MONTE CARLO METHODS and STOCHASTIC VOLATILITY

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

Dipartimento di Scienze delle Decisioni, Università Bocconi
Milano, Italy

Date and classroom	Topic
Mon, 11/23 16:30-18:00, N1-7 velodromo	Normal dynamic linear models
Tue, 11/24 10:30-12:00, SDA-01 via Bocconi	Nonnormal, nonlinear dynamic models
Tue, 11/24 16:30-18:00, N1-7 velodromo	Stochastic volatility models
Wed, 11/25 10:30-12:00, 4-C via Sarfatti	More on Stochastic volatility
Thu, 11/26 10:30-12:00, 4-1 via Sarfatti	Sequential Monte Carlo methods
Fri, 11/27 10:30-11:50, N1-2 velodromo	Particle learning (PL)
Fri, 11/27 12:10-13:30, N1-2 velodromo	More on PL

1st class: Normal dynamic linear models (DLMs)

- Bayesian regression: Close form solution when the prior is conjugate; SIR and Gibbs otherwise.
- DLM: Close form solution when quadruple F,G,V,W is known; FFBS otherwise.
- Basic references: Gamerman and Lopes (2006), West and Harrison (1997), Petris, Petrone and Campagnoli (2009) and Migon, Gamerman, Lopes and Ferreira (2005).

2nd class: Nonnormal, nonlinear dynamic models

- Nonlinear evolution equation, normal or Student's t errors.
- Stochastic volatility AR(1) model.
- Basic references: Carlin, Polson and Stoffer (1992), Jacquier, Polson and Rossi (1994), Kim, Shephard and Chib (1994).

3rd and 4th classes: Stochastic volatility (SV) models

- MCMC methods for Financial Econometrics.
- The Impact of Jumps and Volatility in Returns.
- SV models with jumps.
- Factor stochastic volatility models.
- Basic references: Johannes and Polson (2009), Eraker, Johannes and Polson (2003), Johannes, Polson and Stroud (2009), Lopes and Carvalho (2007), Carvalho and Lopes (2007), Aguilar and West (2000) and Pitt and Shephard (1999).

5th class: Sequential Monte Carlo methods

- Sequential importance sampling (SIS).
- SIS with resampling (SISR).
- Auxiliary particle filter (APF).
- APF + normal kernel density approximation.
- SISR + Sufficient statistics.
- Filters with MCMC updates.
- Basic references: Gordon, Salmond and Smith (1993), Pitt and Shephard (1999), Prado and West (2010), Liu and West (2001), Storvik (2002) and Polson, Stroud and Müller (2008).

6th and 7th classes: Particle learning (PL)

- PL in DLMs and CDLMs.
- PL for nonlinear normal dynamic models.
- PL in non-dynamic problems.
- Basic references: Carvalho, Johannes, Lopes and Polson (2008), Polson, Lopes and Carvalho (2009).

1st class: linear regression and dynamic linear models

Linear regression: Likelihood and prior

The Bayesian approach to the standard multiple linear regression is

$$(y|X, \beta, \sigma^2) \sim N(X\beta, \sigma^2 I_n)$$

where $y = (y_1, \ldots, y_n)$, $X = (x_1, \ldots, x_n)'$ is the $(n \times q)$, design matrix and $q = p + 1$.

The prior distribution of (β, σ^2) is $NIG(b_0, B_0, n_0, s_0^2)$, i.e.

$$\begin{aligned} \beta|\sigma^2 &\sim N(b_0, \sigma^2 B_0) \\ \sigma^2 &\sim IG(n_0/2, n_0 s_0^2/2) \end{aligned}$$

for known hyperparameters b_0, B_0, n_0 and s_0^2.

Linear regression: conditional and marginal posteriors

It can be shown that

$$
\begin{aligned}
(\beta|\sigma^2, y, X) &\sim N(b_1, \sigma^2 B_1) \\
(\sigma^2|\beta, y, X) &\sim IG(n_1/2, n_1 s_{11}^2/2)
\end{aligned}
$$

and

$$
\begin{aligned}
(\sigma^2|y, X) &\sim IG(n_1/2, n_1 s_1^2/2) \\
(\beta|y, X) &\sim t_{n_1}(b_1, s_1^2 B_1)
\end{aligned}
$$

where

$$
\begin{aligned}
n_1 &= n_0 + n \\
B_1^{-1} &= B_0^{-1} + X'X \\
B_1^{-1} b_1 &= B_0^{-1} b_0 + X'y \\
n_1 s_1^2 &= n_0 s_0^2 + (y - Xb_1)'y + (b_0 - b_1)' B_0^{-1} b_0 \\
n_1 s_{11}^2 &= n_0 s_0^2 + (y - X\beta)'(y - X\beta)
\end{aligned}
$$

Linear regression: OLS

The maximum likelihood estimators of β and σ^2 are

$$\hat{\beta} = (X'X)^{-1}X'y \quad \text{and} \quad \hat{\sigma}^2 = s_e^2 = e'e/(n-q),$$

respectively, where $e = y - X\hat{\beta}$.

