An introduction to SMC methods for nonnormal/nonlinear dynamic models

Hedibert Freitas Lopes

Graduate School of Business
University of Chicago
hlopes@ChicagoGSB.edu

2008-09 Program on Sequential Monte Carlo Methods
Opening Workshop, September 7-10, 2008
SAMSI, Research Triangle Park, NC

Outline

- Normal dynamic linear models (NDLM)
- Nonnormal/nonlinear dynamic models
- Sampling importance resampling (SIR)
- Example i: 1st order NDLM
  - Sequential Importance Sampling (SIS)
  - Sequential importance sampling with resampling (SISR)
  - Auxiliary particle filter (APF)
  - APF + parameter learning
- Example ii: Stochastic volatility (SV) model.
- Example iii: Nonlinear normal dynamic model.
- Example iv: Markov switching stochastic volatility (MSSV).
- Example v: SP500 and NDX100 volatility.

Normal linear dynamic model (NDLM)

Normal dynamic linear models (West and Harrison, 1997) are defined by a pair of equations: observation and system equations

$$\begin{aligned} p(y_t|\theta_t) &\equiv N(F_t'\theta_t, \sigma_t^2) \\ p(\theta_t|\theta_{t-1}) &\equiv N(G_t\theta_{t-1}, W_t) \end{aligned}$$

plus $p(\theta_1) \equiv N(a, R)$.

- $\{y_t\}$ conditionally independent given θ_t;
- F_t is a vector of explanatory variables;
- θ_t is the vector of states at time t;
- G_t describes parametric evolution;
- σ_t^2 and W_t are the error variances;

NDLM: sequential Bayes learning

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

$$p(\theta_{t-1}|y^{t-1}) \Longrightarrow p(\theta_t|y^{t-1}) \Longrightarrow p(y_t|\theta_t) \Longrightarrow p(\theta_t|y^t)$$

- Posterior at $t-1$: $p(\theta_{t-1}|y^{t-1}) \equiv N(m_{t-1}, C_{t-1})$
- Evolution
$$p(\theta_t|y^{t-1}) \equiv N(a_t, R_t)$$
with $a_t = G_t m_{t-1}$ and $R_t = G_t C_{t-1} G_t' + W_t$.
- Prediction
$$p(y_t|y^{t-1}) \equiv N(f_t, Q_t)$$
with $f_t = F_t' a_t$ and $Q_t = F_t' R_t F_t + \sigma_t^2$.
- Updating
$$p(\theta_t|y^t) \equiv N(m_t, C_t)$$
with $m_t = a_t + A_t(y_t - f_t)$, $C_t = R_t - A_t A_t' Q_t$ and $A_t = R_t F_t / Q_t$.

NDLM: smoothing

For $t = 1, \ldots, n$,

$$p(\theta_t | y^n) \equiv N(m_t^n, C_t^n)$$

where

$$\begin{aligned} m_t^n &= m_t + C_t G_{t+1}' R_{t+1}^{-1} (m_{t+1}^n - a_{t+1}) \\ C_t^n &= C_t - C_t G_{t+1}' R_{t+1}^{-1} (R_{t+1} - C_{t+1}^n) R_{t+1}^{-1} G_{t+1} C_t. \end{aligned}$$

NDLM: integrated likelihood

In many situations

$$\sigma_t = \sigma \quad W_t = W \quad G_t = G$$

for all t and $\psi = (\sigma, W, G)$ is the vector of *fixed parameters* of the system.

Then, by Bayes' rule

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

which can be approximated by MCMC, for instance.

Nonnormal/nonlinear dynamic models

- Similar to NDLM, nonnormal/nonlinear dynamic models are defined by observation and system equations

$$p(y_t|\theta_t, \psi) \quad \text{and} \quad p(\theta_t|\theta_{t-1}, \psi)$$

plus $p(\theta_1|\psi)$, with ψ known and omitted for now.

- **Evolution** and **updating** are represented by

$$\begin{aligned} p(\theta_t|y^{t-1}) &= \int p(\theta_t|\theta_{t-1}) p(\theta_{t-1}|y^{t-1}) d\theta_{t-1} \\ p(\theta_t|y^t) &\propto p(y_t|\theta_t) p(\theta_t|y^{t-1}) \end{aligned}$$

which are usually **unavailable in closed form**.

Sampling importance resampling (SIR)

Gordon, Salmond and Smith (1993) used SIR ideas (Smith and Gelfand, 1992) to *reinterpret* draws from

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

as draws from

$$p(\theta_t|y^t)$$

Goal: Sample from π

SIR algorithm

1. Draw $\theta_1^*, \ldots, \theta_N^*$ from q
2. Compute (unnormalized) weights $\omega_i = \pi(\theta_i^*)/q(\theta_i^*)$
3. Sample θ_j from $\{\theta_1^*, \ldots, \theta_N^*\}$ with weights $\{\omega_1, \ldots, \omega_N\}$

Output: $\theta_1, \ldots, \theta_N$ from π

Using the prior as proposal

In the Bayesian context, where

$$\pi(\theta) \equiv p(\theta|y) \propto p(\theta)p(y|\theta),$$

a natural (but not necessarily good) choice is

$$q(\theta) = p(\theta),$$

so weights are normalized likelihoods

$$\omega(\theta) \propto p(y|\theta)$$

Example 0.

- Normal likelihood

$$p(y|\theta) \propto \exp\left\{-\frac{n}{2\sigma^2}(\theta - \bar{y})^2\right\}$$

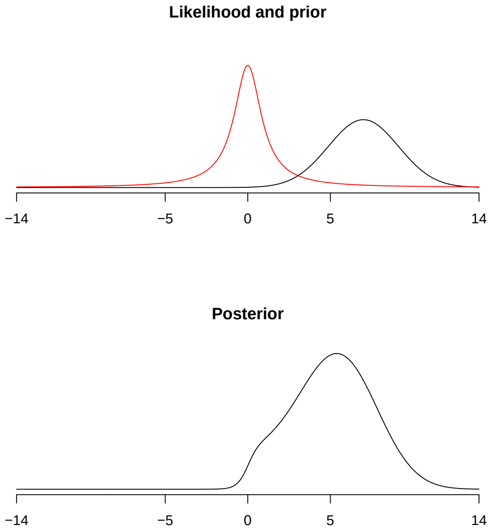
- Cauchy prior

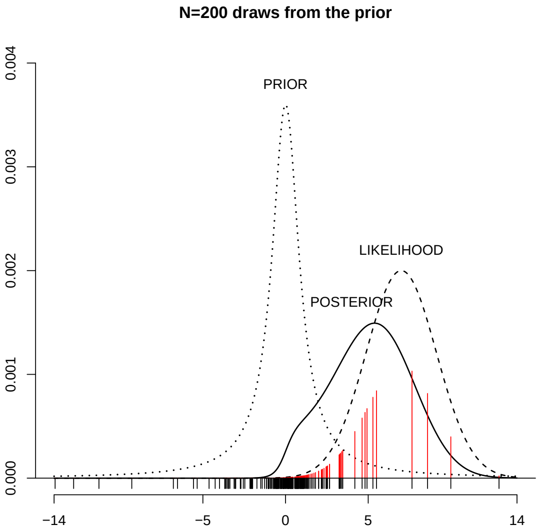
$$p(\theta) \propto \frac{1}{\tau_0^2 + (\theta - \theta_0)^2}$$

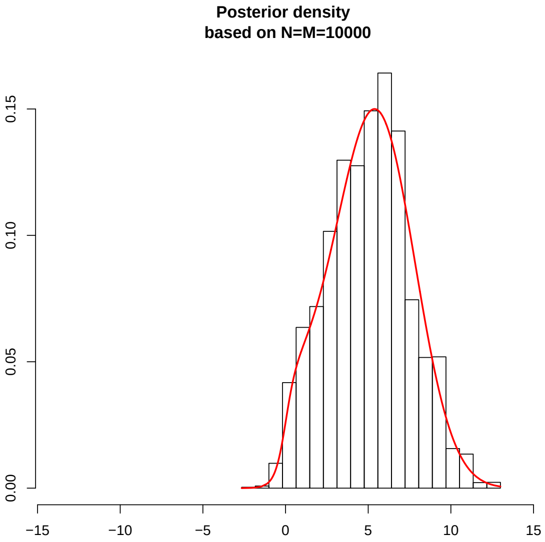
- Posterior

$$p(\theta|y) \propto \frac{\exp\left\{-\frac{n}{2\sigma^2}(\theta - \bar{y})^2\right\}}{\tau_0^2 + (\theta - \theta_0)^2}$$

Assume that $\sigma^2/n = 4.5$, $\bar{y} = 7$, $\theta_0 = 0$ and $\tau_0^2 = 1$.




