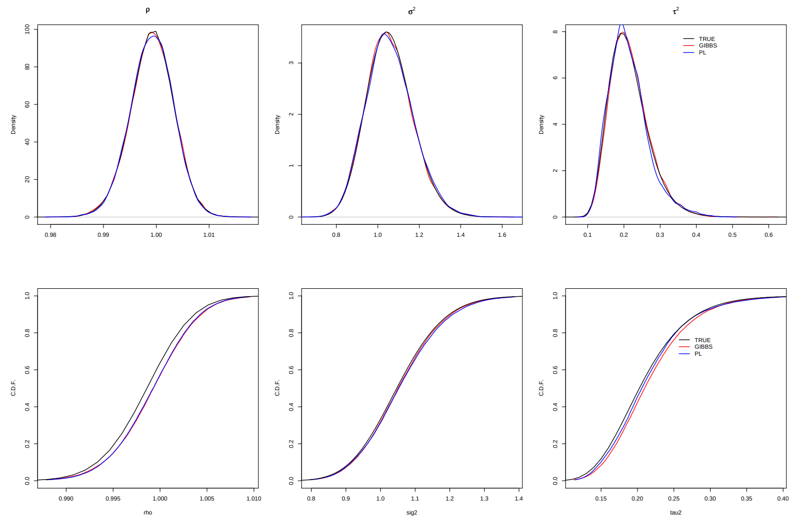
Example iii. PL and MCMC quantiles



TOP: True and PL estimate of percentiles of $p(\theta|y^t)$ for all t. Percentiles are 2.5%,50% and 97.5%.

BOTTOM: True, GIBBS and PL estimates of $F(\theta|y^n)$.

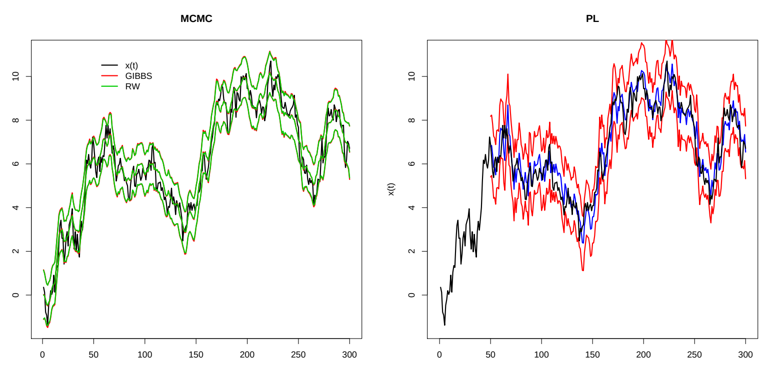
Example iii. PL and GIBBS quantiles



TOP: True, Gibbs and PL estimate of $p(\theta|y^n)$.

BOTTOM: True, Gibbs and PL estimates of $F(\theta|y^n)$.

Example iii. PL and MCMC quantiles

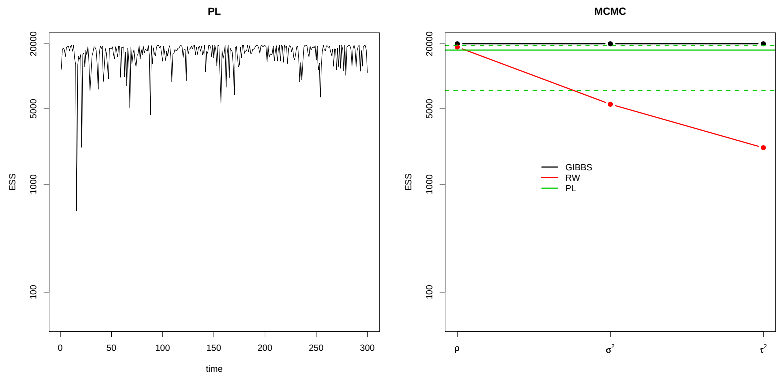


LEFT: Posterior medians and 95% credibility intervals of $p(x_t|y^n)$.

RIGHT: Posterior medians and 95% credibility intervals of $p(x_t|y^t)$.

Example iii. Effective sample sizes

$$ESS_t = M\left(1 + \frac{V(\omega_t)}{E^2(\omega_t)}\right)^{-1} \quad \text{and} \quad ESS = M\left(1 + 2\sum_{k=1}^{\infty}\rho_k\right)^{-1}.$$



ω_t are particle weights.

$\rho_k = cov_\pi(t^{(n)}, t^{(n+k)})/var_\pi(t^{(n)})$ and $t^{(n)} = t(\theta^{(n)})$.

STOCHASTIC VOLATILITY via SEQUENTIAL MONTE CARLO METHODS

Example i: Stochastic volatility

Let y_t, for $t = 1, \ldots, n$, be modeled as

$$\begin{aligned} y_t|x_t &\sim N(0, e^{x_t}) \\ (x_t|x_{t-1}, \theta) &\sim N(\alpha + \beta x_{t-1}, \tau^2) \end{aligned}$$

where $\theta = (\alpha, \phi, \tau^2)$.

Simulation setup: $n = 500$, $\alpha = -0.0031$, $\beta = 0.9951$ and $\tau^2 = 0.0074$ and $x_1 = \alpha/(1-\beta) = -0.632653$ (13% of annualized standard deviation).

Prior setup:

$$\begin{aligned} x_0 &\sim N(m_0, C_0) \qquad \alpha \sim N(\alpha_0, V_\alpha) \\ \beta &\sim N(\beta_0, V_\beta) \qquad \tau^2 \sim IG(n_0/2, n_0\tau_0^2/2) \end{aligned}$$

where $m_0 = 0.0$, $C_0 = 0.1$, $\alpha_0 = -0.0031$, $V_\alpha = 0.01$, $\beta_0 = 0.9951$, $V_\beta = 0.01$, $n_0 = 3$ and $\tau_0^2 = 0.0074$.

LW filter with shrinkage factor a

Particles t: $\{(x_t, \theta)^{(j)}, \omega_t^{(j)}\}_{j=1}^M \sim p(x_t, \theta|y^t)$.