Well known sampling distributions are:

$$\begin{aligned}
(\hat{\beta}|\sigma^2, y, X) &\sim N(\beta, \sigma^2(X'X)^{-1}) \\
(\hat{\beta}|y, X) &\sim t_{n-q}(\beta, s_e^2) \\
(\hat{\sigma}^2|\beta, \sigma^2) &\sim IG((n-q)/2, ((n-q)\sigma^2/2).
\end{aligned}$$

Linear regression: predictive density

The predictive density $p(y|X)$ can be seen as the *marginal likelihood*, i.e.

$$p(y|X) = \int p(y|X,\beta,\sigma^2)p(\beta|\sigma^2)p(\sigma^2)d\beta d\sigma^2 \tag{1}$$

or, by Bayes' theorem, as the *normalizing constant*, i.e.

$$p(y|X) = \frac{p(y|X,\beta,\sigma^2)p(\beta|\sigma^2)p(\sigma^2)}{p(\beta|\sigma^2,y,X)p(\sigma^2|y,X)} \tag{2}$$

which is valid for all (β,σ^2).

Linear regression: predictive density

Closed form solutions are

$$(y|X) \sim t_{n_0}(Xb_0, s_0^2(I_n + XB_0X'))$$

and

$$p(y|X) = \frac{f_N(y; X\beta, \sigma^2 I_n) f_N(\beta; b_0, \sigma^2 B_0) f_{IG}(\sigma^2; n_0/2, n_0 s_0^2/2)}{f_N(\beta; b_1, \sigma^2 B_1) f_{IG}(\sigma^2; n_1/2, n_1 s_1^2/2)}$$

where $f_N(x; \mu, \sigma^2)$ is the density of a normal distribution with mean μ and variance σ^2 evaluated at x and $f_{IG}(x; a, b)$ is the density of an inverse gamma distribution with parameters a and b evaluated at x.

Linear regression: approximating the predictive

Let $\left\{(\beta_i, \sigma_i^2)\right\}_{i=1}^M$ and $\left\{(\tilde{\beta}_i, \tilde{\sigma}_i^2)\right\}_{i=1}^M$ be draws from the prior $p(\beta, \sigma^2)$ and from the posterior $p(\beta, \sigma^2|y, X)$, respectively.

The Monte Carlo and harmonic mean estimators of $p(y|X)$ are

$$\begin{aligned} p_{mc}(y|X) &= \frac{1}{M}\sum_{i=1}^{M} p(y|X, \beta_i, \sigma_i^2) \\ p_{hm}(y|X) &= \left(\frac{1}{M}\sum_{i=1}^{M} p^{-1}(y|X, \tilde{\beta}_i, \tilde{\sigma}_i^2)\right)^{-1}. \end{aligned}$$

Chib's method approximates $p(\sigma^2|y, X)$ from equation (2) by

$$\frac{1}{M}\sum_{i=1}^{M} f_{IG}(\sigma^2; n_1/2, n_1 s_{11}^2(\tilde{\beta}_i)/2)$$

where $n_1 s_{11}^2(\beta) = n_0 s_0^2 + (y - X\beta)'(y - X\beta)$.

Linear regression: sufficient statistics recursions

For $y^t = (y_1, \ldots, y_t)$ and $X^t = (x_1, \ldots, x_t)'$, it follows that

$$\begin{aligned}(\beta|\sigma^2, y^t, X^t) &\sim N(b_t, \sigma^2 B_t) \\ (\sigma^2|y^t, X^t) &\sim IG(n_t/2, n_t s_t^2/2)\end{aligned}$$

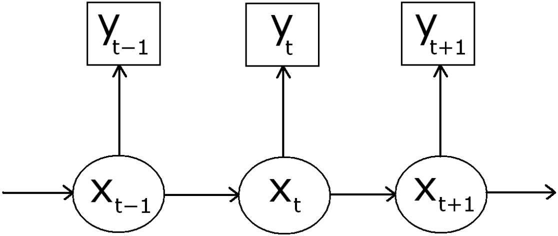
where

$$\begin{aligned}B_t^{-1} &= B_{t-1}^{-1} + x_t x_t' \\ B_t^{-1} b_t &= B_{t-1}^{-1} b_{t-1} + y_t x_t \\ n_t &= n_{t-1} + 1 \\ n_t s_t^2 &= n_{t-1} s_{t-1}^2 + (y_t - b_t' x_t) y_t + (b_{t-1} - b_t)' B_{t-1}^{-1} b_{t-1}\end{aligned}$$

The recursive sufficient statistics are: $x_t x_t'$, $y_t x_t$ and y_t^2.

These recursions will play an important role later on when deriving **sequential Monte Carlo** methods for conditionally Gaussian dynamic linear models, like many stochastic volatility models.

Dynamic linear model: 1st order DLM



The local level model (West and Harrison, 1997) is

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

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

1st order DLM: 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 $m_{t+1} = (1 - A_{t+1})m_t + A_{t+1}y_{t+1}$.