N=200 draws from the prior
PRIOR
LIKELIHOOD
POSTERIOR
0.000
0.001
0.002
0.003
0.004
−14
−5
0
5
14



Posterior density
based on N=M=10000
0.00
0.05
0.10
0.15
−15
−10
−5
0
5
10
15

Example i. 1st order NDLM

- Model

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

- Posterior at $t = 0$

$$(\theta_0|y_0) \sim N(m_0, C_0)$$

- Prior at $t = 1$

$$(\theta_1|y_0) \sim N(m_0, C_0 + \tau^2)$$

- Likelihood

$$p(y_1|\theta_1) \propto \exp\left\{-\frac{1}{2\sigma^2}(y_1 - \theta_1)^2\right\}$$

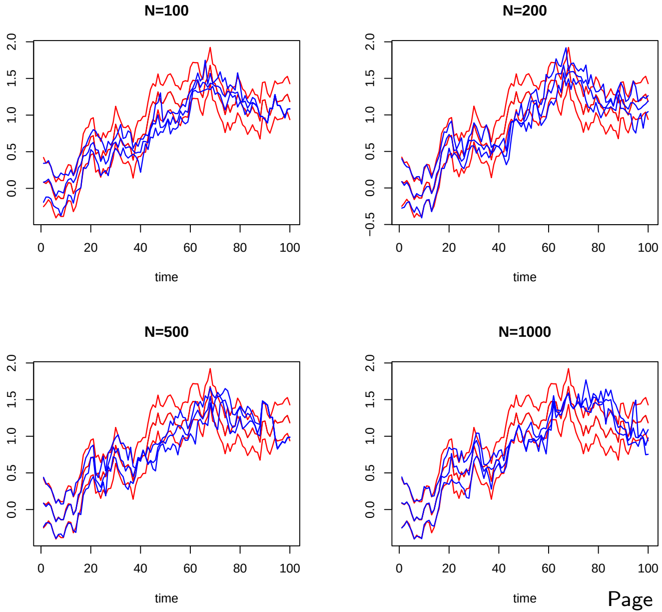
Sequential importance sampling (SIS) filter

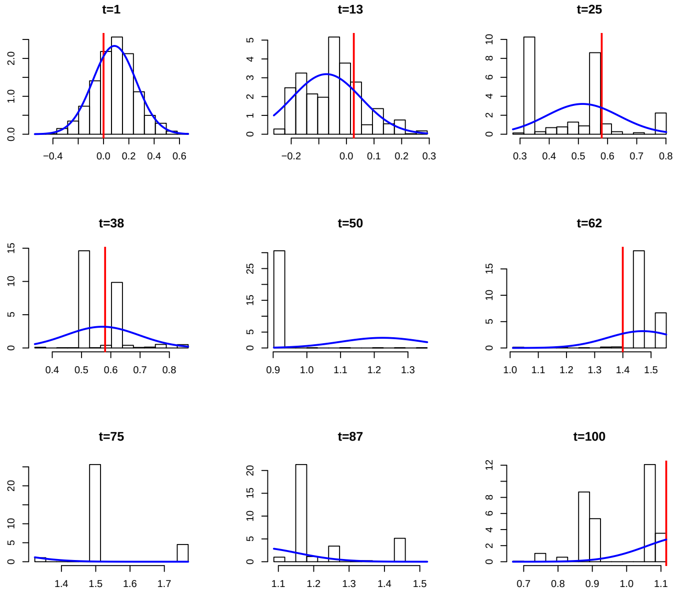
- For $i = 1, \ldots, N$, sample
$$\theta_1^{(i)} \sim N(m_0, C_0 + \tau^2)$$
so
$$\left\{(\theta_1, \omega_0)^{(i)}\right\}_{i=1}^{N} \sim p(\theta_1 | y_0)$$
where $\omega_0^{(i)} = 1/N$ for all i.
- Compute (unnormalized) weights
$$\omega_1^{(i)} = \omega_0^{(i)} p(y_1 | \theta_1^{(i)}) \qquad i = 1, \ldots, N$$
- SIS step:
$$\left\{(\theta_1, \omega_1)^{(i)}\right\}_{i=1}^{N} \sim p(\theta_1 | y_1)$$

For $t = 2, \ldots, n$

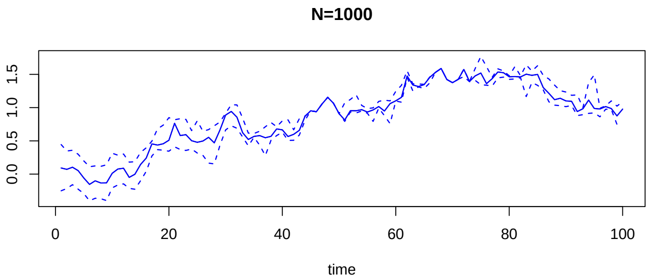
- Sample
$$\theta_t^{(i)} \sim N(\theta_{t-1}^{(i)}, \tau^2) \qquad i = 1, \ldots, N$$
- Compute (unnormalized) weights
$$\omega_t^{(i)} = \omega_{t-1}^{(i)} p(y_t|\theta_t^{(i)}) \qquad i = 1, \ldots, N$$
- SIS step:
$$\left\{(\theta_t, \omega_t)^{(i)}\right\}_{i=1}^{N} \sim p(\theta_t|y^t)$$

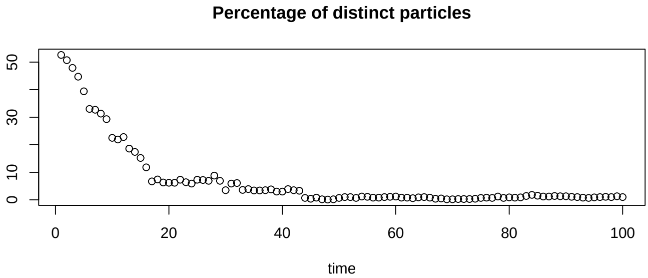
Sample impoverishment

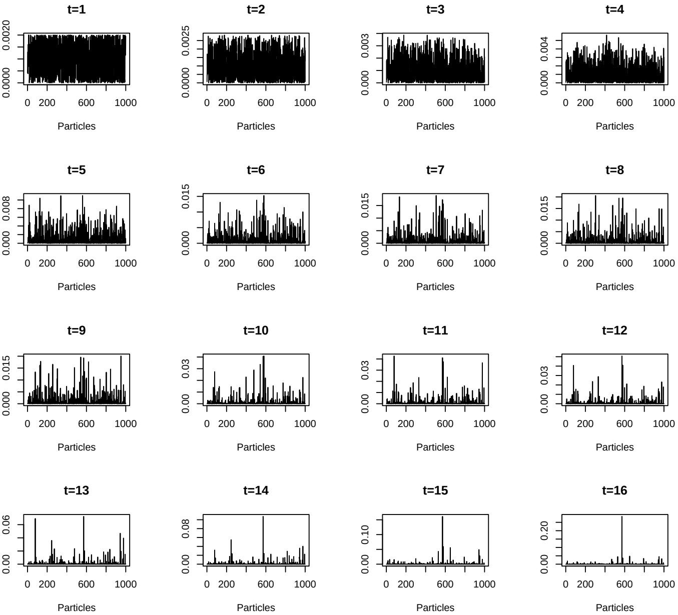




t=1
t=13
t=25
t=38
t=50
t=62
t=75
t=87
t=100


N=1000
1.5
1.0
0.5
0.0
0
20
40
60
80
100
time


Percentage of distinct particles
50
30
10
0
0
20
40
60
80
100
time


t=1
t=2
t=3
t=4
t=5
t=6
t=7
t=8
t=9
t=10
t=11
t=12
t=13
t=14
t=15
t=16
Particles

Sequential importance sampling with resampling (SISR) filter

- Sample

$$\tilde{\theta}_1^{(1)}, \ldots, \tilde{\theta}_1^{(N_0)} \sim p(\theta_1 | y_0) \equiv N(m_0, C_0 + \tau^2)$$

- Compute (unnormalized) weights

$$\omega_1^{(i)} = p(y_1 | \tilde{\theta}_1^{(i)}) \qquad i = 1, \ldots, N_0$$

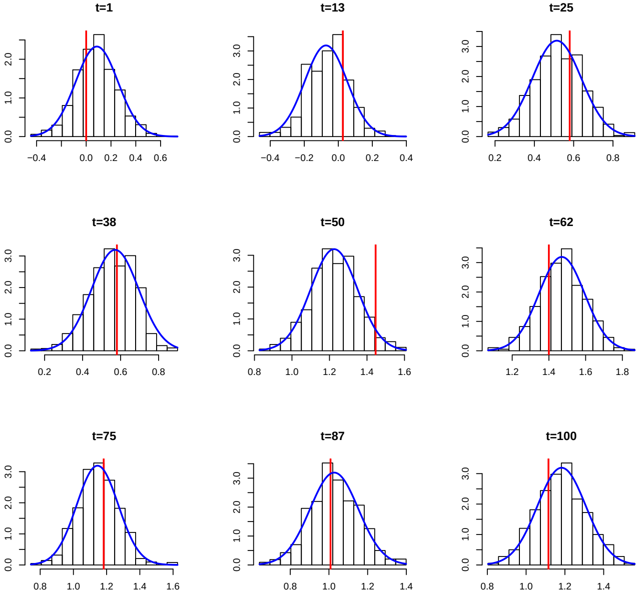
- Use SIR to obtain

$$\theta_1^{(1)}, \ldots, \theta_1^{(N_1)} \sim p(\theta_1 | y_1)$$

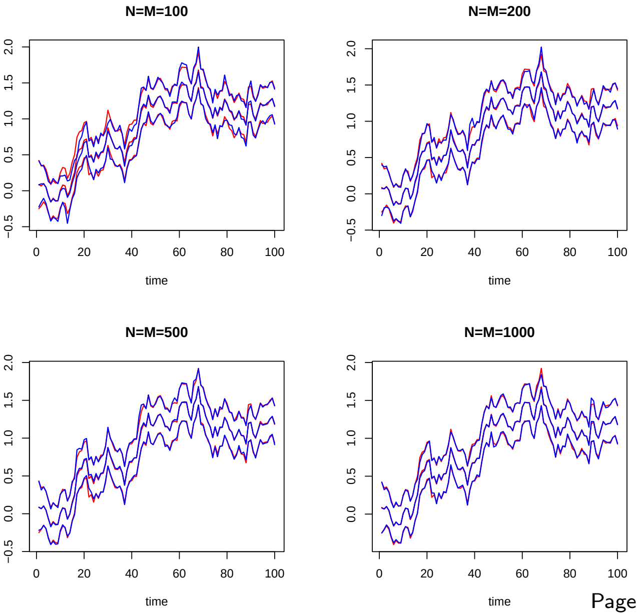
For $t = 2, \ldots, n$

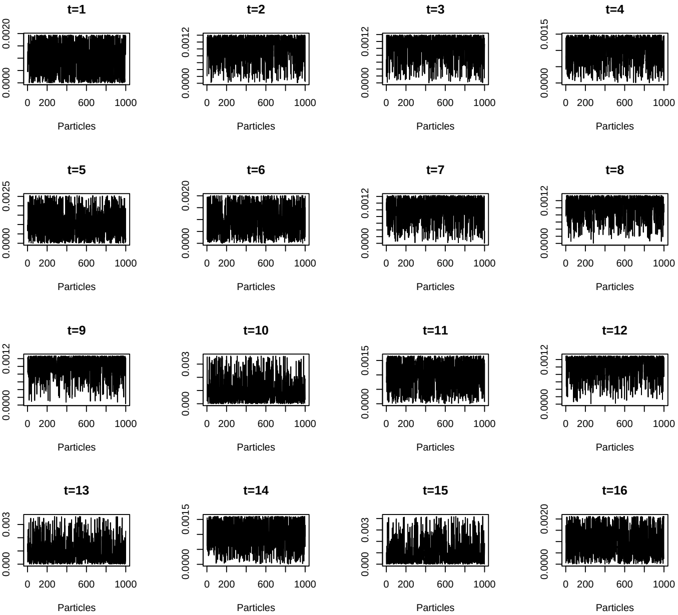
- Sample
$$\tilde{\theta}_t^{(i)} \sim N(\theta_{t-1}^{(i)}, \tau^2) \qquad i = 1, \ldots, N_t$$
- Compute (unnormalized) weights
$$\omega_t^{(i)} = p(y_t | \tilde{\theta}_t^{(i)})$$
- Use SIR to obtain
$$\theta_t^{(1)}, \ldots, \theta_t^{(N_t)} \sim p(\theta_t | y^t)$$