Summary of $p(\theta|y^t)$: $\bar{\theta} \approx E(\theta|y^t)$ and $V \approx V(\theta|y^t)$.

Resample quantities: For $j = 1, \ldots, M$

- Compute $m^{(j)} = a\theta^{(j)} + (1-a)\bar{\theta}$
- Compute $g^{(j)} = \alpha^{(j)} + \phi^{(j)} x_t^{(j)}$

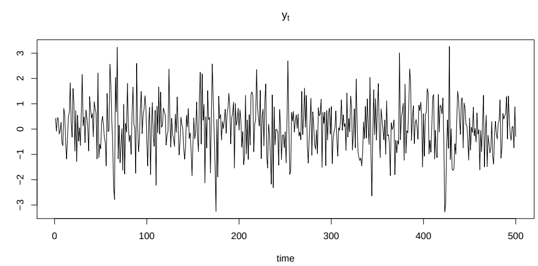
Algorithm: For $l = 1, \ldots, M$

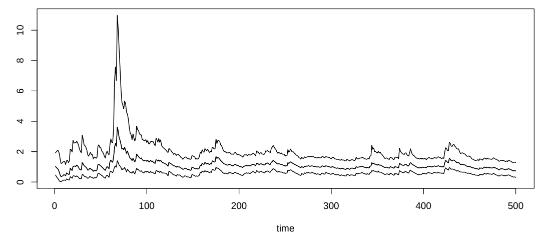
- Draw $k^l \in \{1, \ldots, M\}$, with $P(k^l = j) \propto \omega_t^{(j)} p(y_{t+1}|g^{(j)})$
- Sample $\theta^{(l)}$ from $N(m^{(k^l)}, (1-a^2)V)$
- Sample $x_{t+1}^{(l)}$ from $p(x_{t+1}|x_t^{(k^l)}, \theta^{(l)})$
- Compute weight $\omega_{t+1}^{(l)} \propto p(y_{t+1}|x_{t+1}^{(l)})/p(y_{t+1}|g^{(k^l)})$

Particles at $t+1$: $\{(x_{t+1}, \theta)^{(j)}, \omega_{t+1}^{(j)}\}_{j=1}^M \sim p(x_{t+1}, \theta|y^{t+1})$.

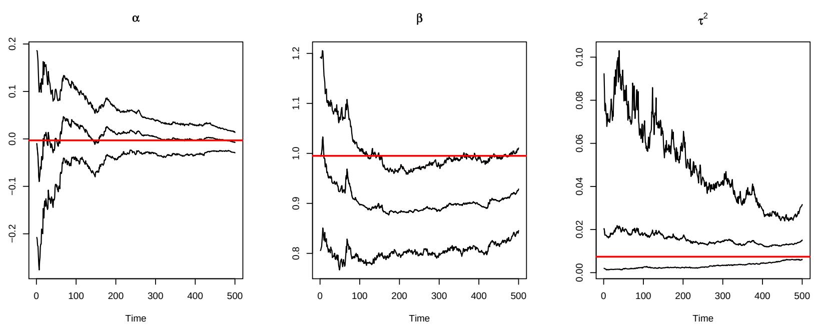
Time series y_t and $p(e^{x_t}|y^t)$

$N = 5000$ and $\theta = (\alpha, \beta, \log(\tau^2))$.





Parameter learning



Example ii: SV-AR(1) via sequential MCMC and LW

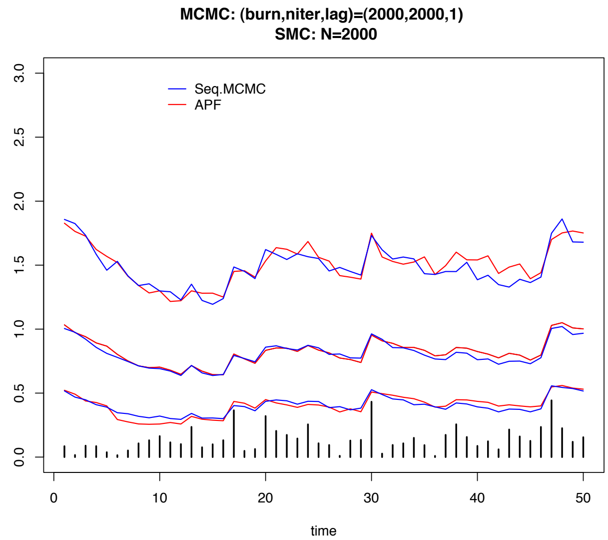
We simulated $n = 50$ observations based on $\alpha = -0.0031$, $\beta = 0.9951$, $\tau^2 = 0.0074$, with $m_0 = 0.0$ and $C_0 = 0.1$.

Also, $x_1 = \alpha/(1 - \beta) = -0.632653$, which corresponds to annualized standard deviations around 13%.

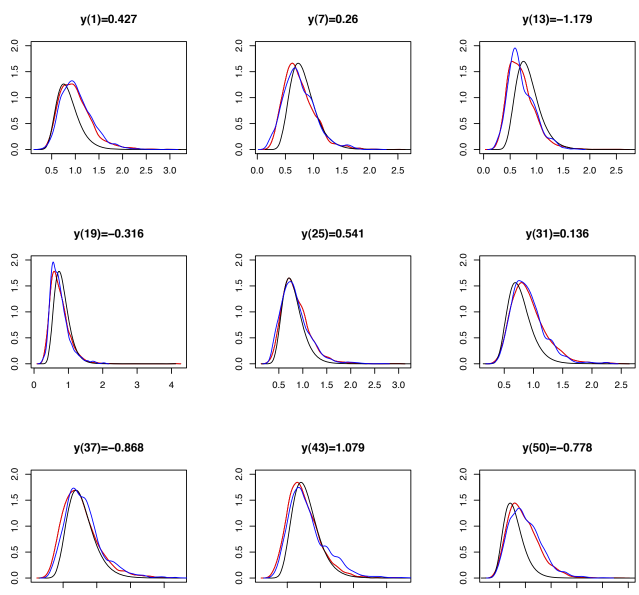
$p(x_t|y^t)$ when θ is known

MCMC: Kim, Shephard and Chib (1994)

SMC: Liu and West (2001) with $\delta = 0.75$ and $a = 0.9521743$.