1st order DLM: 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

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

1st order DLM: 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

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

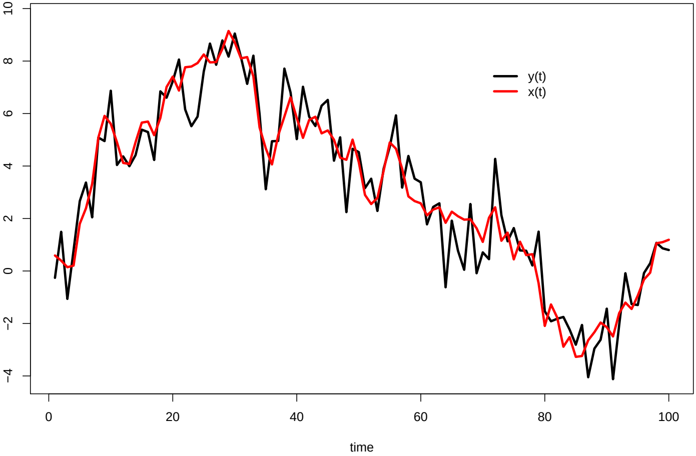
and

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

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).

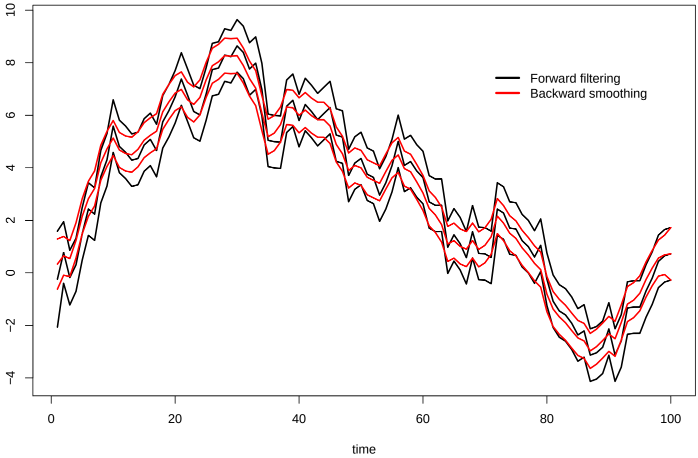
1st order DLM: Simulated data

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



1st order DLM: $p(x_t|y^t,\theta)$ versus $p(x_t|y^n,\theta)$

$m_0 = 0.0$ and $C_0 = 10.0$



1st order DLM: Integrating out states x^n

We showed earlier that

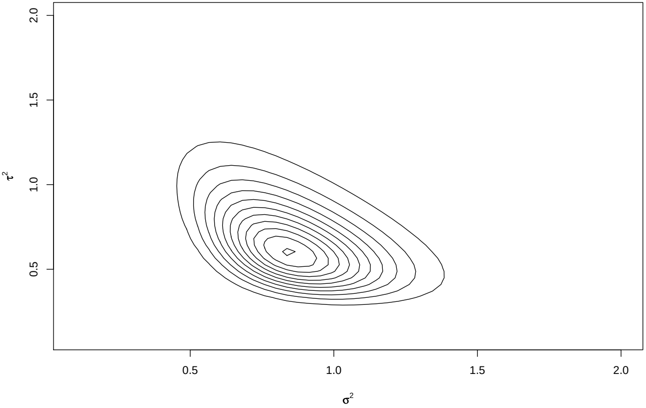
$$(y_t|y^{t-1}) \sim N(m_{t-1}, Q_t)$$

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,

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

1st order DLM: $p(y|\sigma^2, \tau^2)$



1st order DLM: MCMC scheme

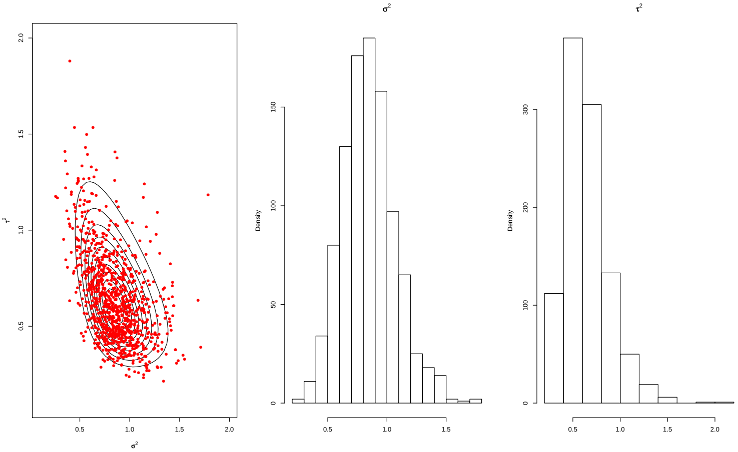
- Sample θ from $p(\theta|y^n, x^n)$

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

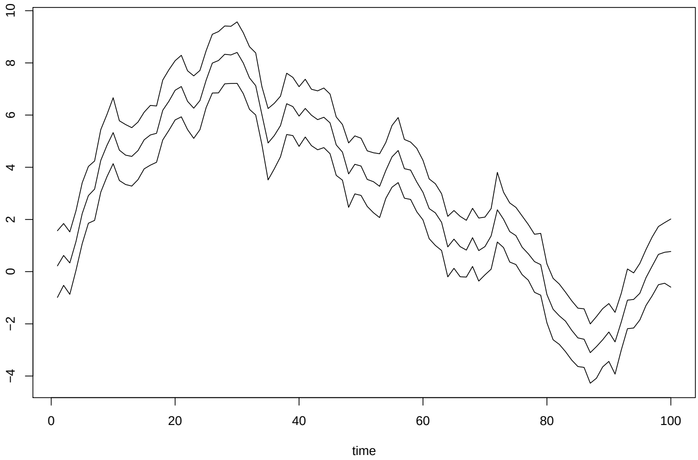
- Sample x^n from $p(x^n|y^n, \theta)$

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

1st order DLM: $p(\sigma^2, \tau^2 | y^n)$



1st order DLM: $p(x_t|y^n)$



Lessons from the 1st order DLM

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.