Example i (cont.) $p(\theta_t|y^t)$

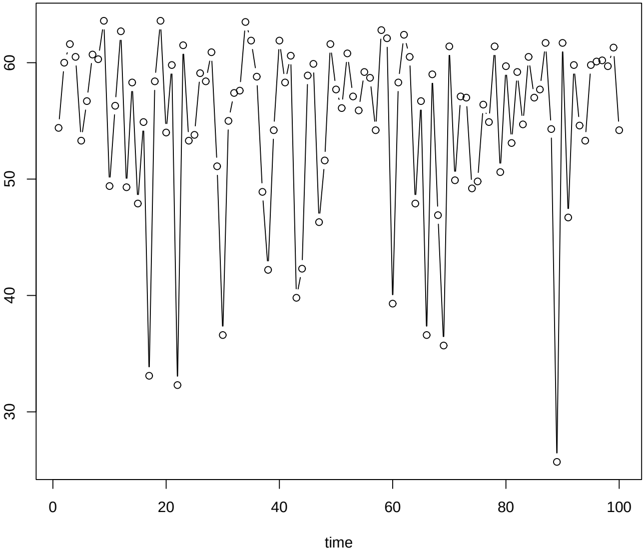


$p(\theta_t|y^t)$




t=1
t=2
t=3
t=4
t=5
t=6
t=7
t=8
t=9
t=10
t=11
t=12
t=13
t=14
t=15
t=16
Particles
0 200 600 1000
0.0000 0.0020
0.0000 0.0012
0.0000 0.0015
0.0000 0.0025
0.000 0.003

Percentage of distinct particles



Auxiliary particle (AP) filter

- $t-1$: posterior draws

$$\left\{(\theta_{t-1}, \omega_{t-1})^{(i)}\right\}_{i=1}^{N} \sim p(\theta_{t-1}|y^{t-1})$$

- t: prior approximation

$$\hat{p}(\theta_t|y^{t-1}) = \sum_{i=1}^{N} p(\theta_t|\theta_{t-1}^{(i)})\omega_{t-1}^{(i)}$$

- t: posterior approximation

$$\hat{p}(\theta_t|y^{t}) = \sum_{i=1}^{N} p(y_t|\theta_t)p(\theta_t|\theta_{t-1}^{(i)})\omega_{t-1}^{(i)}$$

IDEA: Jointly sample θ_t and a latent indicator, k, so that

$$\hat{p}(\theta_t, k|y^t) = p(y_t|\theta_t)p(\theta_t|\theta_{t-1}^{(k)})\omega_{t-1}^{(k)}$$

is the new target distribution.

A possible proposal distribution is

$$q(\theta_t, k|y^t) = p(y_t|\mu_t)p(\theta_t|\theta_{t-1}^{(k)})\omega_{t-1}^{(k)}$$

where, for instance, $\mu_t = E(\theta_t|\theta_{t-1})$.

- Posterior at $t-1$

$$\left\{(\theta_{t-1}, \omega_{t-1})^{(i)}\right\}_{i=1}^{N} \sim p(\theta_{t-1}|y^{t-1})$$

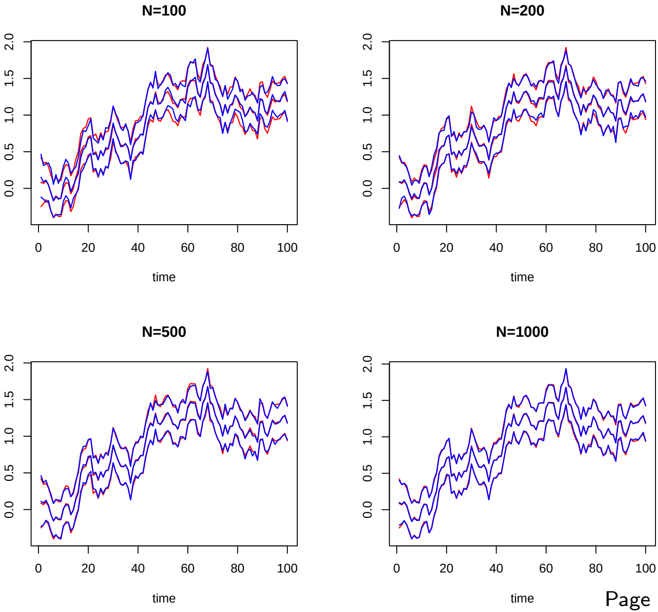
- For $i = 1, \ldots, N$
  - Sample $k \in \{1, \ldots, N\}$ with weight $p(y_t|\mu_t^{(i)})\omega_{t-1}^{(i)}$.
  - Sample $\theta_t^{(i)}$ from $p(\theta_t|\theta_{t-1}^{(k)})$.
  - Compute $\omega_t^{(i)} \propto p(y_t|\theta_t^{(i)})/p(y_t|\mu_t^{(k)})$.

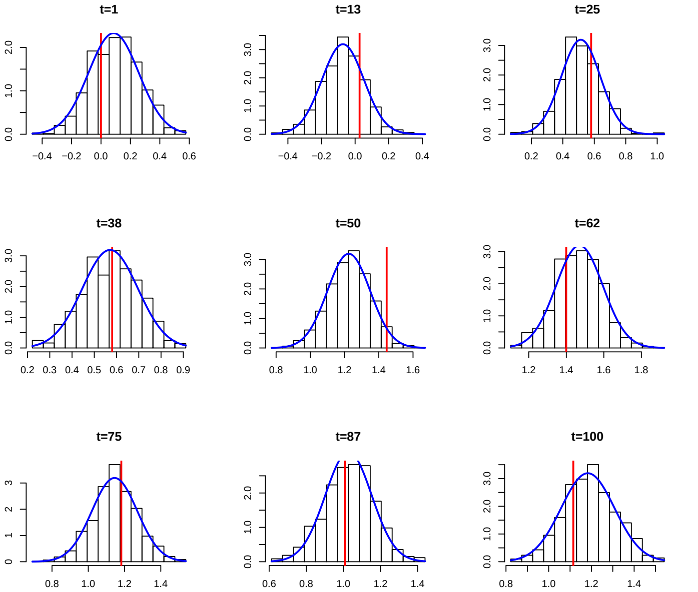
- Posterior at t

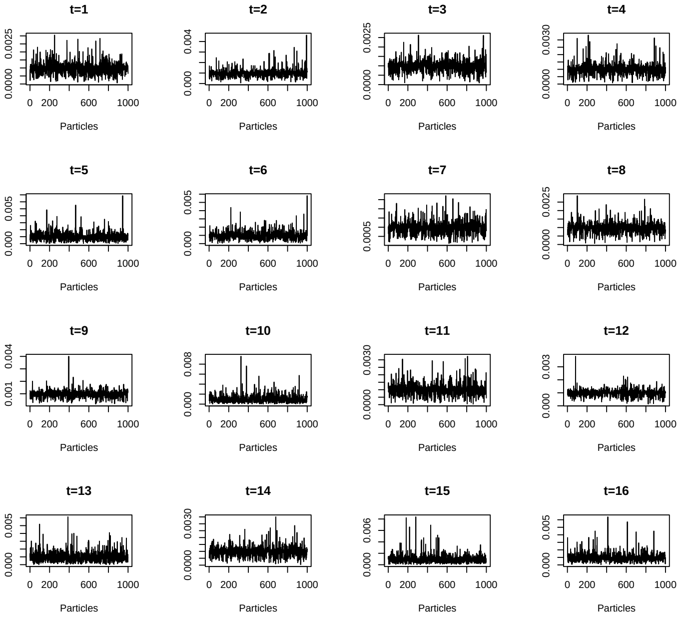
$$\left\{(\theta_t, \omega_t)^{(i)}\right\}_{i=1}^{N} \sim p(\theta_t|y^t)$$

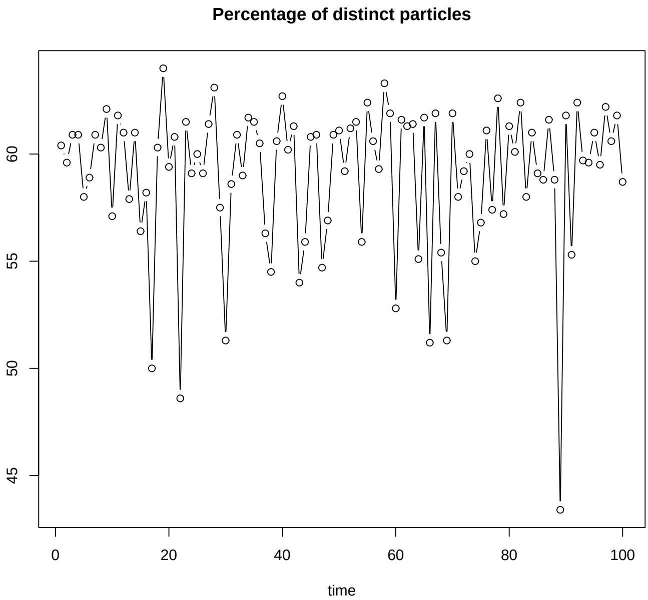
- Maybe add a SIR step to replenish θ_ts.