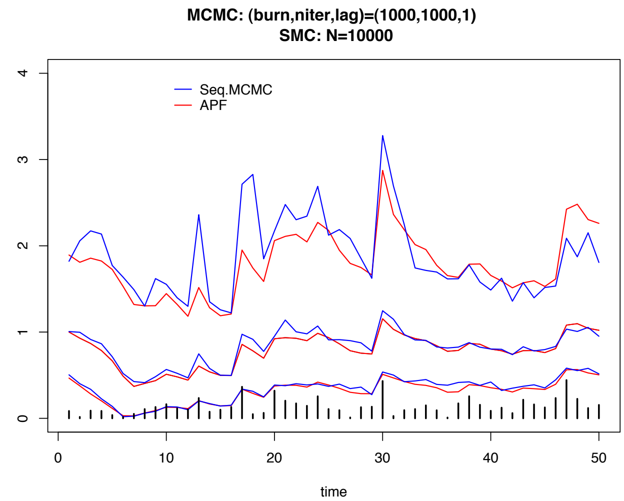


$p(x_t|y^t)$ **when** θ **is known**

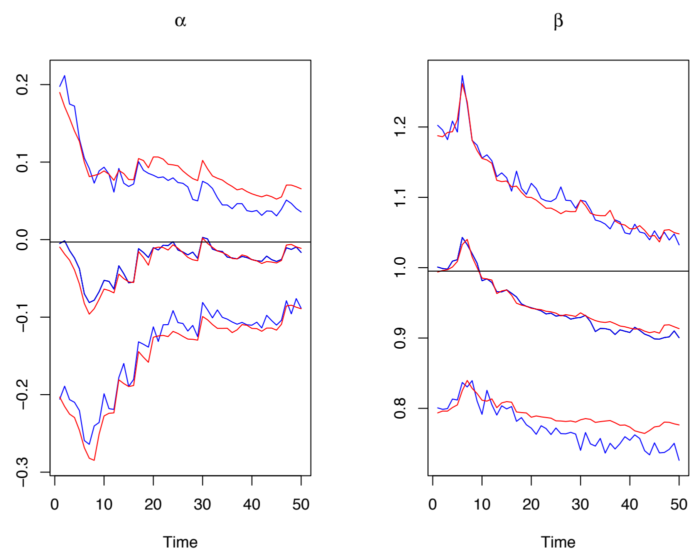


$p(x_t|y^t)$ **when** (α, β) **is unknown**

Prior: $\alpha \sim N(-0.0031, 0.01)$ and $\phi \sim N(0.9951, 0.01)$

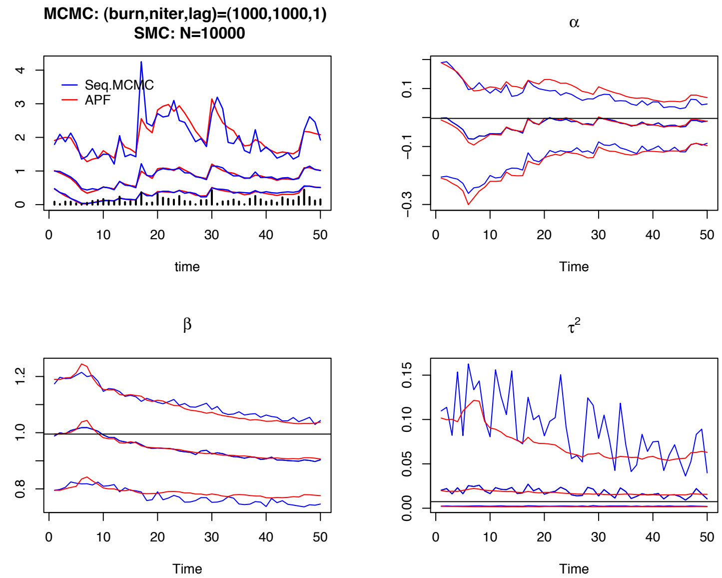


$p(\alpha|y^t)$ **and** $p(\beta|y^t)$

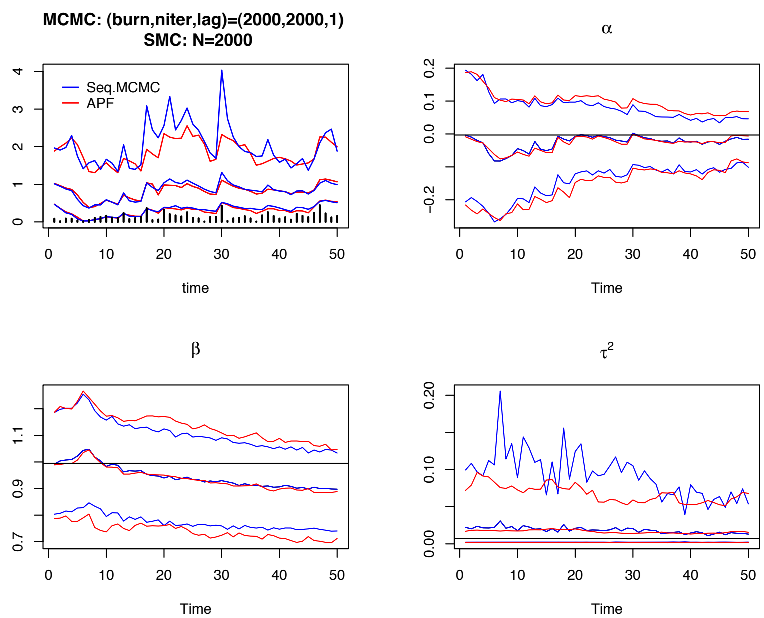


Learning x_t, α, β and τ^2

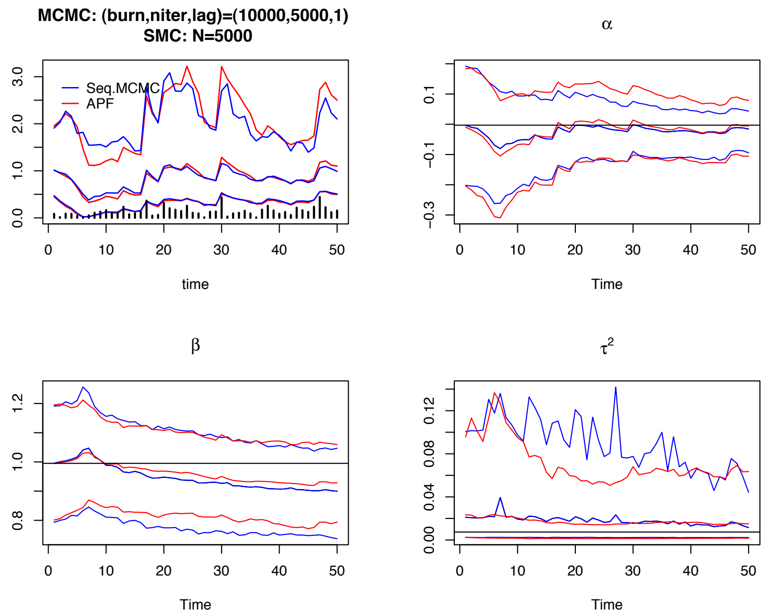
The prior for τ^2 is $IG(1.5, 0.0111)$.