Dynamic linear models

Large class of models with time-varying parameters.

Dynamic linear models are defined by a pair of equations, the *observation equation* and the *evolution/system equation*:

$$
\begin{aligned}
y_t &= F_t'\beta_t + \epsilon_t, \qquad \epsilon_t \sim N(0, V) \\
\beta_t &= G_t\beta_{t-1} + \omega_t, \quad \omega_t \sim N(0, W)
\end{aligned}
$$

- y_t: sequence of observations;
- F_t: vector of explanatory variables;
- β_t: d-dimensional state vector;
- G_t: $d \times d$ evolution matrix;
- $\beta_1 \sim N(a, R)$.

Example: Linear growth model

The linear growth model is slightly more elaborate by incorporation of an extra time-varying parameter β_2 representing the growth of the level of the series:

$$
\begin{aligned}
y_t &= \beta_{1,t} + \epsilon_t \quad \epsilon_t \sim N(0, V) \\
\beta_{1,t} &= \beta_{1,t-1} + \beta_{2,t} + \omega_{1,t} \\
\beta_{2,t} &= \beta_{2,t-1} + \omega_{2,t}
\end{aligned}
$$

where $\omega_t = (\omega_{1,t}, \omega_{2,t})' \sim N(0, W)$ and

$$
\begin{aligned}
F_t &= (1, 0)' \\
G_t &= \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}
\end{aligned}
$$

Prior, updated and smoothed distributions

Prior distributions

$$p(\beta_t|y^{t-k}) \qquad k > 0$$

Updated/online distributions

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

Smoothed distributions

$$p(\beta_t|y^{t+k}) \qquad k > 0$$

Sequential inference

Let $y^t = \{y_1, \ldots, y_t\}$.

Posterior at time $t-1$:

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

Prior at time t:

$$\beta_t|y^{t-1} \sim N(a_t, R_t)$$

with $a_t = G_t m_{t-1}$ and $R_t = G_t C_{t-1} G_t' + W$.

predictive at time t:

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

with $f_t = F_t' a_t$ and $Q_t = F_t' R_t F_t + V$.

Posterior at time t

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

The resulting posterior distribution is

$$\beta_t|y^t \sim N(m_t, C_t)$$

with

$$\begin{aligned} m_t &= a_t + A_t e_t \\ C_t &= R_t - A_t A_t' Q_t \\ A_t &= R_t F_t / Q_t \\ e_t &= y_t - f_t \end{aligned}$$

By induction, these distributions are valid for all times.

Smoothing

In dynamic models, the smoothed distribution $\pi(\beta|y^n)$ is more commonly used:

$$\begin{aligned}\pi(\beta|y^n) &= p(\beta_n|y^n)\prod_{t=1}^{n-1} p(\beta_t|\beta_{t+1},\ldots,\beta_n,y^n)\\ &= p(\beta_n|y^n)\prod_{t=1}^{n-1} p(\beta_t|\beta_{t+1},y^t)\end{aligned}$$

Integrating with respect to $(\beta_1,\ldots,\beta_{t-1})$:

$$\begin{aligned}\pi(\beta_t,\ldots,\beta_n|y^n) &= p(\beta_n|y^n)\prod_{k=t}^{n-1} p(\beta_k|\beta_{k+1},y^t)\\ \pi(\beta_t,\beta_{t+1}|y^n) &= p(\beta_{t+1}|y^n)p(\beta_t|\beta_{t+1},y^t)\end{aligned}$$

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

Smoothing: $p(\beta_t | y^n)$

It can be shown that

$$\beta_t | V, W, y^n \sim N(m_t^n, C_t^n)$$

where

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

Smoothing: $p(\beta|y^n)$

It can be shown that

$$(\beta_t|\beta_{t+1}, V, W, y^n)$$

is normally distributed with mean

$$(G_t' W^{-1} G_t + C_t^{-1})^{-1}(G_t' W^{-1} \beta_{t+1} + C_t^{-1} m_t)$$

and variance $(G_t' W^{-1} G_t + C_t^{-1})^{-1}$.

Forward filtering, backward sampling (FFBS)

Sampling from $\pi(\beta|y^n)$ can be performed by

- Sampling β_n from $N(m_n, C_n)$ and then
- Sampling β_t from $(\beta_t|\beta_{t+1}, V, W, y^t)$, for $t = n-1, \ldots, 1$.

The above scheme is known as the forward filtering, backward sampling (FFBS) algorithm (Carter and Kohn, 1994 and Frühwirth-Schnatter, 1994).