Example i (cont.)




t=1
t=13
t=25
t=38
t=50
t=62
t=75
t=87
t=100


t=1
0.0000
0.0025
0 200 600 1000
Particles
t=2
0.000
0.004
0 200 600 1000
Particles
t=3
0.0000
0.0025
0 200 600 1000
Particles
t=4
0.0000
0.0030
0 200 600 1000
Particles
t=5
0.000
0.005
0 200 600 1000
Particles
t=6
0.000
0.005
0 200 600 1000
Particles
t=7
0.0005
0 200 600 1000
Particles
t=8
0.0000
0.0025
0 200 600 1000
Particles
t=9
0.001
0.004
0 200 600 1000
Particles
t=10
0.000
0.008
0 200 600 1000
Particles
t=11
0.0000
0.0030
0 200 600 1000
Particles
t=12
0.000
0.003
0 200 600 1000
Particles
t=13
0.000
0.005
0 200 600 1000
Particles
t=14
0.0000
0.0030
0 200 600 1000
Particles
t=15
0.000
0.006
0 200 600 1000
Particles
t=16
0.000
0.005
0 200 600 1000
Particles


Percentage of distinct particles
60
55
50
45
0
20
40
60
80
100
time

Liu and West (2001): APF + parameter learning

- Let ψ be the vector of time-invariant parameters, so

$$p(\theta_t, \psi | y^{t-1}) \propto p(y_t|\theta_t, \psi) p(\theta_t|\psi, y^{t-1}) p(\psi|y^{t-1})$$

- As before

$$\hat{p}(\theta_t|\psi, y^{t-1}) = \sum_{i=1}^{N} p(\theta_t|\psi, \theta_{t-1}^{(i)}) \omega_{t-1}^{(i)}$$

where

$$\left\{ (\theta_{t-1}, \psi_{t-1}, \omega_{t-1})^{(i)} \right\}_{i=1}^{N} \sim p(\theta_{t-1}, \psi | y^{t-1})$$

- Liu and West (2001) approximate $p(\psi|y^{t-1})$ by

$$\hat{p}(\psi|y^{t-1}) = \sum_{i=1}^{N} N(\psi; m_{t-1}^{(i)}, h^2 V_{t-1})\omega_{t-1}^{(i)}$$

  where
  - h is a smoothing factor
  - $a^2 = 1 - h^2$ (West, 1993a,b)
  - $m_{t-1}^{(i)} = a\psi_{t-1}^{(i)} + (1-a)\bar{\psi}_{t-1}$
  - $\bar{\psi}_{t-1} = \widehat{E}(\psi|y^{t-1}) = \sum_{i=1}^{N} \psi_{t-1}^{(i)}\omega_{t-1}^{(i)}$
  - $V_{t-1} = \widehat{V}(\psi|y^{t-1}) = \sum_{i=1}^{N}(\psi_{t-1}^{(i)} - \bar{\psi}_{t-1})(\psi_{t-1}^{(i)} - \bar{\psi}_{t-1})'\omega_{t-1}^{(i)}$
- They suggest setting

$$h = 1 - \left(\frac{3\delta - 1}{2\delta}\right)^2$$

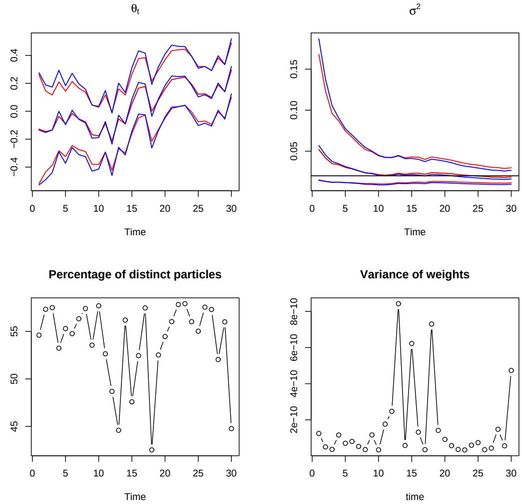
  where δ is Gordon, Salmond and Smith's (1993) artificial evolution discount.

Algorithm

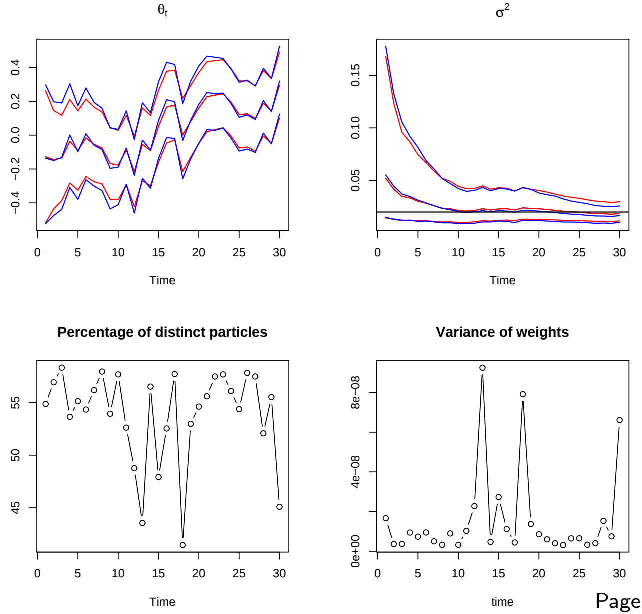
- At time $t-1$
  - $\{(\theta_{t-1}, \psi_{t-1}, \omega_{t-1})^{(i)}\}_{i=1}^{N} \sim p(\theta_{t-1}, \psi | y^{t-1})$
  - Compute $m_{t-1}^{(i)}$
  - Compute $\bar{\psi}_{t-1}$ and V_{t-1}
  - Compute $\mu_t^{(i)} = E(\theta_t | \theta_{t-1}^{(i)}, \psi_{t-1}^{(i)})$
- Repeat the following 4 steps N times
  - Sample k such that $Pr(k = i) \propto p(y_t | \mu_t^{(i)}, m_{t-1}^{(i)}) \omega_{t-1}^{(i)}$
  - Sample $\psi_t \sim N(m_{t-1}^{(k)}, h^2 V_{t-1})$
  - Sample $\theta_t \sim p(\theta_t | \theta_{t-1}^{(k)}, \psi_t)$
  - Compute weight $\omega_t \propto p(y_t | \theta_t, \psi_t) / p(y_t | \mu_t^{(k)}, m_{t-1}^{(k)})$
- $\{(\theta_t, \psi_t, \omega_t)^{(i)}\}_{i=1}^{N} \sim p(\theta_t, \psi | y^t)$
- Maybe add a SIR step to replenish θ_ts and ψ_ts.

Example i (cont.) $N = 100,000$ and $\delta = 0.9$

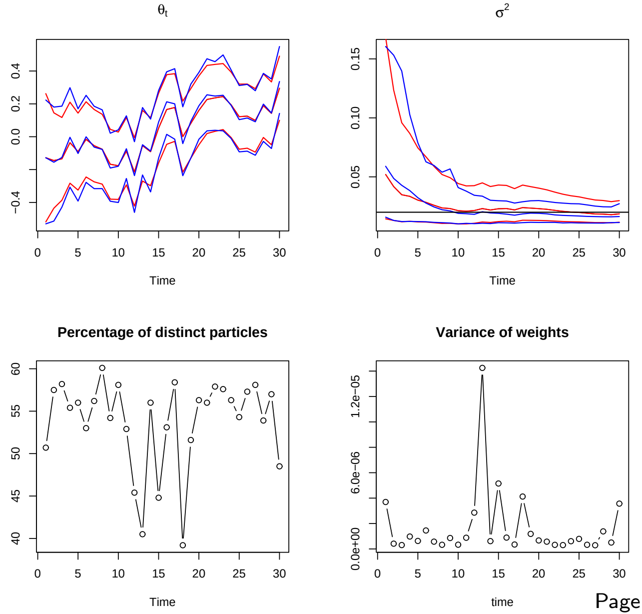
- Simulation: $n = 30$, $\tau^2 = 0.01$ and $\sigma^2 = 0.02$.
- Prior: $\theta_0 \sim N(0.0, 0.1)$ and $\sigma^2 \sim IG(5/2, 5 \times 0.04/2)$.
- SMC: $N = 10^5$, $\delta = 0.9$, $h = 0.1080247$ and $a = 0.9941482$.