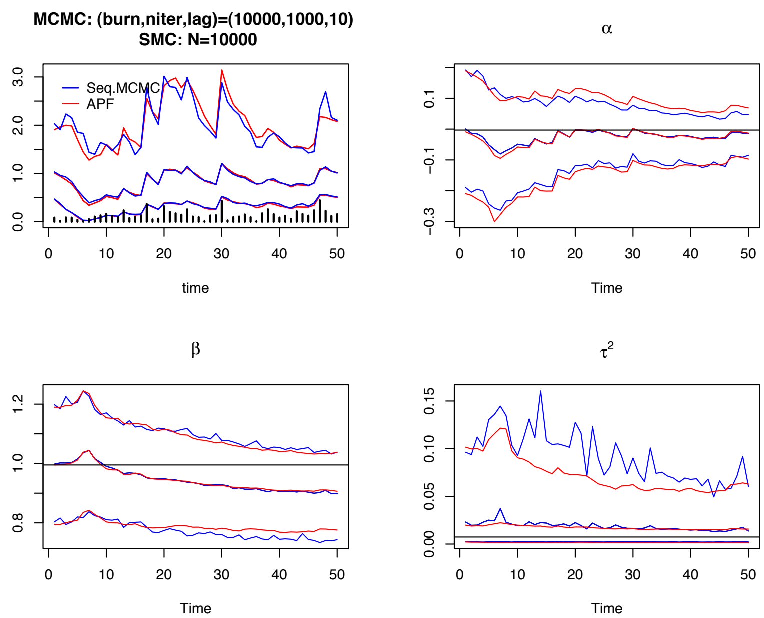
Learning x_t, α, β and τ^2



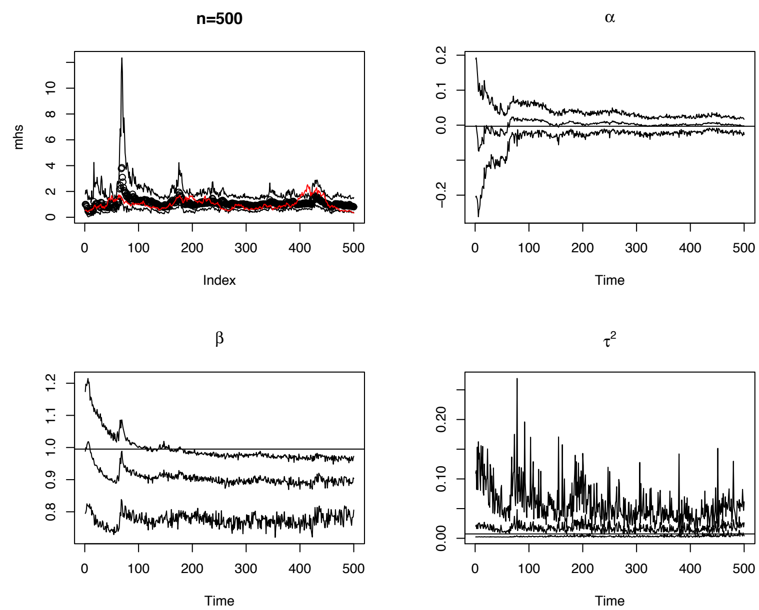
Learning x_t, α, β and τ^2



Learning x_t, α, β and τ^2

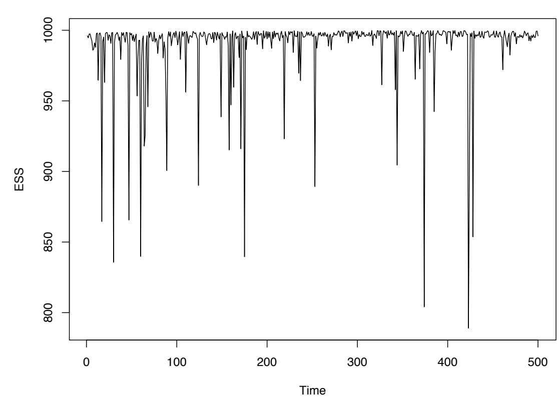


Sequential MCMC when $n = 500$

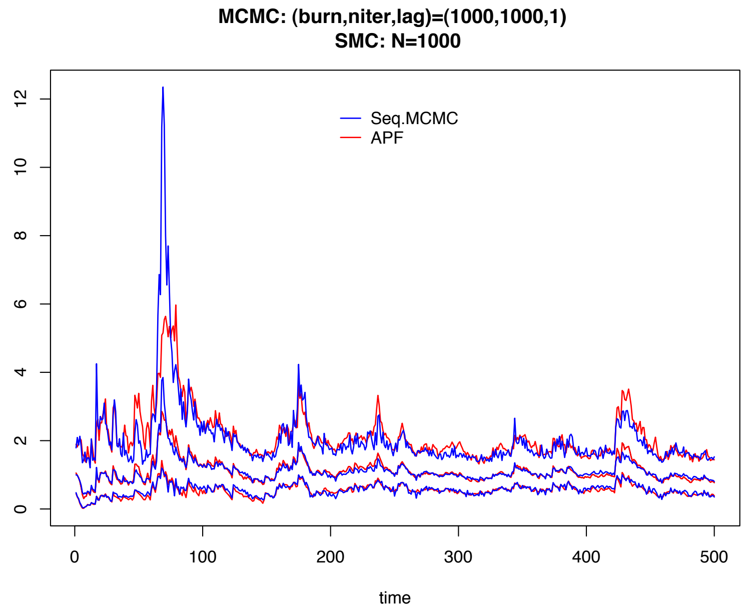


Effective sample size

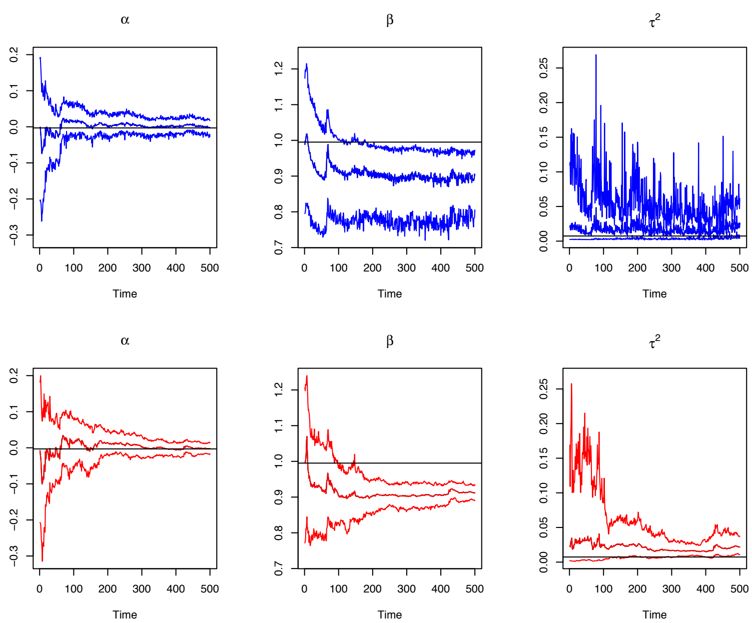
$$ESS_t = \frac{N}{1 + \frac{V(\omega_t)}{E^2(\omega_t)}}$$



Sequential volatilities



Sequential parameter learning



Example iii: Markov switching stochastic volatility

Carvalho and Lopes (2007) adapted APF and LW filters to sequentially estimate states and parameters in Markov switching stochastic volatility (MSSV) models.

Jul 2nd, 97	Thailand devalues the baht by as much as 20%.
Aug 11th, 97	IMF and Thailand set a rescue agreement.
Oct 23rd, 97	Hong Kong's stock index falls 10.4%. South Korea Won weakens.
Dec 2nd, 97	IMF and South Korea set a bailout agreement.
Jun 1st, 98	Russia's stock market crashes.
Jun 20th, 98	IMF gives final approval to a loan package to Russia.
Aug 19th, 98	Russia officially falls into default.
Oct 09th, 98	IMF and World Bank joint meeting + Fed cuts interest rates.
Jan 15th, 99	The real is allowed to float freely by lifting exchange controls.
Feb 2nd, 99	Arminio Fraga is named president of Brazil's Central Bank.

Let the daily returns of the IBOVESPA index, y_t, be modeled by a MSSV model, ie.