Individual sampling from $\pi(\beta_t|\beta_{-t}, y^n)$

Let $\beta_{-t} = (\beta_1, \ldots, \beta_{t-1}, \beta_{t+1}, \ldots, \beta_n)$.

For $t = 2, \ldots, n-1$

$$
\begin{aligned}
\pi(\beta_t|\beta_{-t}, y^n) &\propto p(y_t|\beta_t)\, p(\beta_{t+1}|\beta_t)\, p(\beta_t|\beta_{t-1}) \\
&\propto f_N(y_t; F_t'\beta_t, V) f_N(\beta_{t+1}; G_{t+1}\beta_t, W) \\
&\times f_N(\beta_t; G_t\beta_{t-1}, W) \\
&= f_N(\beta_t; b_t, B_t)
\end{aligned}
$$

where

$$
\begin{aligned}
b_t &= B_t(\sigma^{-2}F_t y_t + G_{t+1}' W^{-1}\beta_{t+1} + W^{-1}G_t\beta_{t-1}) \\
B_t &= (\sigma^{-2}F_t F_t' + G_{t+1}' W^{-1} G_{t+1} + W^{-1})^{-1}
\end{aligned}
$$

for $t = 2, \ldots, n-1$.

For $t = 1$ and $t = n$,

$$\pi(\beta_1|\beta_{-1}, y^n) = f_N(\beta_1; b_1, B_1)$$

and

$$\pi(\beta_n|\beta_{-n}, y^n) = f_N(\beta_t; b_n, B_n)$$

where

$$\begin{aligned}
b_1 &= B_1(\sigma_1^{-2} F_1 y_1 + G_2' W^{-1} \beta_2 + R^{-1} a) \\
B_1 &= (\sigma_1^{-2} F_1 F_1' + G_2' W^{-1} G_2 + R^{-1})^{-1} \\
b_n &= B_n(\sigma_n^{-2} F_n y_n + W^{-1} G_n \beta_{n-1}) \\
B_n &= (\sigma_n^{-2} F_n F_n' + W^{-1})^{-1}
\end{aligned}$$

Sampling from $\pi(V, W|y^n, \beta)$

Assume that

$$\begin{aligned} \phi = V^{-1} &\sim Gamma(n_\sigma/2, n_\sigma S_\sigma/2) \\ \Phi = W^{-1} &\sim Wishart(n_W/2, n_W S_W/2) \end{aligned}$$

Full conditionals

$$\begin{aligned} \pi(\phi|\beta,\Phi) &\propto \prod_{t=1}^{n} f_N(y_t; F_t'\beta_t, \phi^{-1})\ f_G(\phi; n_\sigma/2, n_\sigma S_\sigma/2) \\ &\propto f_G(\phi; n_\sigma^*/2, n_\sigma^* S_\sigma^*/2) \\ \pi(\Phi|\beta,\phi) &\propto \prod_{t=2}^{n} f_N(\beta_t; G_t\beta_{t-1}, \Phi^{-1})\ f_W(\Phi; n_W/2, n_W S_W/2) \\ &\propto f_W(\Phi; n_W^*/2, n_W^* S_W^*/2) \end{aligned}$$

where $n_\sigma^* = n_\sigma + n$, $n_W^* = n_W + n - 1$,

$$\begin{aligned} n_\sigma^* S_\sigma^* &= n_\sigma S_\sigma + \sigma(y_t - F_t'\beta_t)^2 \\ n_W^* S_W^* &= n_W S_W + \Sigma_{t=2}^n (\beta_t - G_t\beta_{t-1})(\beta_t - G_t\beta_{t-1})' \end{aligned}$$

MCMC scheme to sample from $p(\beta, V, W|y^n)$

- Sample V^{-1} from its full conditional

$$f_G(\phi; n_\sigma^*/2, n_\sigma^* S_\sigma^*/2)$$

- Sample W^{-1} from its full conditional

$$f_W(\Phi; n_W^*/2, n_W^* S_W^*/2)$$

- Sample β from its full conditional

$$\pi(\beta|y^n, V, W)$$

  by the FFBS algorithm.

Likelihood for (V, W)

It is easy to see that

$$p(y^n|V, W) = \prod_{t=1}^{n} f_N(y_t|f_t, Q_t)$$

which is the integrated likelihood of (V, W).

Jointly sampling (β, V, W)

(β, V, W) can be sampled jointly by

- Sampling (V, W) from its marginal posterior

$$\pi(V, W|y^n) \propto l(V, W|y^n)\pi(V, W)$$

  by a rejection or Metropolis-Hastings step;
- Sampling β from its full conditional

$$\pi(\beta|y^n, V, W)$$

  by the FFBS algorithm.

Jointly sampling (β, V, W) avoids MCMC convergence problems associated with the posterior correlation between model parameters (Gamerman and Moreira, 2002).

Example: Comparing sampling schemes[1]

First order DLM with $V = 1$

$$\begin{aligned} y_t &= \beta_t + \epsilon_t\,, & \epsilon_t &\sim N(0,1) \\ \beta_t &= \beta_{t-1} + \omega_t\,, & \omega_t &\sim N(0,W), \end{aligned}$$

with $(n, W) \in \{(100, .01), (100, .5), (1000, .01), (1000, .5)\}$.