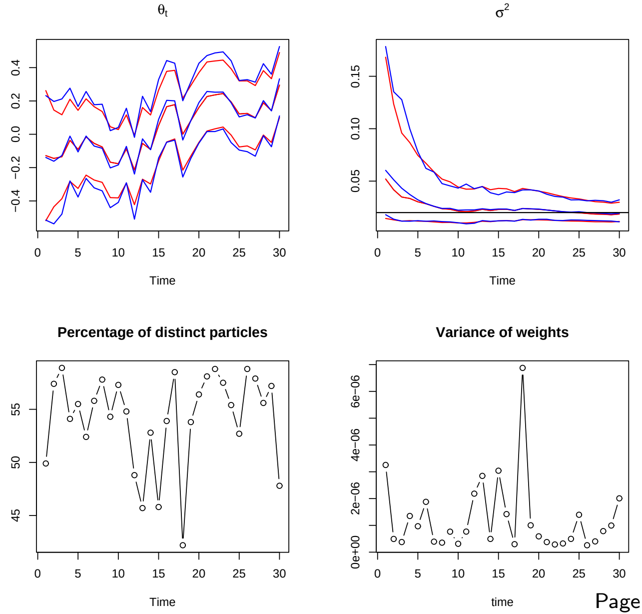
$N = 10,000$ and $\delta = 0.9$



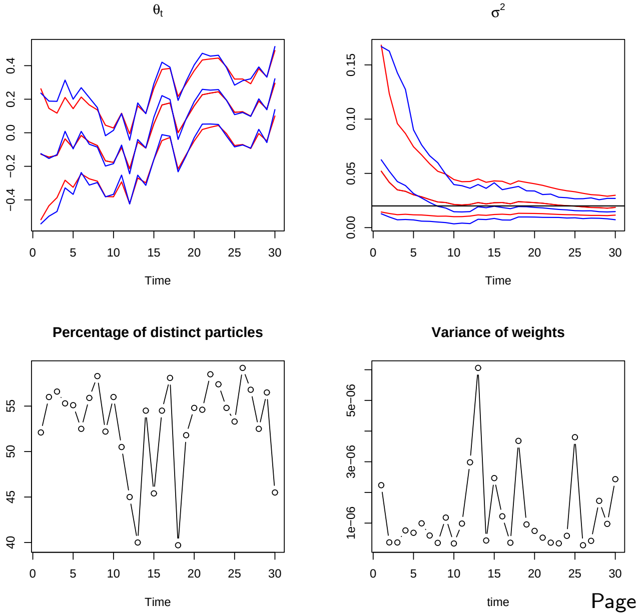
$N = 1,000$ and $\delta = 0.9$



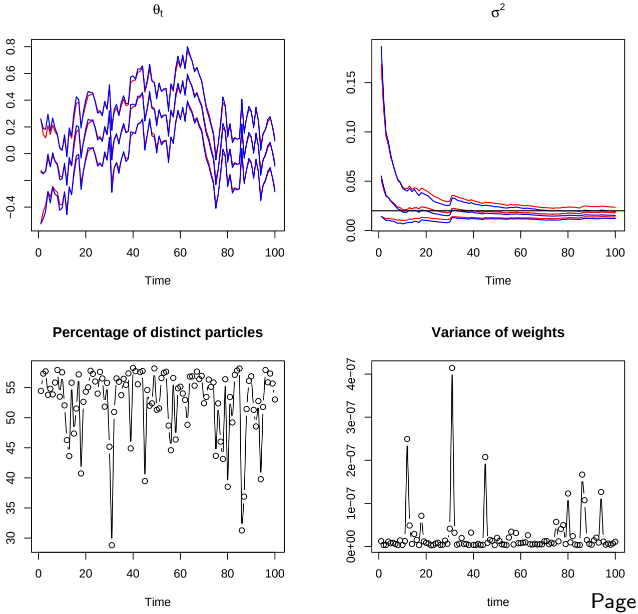
$N = 1,000$, $\delta = 0.75$, $h = 0.3055556$ and $a = 0.9521743$.



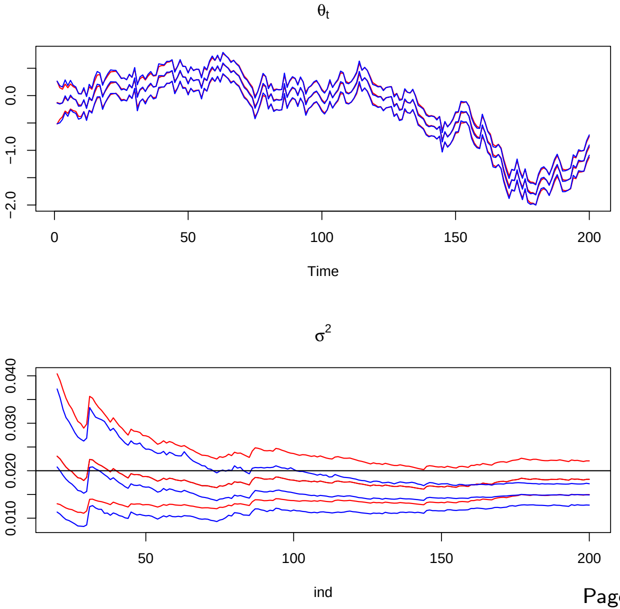
$N = 1,000$, $\delta = 0.5$, $h = 0.75$ and $a = 0.6614378$.



$n = 100$, $N = 10,000$ and $\delta = 0.75$



$n = 200$, $N = 10,000$ and $\delta = 0.75$



Example ii: Stochastic volatility

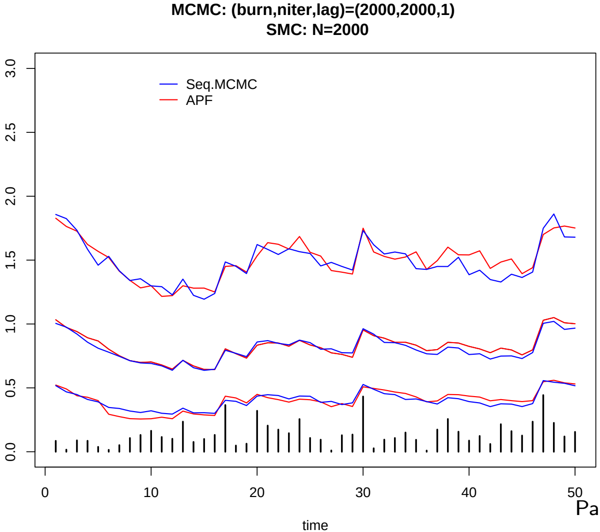
Let y_t, for $t = 1, \ldots, n$ be modeled as

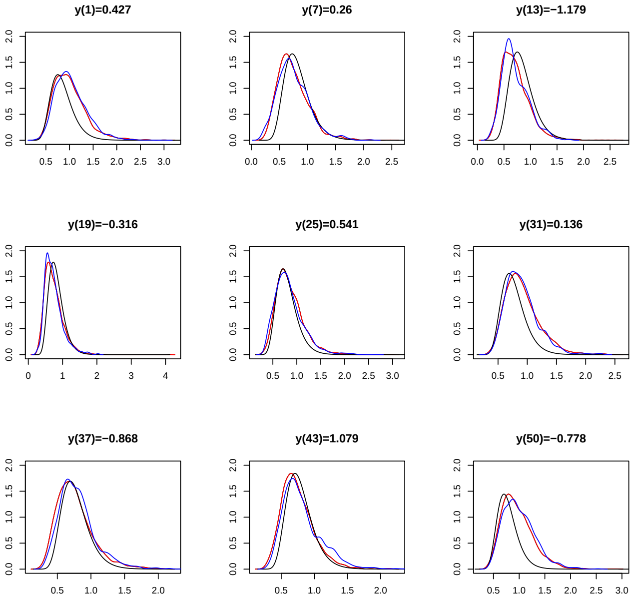
$$
\begin{aligned}
y_t|\theta_t &\sim N(0, e^{\theta_t}) \\
(\theta_t|\theta_{t-1}, \xi) &\sim N(\alpha + \beta\theta_{t-1}, \tau^2)
\end{aligned}
$$

where $\xi = (\alpha, \phi, \tau^2)$ and $\theta_0 \sim N(m_0, V_0)$.

We simulated $n = 50$ observations based on $\alpha = -0.0031$, $\beta = 0.9951$, $\tau^2 = 0.0074$, with $m_0 = 0.0$ and $V_0 = 0.1$. Also, $\theta_1 = \alpha/(1-\beta) = -0.632653$, which corresponds to annualized standard deviations around 13%.

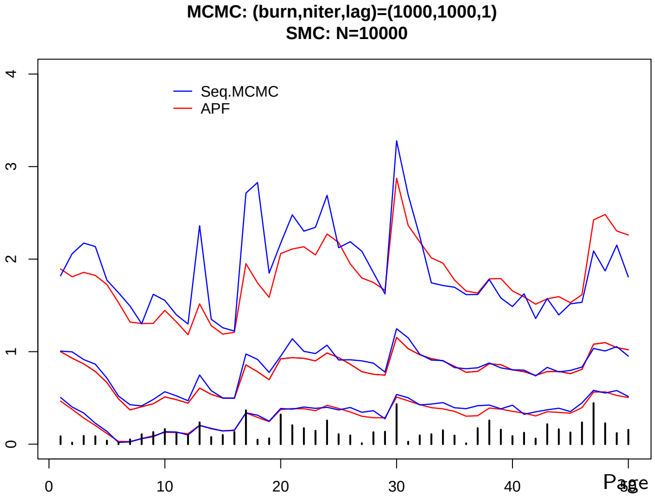
- **Seq.MCMC:** brute force MCMC after each y_t is observed (Kim, Shephard and Chib, 1994).
- **APF:** Auxiliary particle filter
- In both cases α, β and τ^2 are kept fixed.