$$\begin{aligned} y_t|\lambda_t &\sim N(0, \exp(\lambda_t)) \\ (\lambda_t|\lambda_{t-1}, \xi, s_t) &\sim N(\alpha_{s_t} + \phi\lambda_{t-1}, \sigma^2) \end{aligned}$$

where $\xi = (\alpha, \phi, \sigma^2)$, $\alpha = (\alpha_1, \ldots, \alpha_k)$ and regime variables s_t following a k-state first order Markov process,

$$p_{ij} = Pr(s_t = j|s_{t-1} = i) \qquad \text{for } i, j = 1, \ldots, k$$

and $P = (p_{11}, \ldots, p_{1k-1}, \ldots, p_{k1}, \ldots, p_{k,k-1})$.

Particle filter

- *Step 0:* $\left\{\lambda_t^{(j)}, s_t^{(j)}, w_t^{(j)}\right\}_{j=1}^{M} \sim p(\lambda_t, s_t, \theta|D_t)$
- *Step 1:* For $j = 1, \ldots, M$,

$$\begin{aligned} \tilde{s}_{t+1}^{(j)} &= \arg\max_{l \in 1,\ldots,k} Pr(s_{t+1} = l|s_t = s_t^{(j)}) \\ \mu_{t+1}^{(j)} &= \alpha_{\tilde{s}_{t+1}^{(j)}}^{(j)} + \phi_t^{(j)}\lambda_t^{(j)} \end{aligned}$$

- *Step 2:* For $l = 1, \ldots, M$
  1. Sample k^l from $\{1, \ldots, k\}$, with $Pr(k^l) \propto \; p(y_{t+1}|\mu_{t+1}^{(k^l)})w_t^{(k^l)}$
  2. Sample $\theta_{t+1}^{(l)}$ from $N(m_t^{(k^l)}, b^2V_t)$
  3. Sample $s_{t+1}^{(l)}$ from $1, \ldots, k$ with $Pr(s_{t+1}^{(l)}) = Pr(s_{t+1}^{(l)}|s_t^{(k^l)})$
  4. Sample $\lambda_{t+1}^{(l)}$ from $p(\lambda_{t+1}|\lambda_t^{(k^l)}, s_{t+1}^{(l)}, \theta_{t+1}^{(l)})$
- *Step 3:* For $l = 1, \ldots, M$, compute new weights

$$w_{t+1}^{(l)} \propto p(y_{t+1}|\lambda_{t+1}^{(l)})/p(y_{t+1}|\mu_{t+1}^{(k^l)})$$