400 runs: 100 replications per combination.

Priors: $\beta_1 \sim N(0, 10)$ and V and W have inverse Gammas with means set at true values and coefficients of variation set at 10.

Posterior inference: based on 20,000 MCMC draws.

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

Schemes

Scheme I: Sampling $\beta_1, \ldots, \beta_n$, V and W from their conditionals.
Scheme II: Sampling β, V and W from their conditionals.
Scheme III: Jointly sampling (β, V, W).

Scheme	n=100	n=1000
II	1.7	1.9
III	1.9	7.2

Computing times relative to scheme I. For instance, when $n = 100$ it takes almost 2 times as much to run scheme III.

		Scheme		
W	n	I	II	III
0.01	1000	242	8938	2983
0.01	100	3283	13685	12263
0.50	1000	409	3043	963
0.50	100	1694	3404	923

Sample averages (based on the 100 replications) of effective sample size n_{eff} based on V (see the explanation over the next few pages).

Effective sample size

For a given θ, let $t^{(n)} = t(\theta^{(n)})$, $\gamma_k = Cov_\pi(t^{(n)}, t^{(n+k)})$, the variance of $t^{(n)}$ as $\sigma^2 = \gamma_0$, the autocorrelation of lag k as $\rho_k = \gamma_k/\sigma^2$ and $\tau_n^2/n = Var_\pi(\bar{t}_n)$. It can be shown that, as $n \to \infty$,

$$\tau_n^2 = \sigma^2 \left(1 + 2\sum_{k=1}^{n-1} \frac{n-k}{n}\rho_k\right) \to \sigma^2 \underbrace{\left(1 + 2\sum_{k=1}^{\infty} \rho_k\right)}_{\text{inefficiency factor}} .$$

The *inefficiency factor* measures how far $t^{(n)}$s are from being a random sample and how much $Var_\pi(\bar{t}_n)$ increases because of that.

The *effective sample size* is defined as

$$n_{\text{eff}} = \frac{n}{1 + 2\sum_{k=1}^{\infty} \rho_k}$$

or the size of a random sample with the same variance.

2nd class: Nonnormal, nonlinear dynamic models

Dynamic generalized linear model

Dynamic generalized models were introduced by West, Harrison and Migon (1985).

The model is

$$
\begin{aligned}
f(y_t|\theta_t) &= a(y_t)\exp\{y_t\theta_t + b(\theta_t)\} \\
E(y_t|\theta_t) &= \mu_t \\
g(\mu_t) &= F_t'\beta_t \\
\beta_t &= G_t\beta_{t+1} + w_t
\end{aligned}
$$

with $w_t \sim N(0, W_t)$ and the link function g is again differentiable.

The model is completed with a prior $\beta_1 \sim N(a, R)$.

It combines the prior specification of normal dynamic models with the observational structure of generalized linear models.

Dynamic binomial and Poisson regressions

Dynamic logistic regression with a series of binomial observations y_t with respective success probabilities π_t dynamically related to explanatory variables $x = (x_1, \ldots, x_d)'$ through the logistic link $\text{logit}(\pi_t) = x_t' \beta_t$.

Poisson counts with means λ_t dynamically related through multiplicative perturbations $\lambda_t = \lambda_{t-1} w_t^*$. After a logarithmic transformation, one obtains $\log \lambda_t = \log \lambda_{t-1} + w_t$ with $w_t = \log w_t^*$.

Posterior inference via MCMC

Assuming that the variances of the system disturbances are constant, the model parameters are given by the state parameters $\beta = (\beta_1, \ldots, \beta_n)'$ and the system variance $W = \Phi^{-1}$.

The model is specified with the observation and system equations and completed with the independent prior distributions $\beta_1 \sim N(a, R)$ and $\Phi \sim W(n_W/2, n_W S_W/2)$.

The posterior distribution is given by

$$\pi(\beta, \Phi) \propto \prod_{t=1}^{n} f(y_t|\beta_t) \prod_{i=2}^{n} p(\beta_t|\beta_{t-1}, \Phi)\, p(\beta_1)p(\Phi)\ .$$

Full conditional for Φ

$$
\begin{aligned}
\pi_\Phi(\Phi) &\propto \prod_{t=2}^{n} p(\beta_t|\beta_{t-1},\Phi)\, p(\Phi) \\
&\propto \prod_{t=2}^{n} |\Phi|^{1/2} \exp\left\{-\frac{1}{2}\mathrm{tr}[(\beta_t - G_t\beta_{t-1})(\beta_t - G_t\beta_{t-1})'\Phi]\right\} \\
&\times |\Phi|^{[n_W-(p+1)]/2} \exp\left\{-\frac{1}{2}\mathrm{tr}(n_W S_W \Phi)\right\} \\
&\propto |\Phi|^{[n_W^*-(d+1)]/2} \exp\left\{-\frac{1}{2}\mathrm{tr}\left[(n_W^* S_W^*)\,\Phi\right]\right\} .
\end{aligned}
$$

that is the density of the $W(n_W^*/2, n_W^* S_W^*/2)$ distribution with

$$
\begin{aligned}
n_W^* &= n_W + n - 1 \\
n_W^* S_W^* &= n_W S_W + \sum_{t=2}^{n}(\beta_t - G_t\beta_{t-1})(\beta_t - G_t\beta_{t-1})'
\end{aligned}
$$