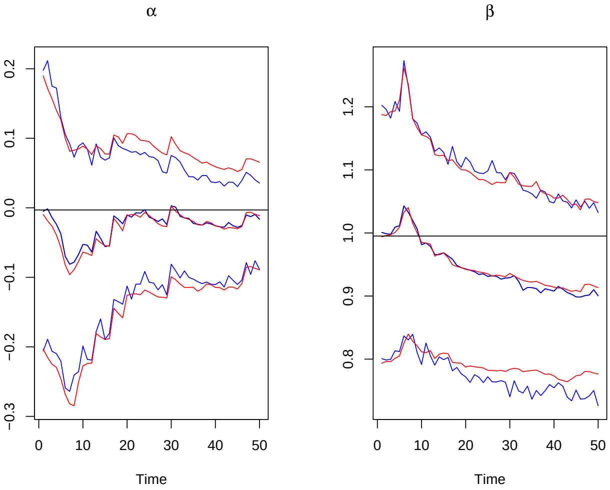



y(1)=0.427
y(7)=0.26
y(13)=−1.179
y(19)=−0.316
y(25)=0.541
y(31)=0.136
y(37)=−0.868
y(43)=1.079
y(50)=−0.778

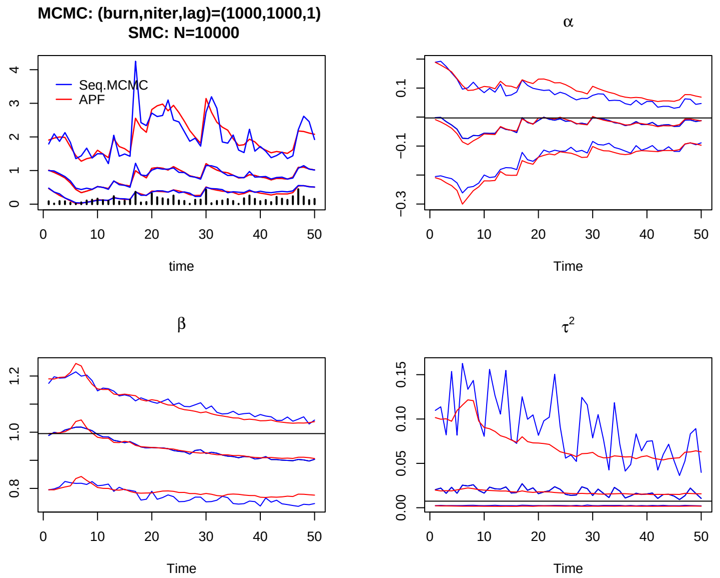
Learning α and β

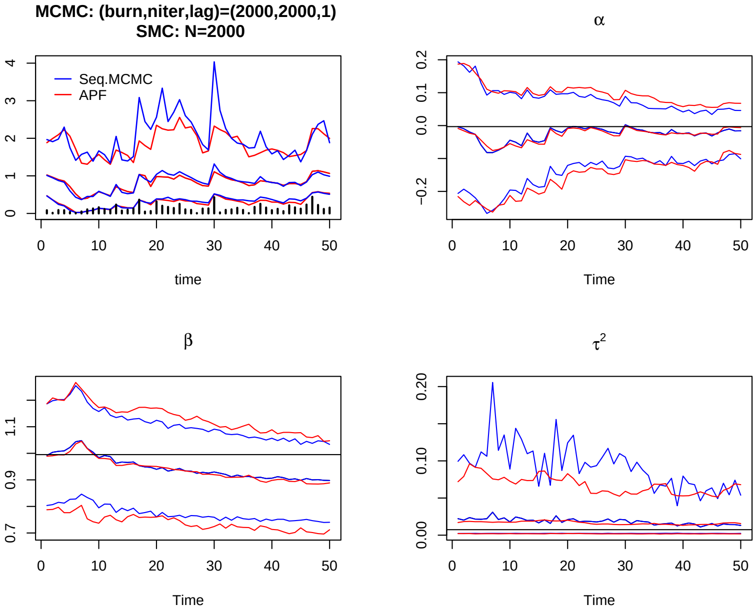
- $\delta = 0.75$, $h = 0.3055556$ and $a = 0.9521743$
- $\theta_0 \sim N(0, 0.1)$, $\alpha \sim N(-0.0031, 0.01)$ and $\phi \sim N(0.9951, 0.01)$

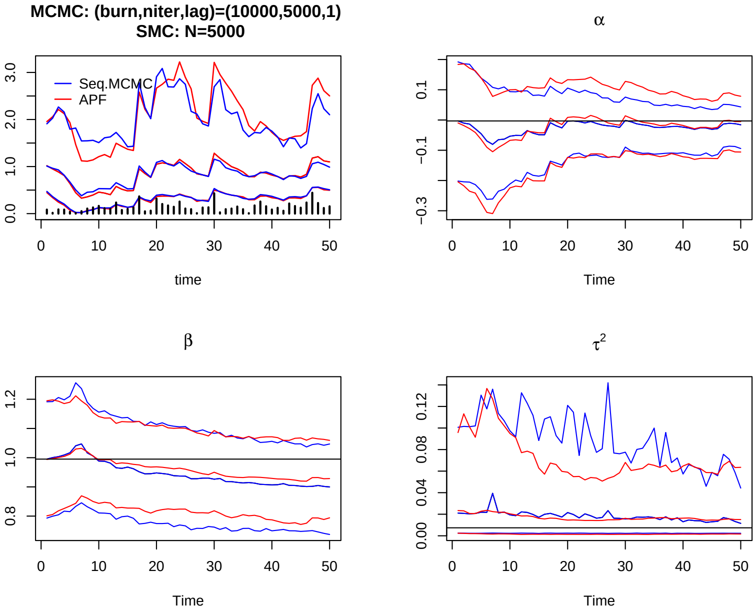


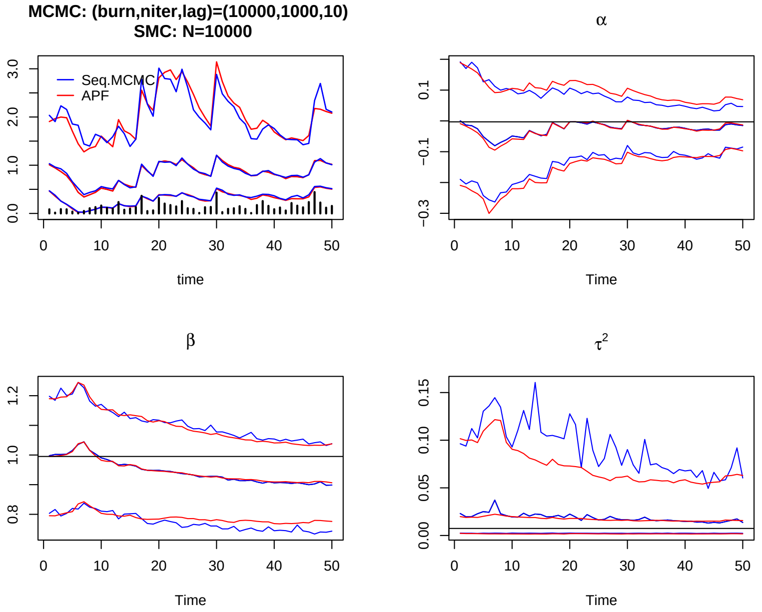

α
β
0.2
0.1
0.0
−0.1
−0.2
−0.3
1.2
1.1
1.0
0.9
0.8
0
10
20
30
40
50
Time
Time

Learning α, β and τ^2, with $\tau^2 \sim IG(1.5, 0.0111)$

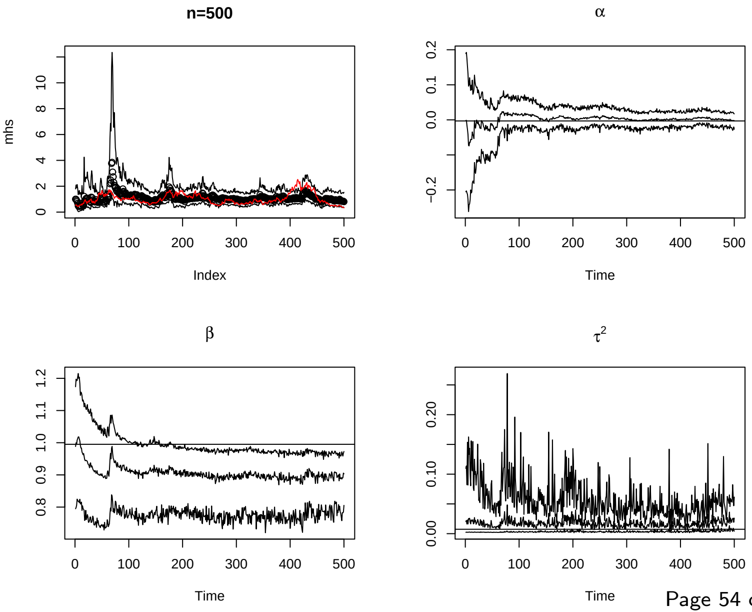



MCMC: (burn,niter,lag)=(2000,2000,1)
SMC: N=2000
Seq.MCMC
APF
time
α
Time
β
Time
τ²
Time


MCMC: (burn,niter,lag)=(10000,5000,1)
SMC: N=5000
Seq.MCMC
APF
time
α
Time
β
Time
τ2
Time


MCMC: (burn,niter,lag)=(10000,1000,10)
SMC: N=10000
Seq.MCMC
APF
time
α
Time
β
Time
τ2
Time

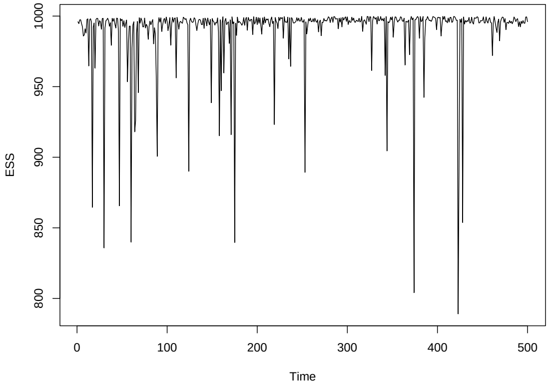
$n = 500$: MCMC



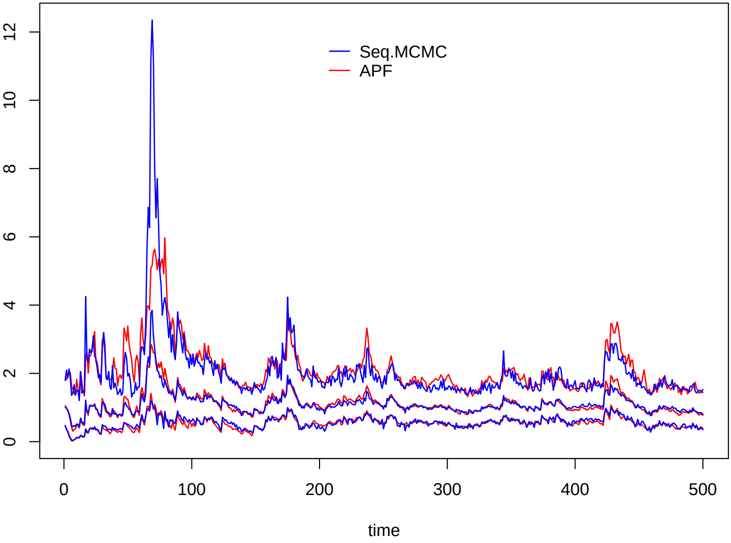
n=500
mhs
0
2
4
6
8
10
0
100
200
300
400
500
Index
α
−0.2
0.0
0.1
0.2
Time
β
0.8
0.9
1.0
1.1
1.2
Time
τ2
0.00
0.10
0.20
Time