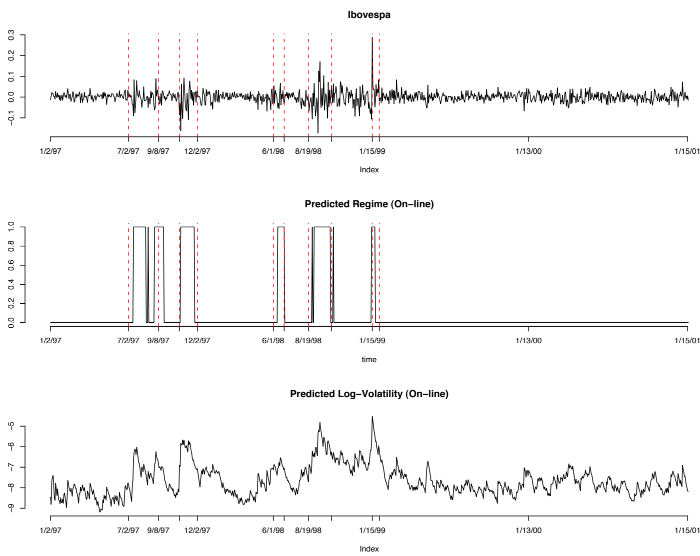
- *Step 4:* $\left\{\lambda_{t+1}^{(j)}, S_{t+1}^{(j)}, w_{t+1}^{(j)}\right\}_{j=1}^{M} \sim p(\lambda_{t+1}, S_{t+1}, \theta|D_{t+1})$.

Currency crisis

Carvalho and Lopes (2007) used IBOVESPA daily data from January 2nd, 1997 to January, 16th 2001 (1000 observations).

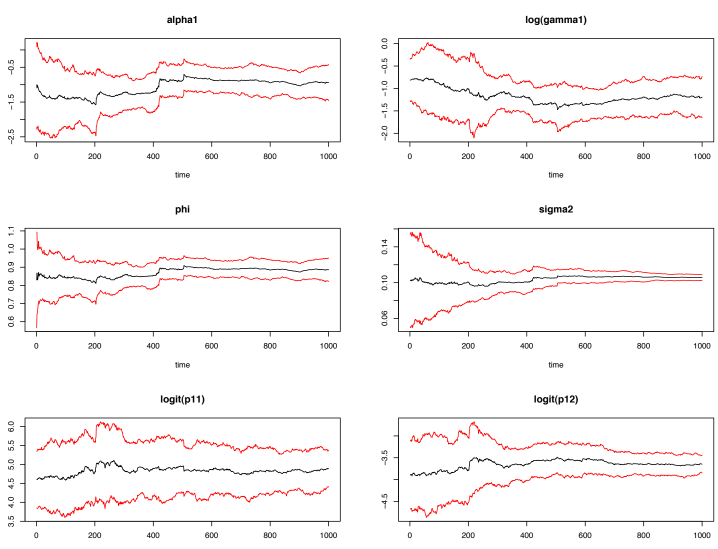
Fitting regime shifts

The vertical lines indicate key market events.

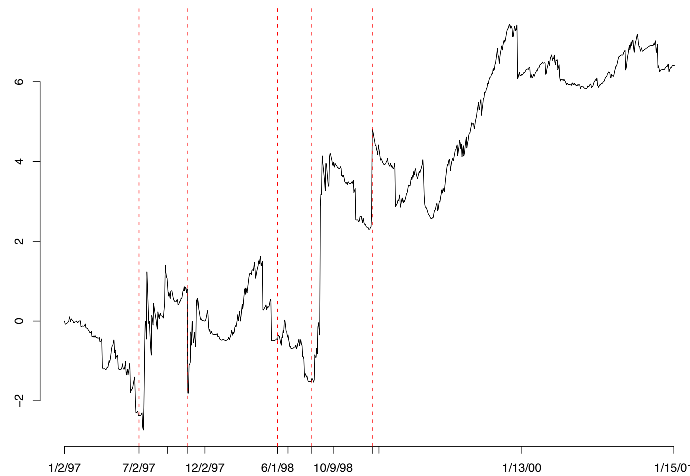


Sequential inference for fixed parameters

Posterior mean, 5% and 95% quantiles of θ.



Sequential Bayes factor: MSSV vs SV



Particle Learning

This example was kindly prepared by my PhD student Samir Warty.

Recall the AR(1) stochastic volatility model:

$$y_{t+1} = \exp\left(\frac{x_{t+1}}{2}\right)\epsilon_{t+1}$$
$$x_{t+1} = \alpha + \beta x_t + \tau \nu_{t+1}$$

where $(\epsilon_t, \nu_t) \sim N(0_2, I_2)$ and $\theta = (\alpha, \beta, \tau)$.

Prior distribution:

$$\tau^2 | s_0 \sim \mathcal{IG}\left(\frac{n_0}{2}, \frac{n_0 S_0}{2}\right)$$
$$\alpha, \beta | \tau^2, s_0 \sim N(m_0, \tau^2 C_0)$$

where $s_0 = (n_0, S_0, m_0, C_0)$.

Data augmentation argument

Following Kim, Shephard and Chib's (1998) idea:

$$z_{t+1} \equiv \log y_{t+1}^2 = x_{t+1} + \log \epsilon_{t+1}^2 \approx x_{t+1} + u_t$$

where

$$u_t \sim \sum_{i=1}^{7} \pi_i \mathcal{N}(\mu_i, \sigma_i^2)$$

Particle learning (PL) uses augments the state vector to include $\lambda_{t+1} \in \{1, \ldots, 7\}$, the component of the Normal mixture approximation.

Let s_t denote the set of sufficient statistics for (α, β, τ^2) at time t.

Algorithm

- Resample old particles $c_t = (x_t, s_t, \theta)$ with weights

$$w_t \propto p(z_{t+1}|c_t) = \sum_{i=1}^{7} \pi_i f_N(z_{t+1}; \mu_i + \alpha + \beta x_t, \sigma_i^2 + \tau^2)$$

- Propagate new states x_{t+1} from

$$p(x_{t+1}|c_t, z_{t+1}) = \sum_{i=1}^{7} \pi_i f_N(x_{t+1}; \gamma_i, \omega_i)$$

where

$$\begin{aligned} \omega_i &= (\sigma_i^{-2} + \tau^{-2})^{-1} \\ \gamma_i &= \omega_i(\sigma_i^{-2}(z_{t+1} - \mu_i) + \tau^{-2}(\alpha + \beta x_t)) \end{aligned}$$

Algorithm (cont.)