Full conditionals for β

For block β

$$\begin{aligned}
\pi_\beta(\beta) &\propto \prod_{t=1}^{n} f(y_t|\beta_t) \prod_{t=2}^{n} p(\beta_t|\beta_{t-1},\Phi)\, p(\beta_1) \\
&\propto \exp\left\{ \sum_{t=1}^{n} [y_t\theta_t + b(\theta_t)] - \frac{1}{2}\sum_{t=1}^{n} (\beta_t - G_t\beta_{t-1})'\Phi(\beta_t - G_t\beta_{t-1}) \right\} .
\end{aligned}$$

For block β_t, $t = 2, \ldots, n-1$

$$\begin{aligned}
\pi_t(\beta_t) &\propto f(y_t|\beta_t)\, p(\beta_t|\beta_{t-1},\Phi) p(\beta_{t+1}|\beta_t,\Phi) \\
&\propto \exp\{y_t\theta_t + b(\theta_t)\} \exp\left\{ -\frac{1}{2} [(\beta_t - G_t\beta_{t-1})'\Phi(\beta_t - G_t\beta_{t-1}) \right. \\
&+ \left. (\beta_{t+1} - G_{t+1}\beta_t)'\Phi(\beta_{t+1} - G_{t+1}\beta_t)]\right\} .
\end{aligned}$$

Similar results follow for blocks β_1 and β_n.

Sampling schemes

Knorr-Held (1997) suggested the use of independence chains with prior proposals.

Shephard and Pitt (1997) used independence chains with proposals based on both prior and a normal approximation to the likelihood.

Ravines (2005) used independence normal proposals for the block β with moments given by the approximation of West, Harrison and Migon (1985).

Singh and Roberts (1982) and Fahrmeir and Wagenpfeil (1997) extended to the dynamic setting the method of mode evaluation for static regression.

An alternative previously discussed is the reparametrization in terms of the system disturbances w_t (Gamerman, 1998)

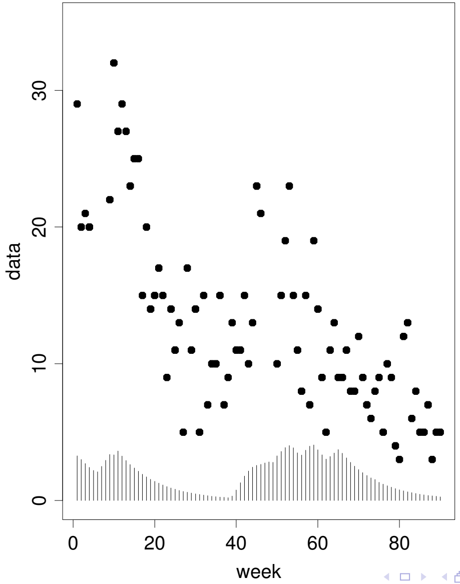
Example: Advertising awareness

Samples of $n_t = 66$ people were selected at random every week for an opinion poll and asked whether they remembered having seen a given advertising campaign on TV. A weekly cumulative measure of campaign expenditure was constructed.

Following Migon and Harrison (1985), the model used for this problem was a dynamic logistic regression

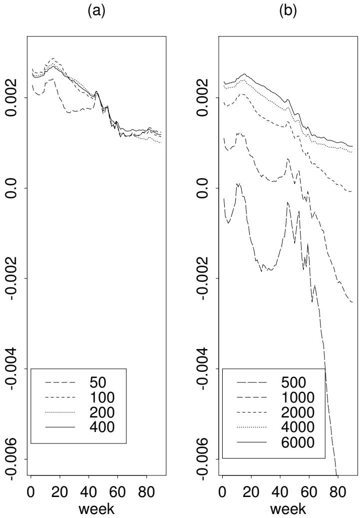
$$\begin{aligned} y_t &\sim bin(n_t, \pi_t) \\ \mu_t &= n_t \pi_t \\ \text{logit}(\pi_t) &= \beta_{1t} + \beta_{2t} x_t \\ \beta_t | \beta_{t-1} &\sim N(\beta_{t-1}, W) \end{aligned}$$

Advertising awareness: Data

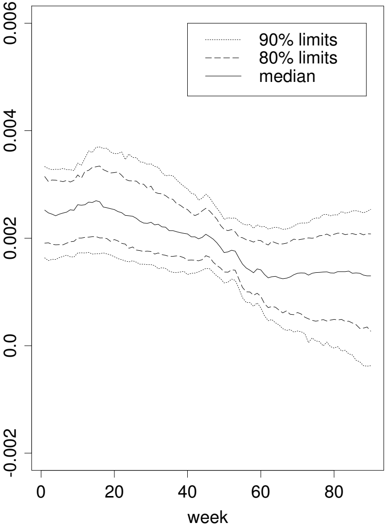


Advertising awareness: Comparing MCMC schemes

Average trajectory of β_{2t} in 500 parallel chains with number of iterations for sampling from: (a) system disturbances; (b) state parameters.