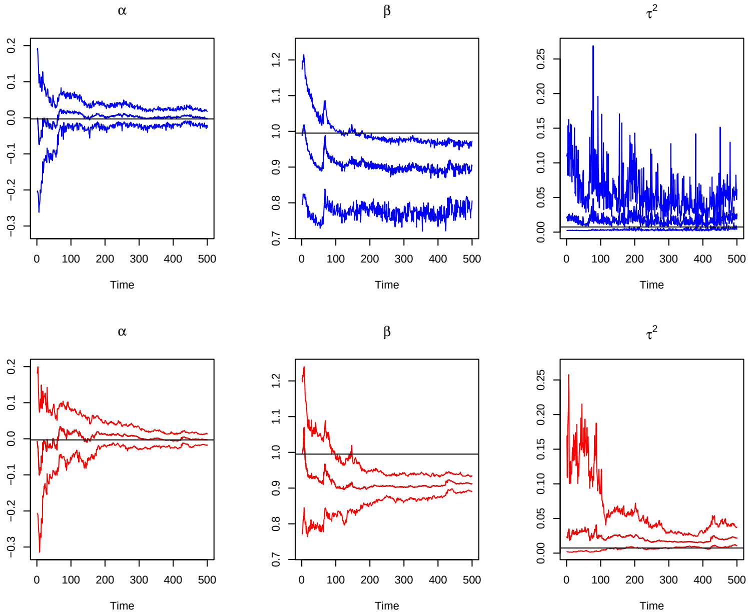
Effective sample size (ESS)

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



MCMC: (burn,niter,lag)=(1000,1000,1)
SMC: N=1000




α
β
τ²
α
β
τ²
Time

Example iii: Nonlinear model

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

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

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

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

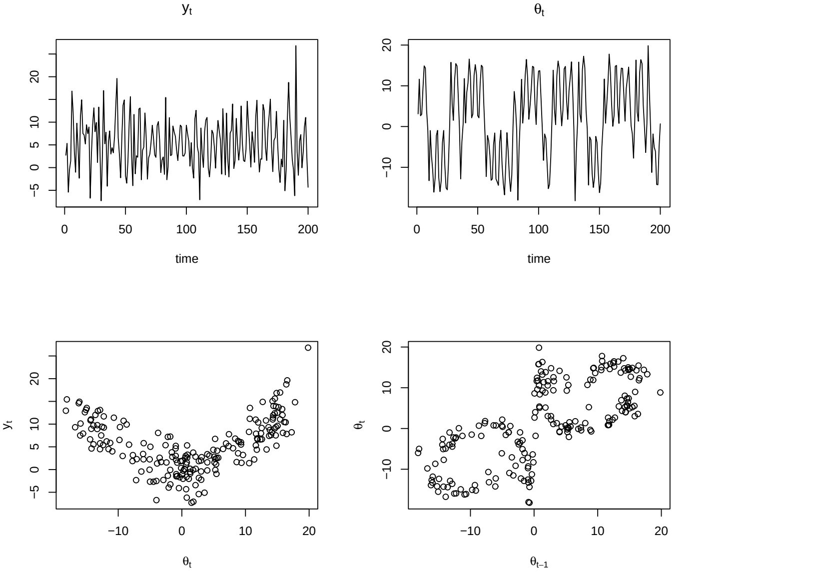
$$\begin{aligned} \theta_0 &\sim N(m_0, V_0) \\ \xi &\sim N(c_0, C_0) \\ \sigma^2 &\sim IG(a_0, A_0) \\ \tau^2 &\sim IG(b_0, B_0) \end{aligned}$$

Simulation set up

We simulated $n = 200$ observations based on $\xi = (0.5, 25, 8)'$, $\sigma^2 = 10$, $\tau^2 = 1$ and $\theta_0 = 0.1$.

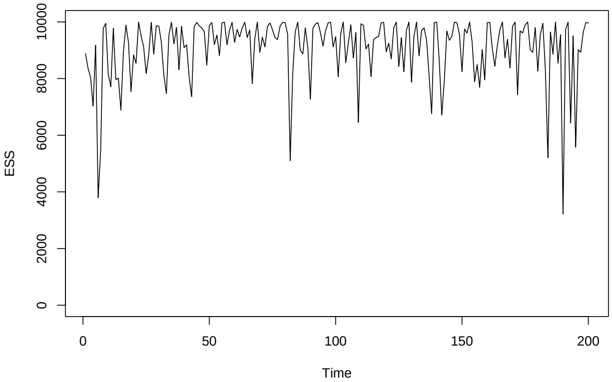
Prior hyperparameters:

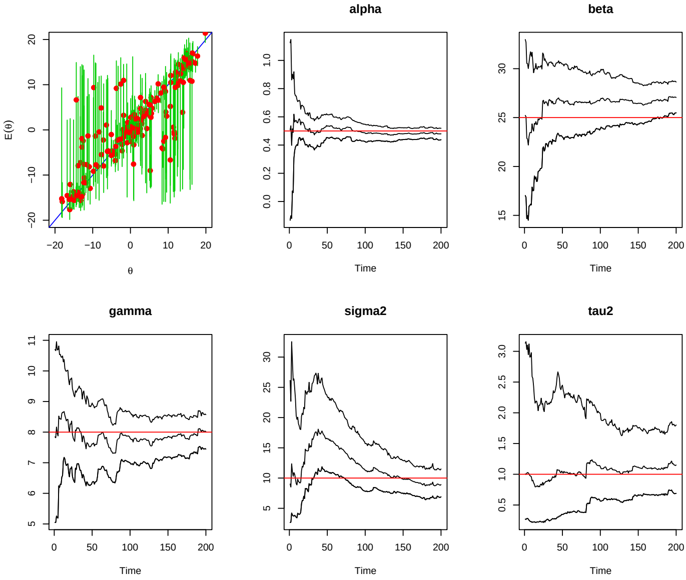
$$
\begin{aligned}
m_0 &= 0.0 \quad \text{and} \quad V_0 = 5 \\
c_0 &= (0.5, 25, 8)' \quad \text{and} \quad C_0 = \text{diag}(0.1, 16, 2) \\
a_0 &= 3 \quad \text{and} \quad A_0 = 20 \\
b_0 &= 3 \quad \text{and} \quad B_0 = 2
\end{aligned}
$$


y_t
θ_t
time
time
y_t
θ_t
θ_{t-1}
θ_t

Effective sample size

$$(N, \delta, h, a) = (10000, 0.75, 0.31, 0.95)$$




E(θ)
θ
alpha
beta
gamma
sigma2
tau2
Time

Example iv: Markov switching stochastic volatility (Carvalho and Lopes, 2007)

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

- *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 $N(m_t^{(k^l)}, b^2 V_t)$
  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

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

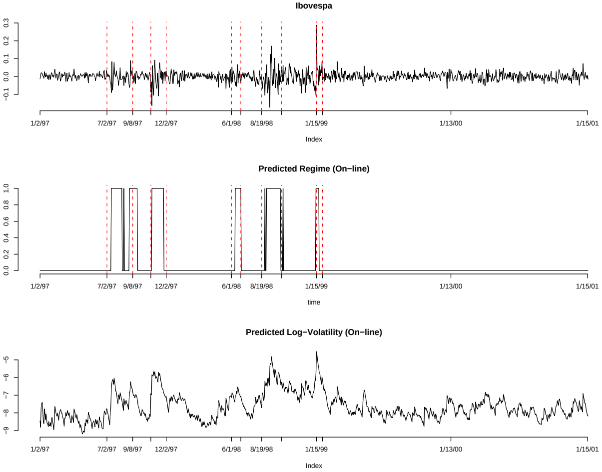
- *Step 4:* $\left\{\lambda_{t+1}^{(j)}, S_{t+1}^{(j)}, w_{t+1}^{(j)}\right\}_{j=1}^{M} \sim p(\lambda_{t+1}, S_{t+1}, \theta | D_{t+1})$.
- *Step 5:* Resample

Data from 01/02/1997 to 01/16/2001 (1000 observations).

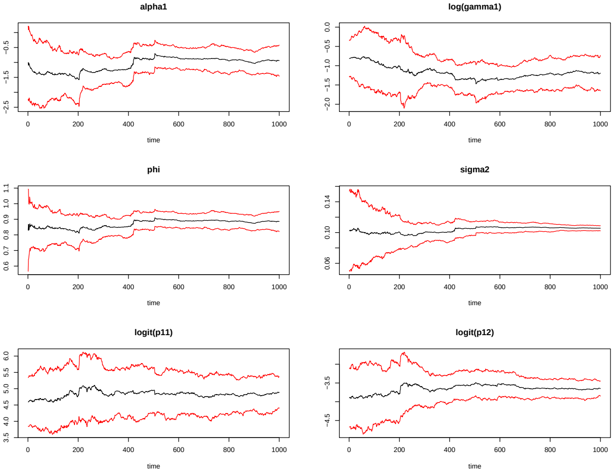
07/02/1997	Thailand devalues the baht by as much as 20%.
08/11/1997	IMF and Thailand set a rescue agreement.
10/23/1997	Hong Kong's stock index falls 10.4%. South Korea Won starts to weaken.
12/02/1997	IMF and South Korea set a bailout agreement.
06/01/1998	Russia's stock market crashes.
06/20/1998	IMF gives final approval to a loan package to Russia.
08/19/1998	Russia officially falls into default.
10/09/1998	IMF and World Bank joint meeting to discuss the global economic crisis.
	The Fed cuts interest rates.
01/15/1999	The Brazilian government allows its currency, the real,
	to float freely by lifting exchange controls.
02/02/1999	Arminio Fraga is named president of Brazil's Central Bank.