- Update sufficient statistics $s_{t+1} = (n_{t+1}, S_{t+1}, m_{t+1}, C_{t+1})$

$$\begin{aligned} n_{t+1} &= n_t + 1 \\ n_{t+1}S_{t+1} &= n_t S_t + \frac{(x_{t+1} - X_t m_t)^2}{1 + X_t C_t X_t'} \\ C_{t+1}^{-1} &= C_t^{-1} + X_t' X_t \\ C_{t+1}^{-1} m_{t+1} &= C_t^{-1} m_t + X_t' x_{t+1} \end{aligned}$$

where $X_t = (1, x_t)$.

- Sample parameters

$$\begin{aligned} \tau^2|s_t &\sim \mathcal{IG}\left(\frac{n_{t+1}}{2}, \frac{n_{t+1}S_{t+1}}{2}\right) \\ \alpha, \beta|\tau^2, s_t &\sim \mathcal{N}(m_{t+1}, \tau^2 C_{t+1}) \end{aligned}$$

Resampling weights

$$\begin{aligned} w_t &\propto p(z_{t+1}|c_t, \lambda_t, z^t) \\ &\propto \sum_{i=1}^{7} \int_R p(z_{t+1}|x_t, s_t, \theta, \lambda_{t+1} = i, \lambda_t, z^t, x_{t+1}) p(x_{t+1}|x_t, s_t, \theta, \lambda_{t+1} = i, \lambda_t) dx_{t+1} \\ &\qquad \text{(Marginalization over data augmentation)} \\ &\propto \sum_{i=1}^{7} \int_R p(z_{t+1}|\lambda_{t+1} = i, x_{t+1}) p(x_{t+1}|x_t, \theta) dx_{t+1} \qquad \text{(Conditional independence)} \\ &\propto \sum_{i=1}^{7} \int_R f_N(z_{t+1}; \mu_i + x_{t+1}, \sigma_i^2) f_N(x_{t+1}; \alpha + \beta x_t, \tau^2) dx_{t+1} \\ &\propto \sum_{i=1}^{7} \pi_i f_N(z_{t+1}; \mu_i + \alpha + \beta x_t, \sigma_i^2 + \tau^2) \end{aligned}$$

Posterior distribution for new states

$$
\begin{aligned}
p(x_{t+1}|c_t,\lambda_t,z_{t+1}) &= \sum_{i=1}^{7}\pi_i p(x_{t+1}|c_t,\lambda_{t+1}=i,\lambda_t,z_{t+1}) && \text{(Marginalization over data augmentation)}\\
&= \sum_{i=1}^{7}\pi_i \frac{p(z_{t+1}|x_t,s_t,\theta,\lambda_{t+1}=i,\lambda_t,x_{t+1})p(x_{t+1}|x_t,s_t,\theta,\lambda_{t+1}=i,\lambda_t)}{p(z_{t+1}|x_t,s_t,\theta,\lambda_{t+1}=i,\lambda_t)} && \text{(Bayes theorem)}\\
&= \sum_{i=1}^{7}\pi_i \frac{p(z_{t+1}|\lambda_{t+1}=i,x_{t+1})p(x_{t+1}|x_t,\theta)}{p(z_{t+1}|x_t,\theta,\lambda_{t+1}=i)} && \text{(Conditional independence)}\\
&= \sum_{i=1}^{7}\pi_i \frac{f_N(z_{t+1};\mu_i+x_{t+1},\sigma_i^2)f_N(x_{t+1};\alpha+\beta x_t,\tau^2)}{f_N(z_{t+1};\mu_i+\alpha+\beta x_t,\sigma_i^2+\tau^2)}\\
&= \sum_{i=1}^{7}\pi_i f_N(x_{t+1};\gamma_i,\omega_i)
\end{aligned}
$$

where $\omega_i = (\sigma_i^{-2}+\tau^{-2})^{-1}$ and $\gamma_i = \omega_i(\sigma_i^{-2}(z_{t+1}-\mu_i)+\tau^{-2}(\alpha+\beta x_t))$.

Recursive sufficient statistics

$$
\begin{aligned}
n_{t+1}S_{t+1} &= n_tS_t + (x_{t+1} - X_t(C_t^{-1}+X_t'X_t)^{-1}(C_t^{-1}m_t + X_t'x_{t+1}))'x_{t+1}\\
&\quad + (m_t - (C_t^{-1}+X_t'X_t)^{-1}(C_t^{-1}m_t+X_t'x_{t+1}))'C_t^{-1}m_t\\
&= n_tS_t + (x_{t+1}'x_{t+1} - x_{t+1}'X_t(C_t^{-1}+X_t'X_t)^{-1}(C_t^{-1}m_t+X_t'x_{t+1}))\\
&\quad + (m_t'(C_t^{-1})'m_t - m_t'(C_t^{-1})'(C_t^{-1}+X_t'X_t)^{-1}(C_t^{-1}m_t+X_t'x_{t+1}))\\
&= n_tS_t + (x_{t+1}'x_{t+1} - x_{t+1}'X_t\left(C_t - \frac{C_tX_t'X_tC_t}{1+X_tC_tX_t'}\right)(C_t^{-1}m_t+X_t'x_{t+1}))\\
&\quad + (m_t'C_t^{-1}m_t - m_t'C_t^{-1}\left(C_t - \frac{C_tX_t'X_tC_t}{1+X_tC_tX_t'}\right)(C_t^{-1}m_t+X_t'x_{t+1}))\\
&= n_tS_t + x_{t+1}'\left(1-c+\frac{c^2}{1+c}\right)x_{t+1} - x_{t+1}'\left(1-\frac{c}{1+c}\right)X_tm_t\\
&\quad - m_t'X_t'\left(1-\frac{c}{1+c}\right)x_{t+1} + \frac{m_t'X_t'X_tm_t}{1+c} && \text{(where } c \equiv X_tC_tX_t'\text{)}\\
&= n_tS_t + \left(\frac{1}{1+c}\right)(x_{t+1}'x_{t+1} - 2x_{t+1}'X_tm_t + m_t'X_t'X_tm_t)\\
&= n_tS_t + \left(\frac{1}{1+c}\right)(x_{t+1}-X_tm_t)'(x_{t+1}-X_tm_t)
\end{aligned}
$$

where

$$
(C_t^{-1}+X_t'X_t)^{-1} = \left(C_t - \frac{C_tX_t'X_tC_t}{1+X_tC_tX_t'}\right)
$$

when X_t is a vector.

REFERENCES

A few references on MC and MCMC

1. **Casella and George (1992) Explaining the Gibbs sampler. The American Statistician,46,167-74.**
2. **Chib and Greenberg (1995) Understanding the Metropolis-Hastings algorithm. The American Statistician, 49, 327-35.**
3. Gelfand and Smith (1990) Sampling-Based Approaches to Calculating Marginal Densities, JASA, 85, 398-409.
4. Geman and Geman (1984) Stochastic Relaxation, Gibbs Distributions and the Bayesian Restoration of Images. IEEE Transactions on Pattern Analysis and Machine Intelligence, 6, 721-41.
5. Geweke (1989) Bayesian inference in econometric models using Monte Carlo integration. Econometrica, 57, 1317-39.
6. Gilks and Wild (1992) Adaptive rejection sampling for Gibbs sampling. Applied Statistics, 41, 337-48. Computational Statistics and Data Analysis, 51, 4526-42.
7. Hastings (1970) Monte Carlo Sampling Methods Using Markov Chains and Their Applications. Biometrika, 57, 97-109.
8. Metropolis, Rosenbluth, Rosenbluth, Teller and Teller (1953) Equation of State Calculations by Fast Computing Machines. Journal of Chemical Physics, Number 21, 1087-92.
9. **Smith, A. F. M. and Gelfand, A. E. (1992) Bayesian statistics without tears: a sampling-resampling perspective. American Statistician, 46, 84-8.**

A few references on sequential Monte Carlo methods

1. **Carvalho, Johannes, Lopes and Polson (2008) Particle learning and smoothing. University of Chicago Graduate School of Business.**
2. Carvalho and Lopes (2007) Simulation-based sequential analysis of Markov switching stochastic volatility models,Computational Statistics and Data Analysis, 51, 4526-4542.
3. Doucet, de Freitas, and Gordon, 2001, Sequential Monte Carlo Methods in Practice, Springer, New York.
4. Fearnhead, P. (2002). Markov chain Monte Carlo, sufficient statistics, and particle filters. Journal of Computational and Graphical Statistics, 11, 848-862.
5. **Gordon, Salmond, and Smith, 1993, Novel Approach to Nonlinear/Non-Gaussian Bayesian State Estimation, IEE Proceedings, F-140, 107-113.**
6. Liu and West (2001). Combined parameters and state estimation in simulation-based filtering. In Sequential Monte Carlo Methods in Practice. Springer-Verlag New York.
7. **Pitt and Shephard (1999) Filtering via Simulation: Auxiliary Particle Filter, Journal of the American Statistical Association, 94, 590-599.**
8. Polson, Stroud and Muller (2008). Practical filtering with sequential parameter learning. Journal of the Royal Statistical Society, Series B, 70, 413-428.
9. **Storvik (2002). Particle filters in state space models with the presence of unknown static parameters, IEEE. Trans. of Signal Processing, 50, 281-289.**

A few references on dynamic models

1. **Carlin, Polson and Stoffer (1992) A Monte Carlo approach to nonnormal and nonlinear state-space modeling. Journal of the American Statistical Association, 87, 493-500.**
2. Carter and kohn (1994) On Gibbs sampling for state space models, Biometrika, 81, 541-553.
3. Fruhwirth-Schnatter (1994) Data augmentation and dynamic linear models, Journal of Time Series Analysis, 15, 183-102.
4. **Migon, Gamerman, Lopes and Ferreira (2005) Dynamic models. In Dey and Rao (Eds.) Handbook of Statistics, Volume 25.**
5. West and Harrison (1989/1997) Bayesian Forecasting and Dynamic Models. New York: Springer-Verlag.

A few references on stochastic volatility models

1. Berg, Meyer and Yu (2004), Deviance Information Criterion for Comparing Stochastic Volatility Models, Journal of Business and Economic Statistics, 22, 107-20.

2. **Eraker, Johannes and Polson (2003) The Impact of Jumps in Volatility and Returns, Journal of Finance, 2003, 58, 1269-300.**

3. Jacquier, Polson and Rossi (1994) Bayesian analysis of stochastic volatility models. Journal of Business and Economic Statistics, 12, 371-415.

4. Jensen and Maheu (2008) Bayesian semiparametric stochastic volatility modeling. Working paper 2008-15, Federal Reserve Bank of Atlanta.

5. **Johannes and Polson (2006) MCMC Methods for Financial Econometrics, Handbook of Financial Econometrics, Yacine Aït-Sahalia and Lars Hansen.**

6. **Kim, Shephard and Chib (1998) Stochastic volatility: Likelihood inference and comparison with ARCH models, Review of economic studies, 65, 36193.**

7. Lopes and Polson (2009) Extracting SP500 and NASDAQ volatility: The credit crisis of 2007-2008, Handbook of Applied Bayesian Analysis.

8. Lopes and Carvalho (2007) Factor stochastic volatility with time varying loadings and Markov switching regimes, Journal of Statistical Planning and Inference, 37, 3082-3091.

9. Lopes and Salazar (2006) Time series mean level and stochastic volatility modeling by smooth transition autoregressions: a Bayesian approach, In Fomby, T.B. (Ed.) Advances in Econometrics: Econometric Analysis of Financial and Economic Time Series/Part B, 2006, Volume 20, 229-242.

10. Polson, Stroud and Muller (2008) Practical Filtering with Sequential Parameter Learning, Journal of the Royal Statistical Society. Series B: Statistical Methodology, 70, pp. 413-28.