Table: Currency crisis – Some key dates

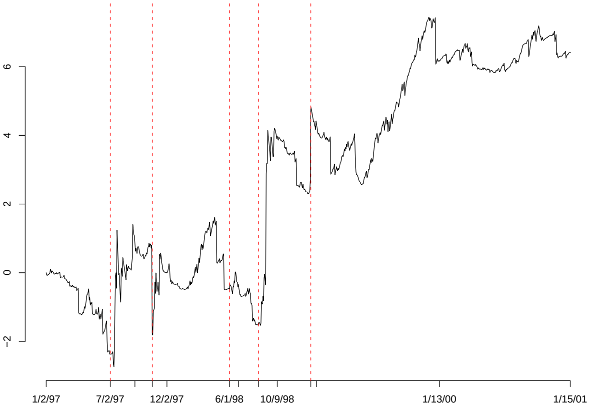
Ibovespa index (*top*), estimated (posterior mean) s_t (*center*), and λ_t (*bottom*). The vertical lines indicate key market events that identify what agents in the markets would refer to as the beginning and end of the crisis



Posterior mean, 5% and 95% quantiles of θ for the Ibovespa data $\alpha_1, \log(\gamma_1), \phi, \sigma^2, \log(\frac{p_{11}}{1-p_{11}})$, and $\log(\frac{p_{21}}{1-p_{21}})$.



Ibovespa: Bayes Factor – MSSV *vs.* SV.



Example v: S&P500 and NDX100 volatility: Credit Crisis of 2007 (Lopes and Polson, 2008)

- Description of the different volatility indices and how they are related to option prices and the market price of volatility risk.

- Daily return data for the Standard and Poor's SP500 (SPY) stock index and the Nasdaq 100 (QQQQ) index.

- The corresponding option volatility indices are the VIX and the VXN; again these are available on a continuous basis.

- Description of the three competing models that we use for price dynamics: pure stochastic volatility (SV), SV with jumps (SVJ) and Garch(1,1).

- Parameter estimates will be determined from a longer historical period for 2002-2006.

- Our particle filtering approach will then be implemented for the year of 2007.

SV model

A common approach to modeling asset prices, is to assume that the equity index S_t and its stochastic variance V_t jointly solve

$$
\begin{aligned}
\frac{dS_t}{S_t} &= \sqrt{V_{t+1}}dB_t \\
d\log V_t &= \kappa_v(\theta_v - \log V_t) + \sigma_v dB_t^V
\end{aligned}
$$

To analyze this model in an online fashion it is common to using an Euler discretization of the above model for continuously compounded returns $Y_{t+1} = \log\left(S_{t+1}/S_t\right)$

$$
\begin{aligned}
Y_{t+1} &= \sqrt{V_{t+1}}\varepsilon_{t+1} \\
X_{t+1} &= \alpha_v + \beta_v X_t + \sigma_v \eta_{t+1}
\end{aligned}
$$

Here $X_t = \log V_t$ and ε_{t+1} and η_{t+1} are normally distributed, serially and contemporaneously independent shocks and we define $\alpha = \kappa_v\theta_v$ and $\beta_v = 1 - \kappa_v$.

SVJ model

We assume that an equity index S_t and its stochastic variance V_t jointly solve

$$
\begin{aligned}
\frac{dS_t}{S_t} &= \sqrt{V_{t+1}}dB_t + d\left(\sum_{s=N_t}^{N_{t+1}} Z_s\right) \\
d\log V_t &= \kappa_v(\theta_v - \log V_t) + \sigma_v dB_t^V
\end{aligned}
$$

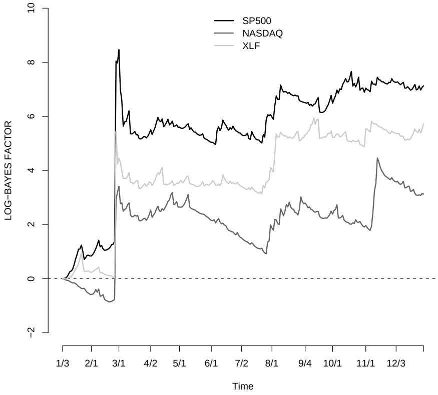
where the additional term describes the jump process.

Using an Euler discretization of this continuous time process leads to the log stochastic volatility jump (SVJ) model

$$
\begin{aligned}
Y_{t+1} &= \sqrt{V_{t+1}}\varepsilon_{t+1} + J_{t+1}Z_{t+1} \\
X_{t+1} &= \alpha_v + \beta_v X_t + \sigma_v \eta_{t+1} \\
J_{t+1} &\sim Ber(\lambda) \\
Z_{t+1} &\sim N(\mu_z, \sigma_z^2)
\end{aligned}
$$

where $Y_{t+1} = \log\left(S_{t+1}/S_t\right)$ are log-returns with ε_{t+1} and η_{t+1} normally distributed, serially and contemporaneously independent shocks.

Sequential (log) Bayes factor, $BF(M_1, M_0)$: $M_1 \equiv$SVJ model
$M_0 \equiv$SV model.



Perfect adaptation & smoothing

- Reversing Bayes rule (resample-propagate)

$$
\begin{aligned}
p(\theta_t|y^{t+1}) &\propto p(y_{t+1}|\theta_t)p(\theta_t|y^t) \\
p(\theta_{t+1}|y^{t+1}) &= \int p(\theta_{t+1}|\theta_t, y_{t+1})p(\theta_t|y^{t+1})d\theta_t
\end{aligned}
$$

- Godsill, Doucet, West (2004)

1. Choose $\tilde{\theta}_T = \theta_T^{(i)}$ w.p. $\omega_T^{(i)}$
2. For $t = T-1, \ldots, 1$:
   - Calculate $\omega_{t|t+1}^{(i)} \propto \omega_t^{(i)} p(\tilde{\theta}_{t+1}|\theta_t^{(i)}) \;\; \forall i$
   - Choose $\tilde{\theta}_t = \theta_t^{(i)}$ w.p. $\omega_{t|t+1}^{(i)}$.
3. $\tilde{\theta}_{1:T} = (\tilde{\theta}_1, \ldots, \tilde{\theta}_T) \sim p(\theta_{1:T}|y^T)$.

Don't miss Carlos Carvalho's talk, Tuesday morning!

Cited references

Carvalho, C. and Lopes, H. (2007) Simulation-based sequential analysis of Markov switching stochastic volatility models. *Computational Statistics and Data Analysis*, **51**, 4526-42.

Godsill, S. J., Doucet, A. and West, M. (2004) Monte Carlo smoothing for nonlinear time series. *Journal of the American Statistical Association*, **99**, 156-168.

Gordon, N., Salmond, D., Smith, A. (1993). Novel approach to nonlinear/non-Gaussian Bayesian state estimation. *IEE Proc. F*, **140**, 10713.

Liu, J. and West, M. (2001) Combined parameter and state estimation in simulation based filtering. In *Sequential Monte Carlo in Practice* (eds N. de Freitas, N. Gordon and A. Doucet). Springer-Verlag, Chapter 10, 197-223.

Lopes, H. F. and Polson, N. (2008) Extracting S&P500 and NDX100 volatility: Credit Crisis of 2007 *The Handbook of Applied Bayesian Analysis*, Eds. Tony O'Hagan & Mike West. Oxford University Press

Migon, H., Gamerman, D., Lopes, H. and Ferreira, M. (2005) Dynamic models, In Dey, D. and Rao, C.R. (Eds.) *Handbook of Statistics, Volume 25: Bayesian Thinking, Modeling and Computation*, Chapter 19, 553-88.

Smith, A. F. M. and Gelfand, A. E. (1992) Bayesian statistics without tears: a sampling-resampling perspective. *American Statistician*, **46**, 84-8.

West, M. (1993a). Approximating posterior distributions by mixtures, *Journal of the Royal Statistical Society, Series B*, **55**, 409-22.

West M. (1993b). Mixtures models, Monte Carlo, Bayesian updating and dynamic models. *Computer Science and Statistics*, **24**, 325333.

Important additional (more recent) reading

Sequential Monte Carlo Methods: `http://www-sigproc.eng.cam.ac.uk/smc/`

Doucet, A., Godsill, S. and Andrieu, C. (2000). On sequential Monte Carlo sampling methods for Bayesian filtering. *Statistics and Computing*, **10**, 197208.

Doucet, A., de Freitas, N. and Gordon, N. (2001) *Sequential Monte Carlo Methods in Practice*, Springer, New York.

Del Moral, P. Jasra, A. and Doucet, A. (2006) Sequential Monte Carlo for Bayesian Computation (with discussion). *Bayesian Statistics 8*, Oxford University Press.

Johansen, A. M. and Doucet, A. A Note on Auxiliary Particle Filters. *Statistics and Probability Letters*.

Fearnhead, P. (2002). Markov chain Monte Carlo, sufficient statistics, and particle filters. *Journal of Computational and Graphical Statistics*, **11**, 848-62.

Polson, N. G., Stroud, J. R. and Müller, P. (2008). Practical Filtering with Sequential Parameter Learning. *Journal of the Royal Statistical Society, Series B*, **70**, 413-428.

Storvik, G. (2002). Particle filters in state space models with the presence of unknown static parameters. *IEEE Trans. Signal Process*, **50**, 2819.

Workshop talks:

- **Carlos Carvalho, Particle Learning and Smoothing**
- **Eric Moulines, Theory of Sequential Monte Carlo**
- **Arnaud Doucet, Particle Markov Chain Monte Carlo**