Advertising awareness: Expenditure coefficient β_{2t}



Example: Nonlinear, normal dynamic model

Let y_t, for $t = 1, \ldots, n$, be generated by the following nonlinear dynamic model

$$
\begin{aligned}
(y_t | x_t, \psi) &\sim N(x_t^2/20, \sigma^2) \\
(x_t | x_{t-1}, \psi) &\sim N(G'_{x_{t-1}}\theta, \tau^2) \\
x_0 &\sim N(m_0, C_0)
\end{aligned}
$$

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

Prior distribution

$$
\begin{aligned}
\sigma^2 &\sim IG(n_0/2, n_0\sigma_0^2/2) \\
\theta|\tau^2 &\sim N(\theta_0, \tau^2 V_0) \\
\tau^2 &\sim IG(\nu_0/2, \nu_0\tau_0^2/2)
\end{aligned}
$$

Sampling $(\psi|x_{0:n}, y^n)$

Let $y^n = (y_1, \ldots, y_n)$ and $x_{0:n} = (x_0, \ldots, x_n)'$.

It follows that

$$
\begin{aligned}
(\theta, \tau^2|x_{0:n}) &\sim N(\theta_1, \tau^2 V_1) IG(\nu_1/2, \nu_1\tau_1^2/2) \\
(\sigma^2|y^n, x^n) &\sim IG(n_1/2, n_1\sigma_1^2/2)
\end{aligned}
$$

where $\nu_1 = \nu_0 + n$, $n_1 = n_0 + n$

$$
\begin{aligned}
Z &= (G_{x_0}, \ldots, G_{x_{n-1}})' \\
V_1^{-1} &= V_0^{-1} + Z'Z \\
V_1^{-1}\theta_1 &= V_0^{-1}\theta_0 + Z'x_{1:n} \\
\nu_1\tau_1^2 &= \nu_0\tau_0^2 + (y - Z\theta_1)'(y - Z\theta_1) + (\theta_1 - \theta_0)'V_0^{-1}(\theta_1 - \theta_0) \\
n_1\sigma_1^2 &= n_0\sigma_0^2 + \sum_{t=1}^{n}(y_t - x_t^2/20)^2
\end{aligned}
$$

Sampling $x_1, \ldots, x_n$

Let $x_{-t} = (x_0, \ldots, x_{t-1}, x_{t+1}, \ldots, x_n)$, for $t = 1, \ldots, n-1$, $x_{-0} = x^n$, $x_{-n} = x_{0:(n-1)}$ and $y_0 = \emptyset$.

For $t = 0$

$$p(x_0|x_{-0}, y_0, \psi) \propto f_N(x_0; m_0, C_0) f_N(x_1; G'_{x_0}\theta, \tau^2)$$

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

$$p(x_t|x_{-t}, y_t, \psi) \quad \propto \quad f_N(y_t; x_t^2/20, \sigma^2) f_N(x_t; G'_{x_{t-1}}\theta, \tau^2) f_N(x_{t+1}; G'_{x_t}\theta, \tau^2)$$

For $t = n$

$$p(x_n|x_{-n}, y_n, \psi) \propto f_N(y_n; x_n^2/20, \sigma^2) f_N(x_n; G'_{x_{n-1}}\theta, \tau^2)$$

Metropolis-Hastings algorithm

A simple random walk Metropolis algorithm with tuning variance v_x^2 would work as follows. For $t = 0, \ldots, n$

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

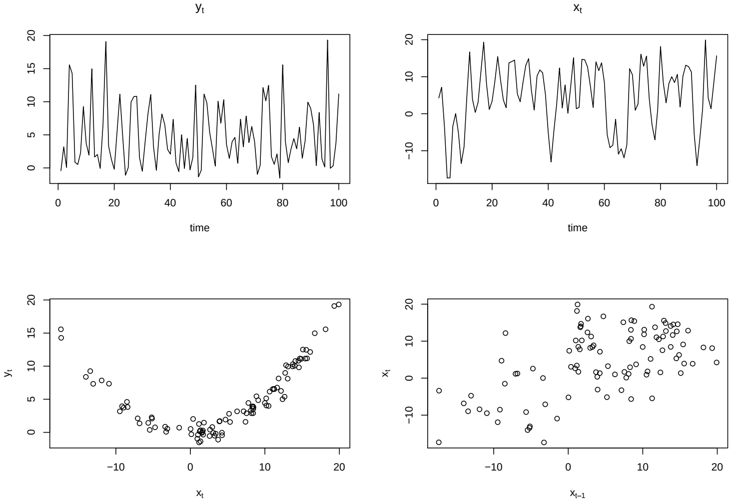
$$\alpha = \min\left\{1, \frac{p(x_t^*|x_{-t}, y_t, \psi)}{p(x_t^{(j)}|x_{-t}, y_t, \psi)}\right\}$$

4. New state:

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

Simulation set up

We simulated $n = 100$ observations based on $\theta = (0.5, 25, 8)'$, $\sigma^2 = 1$, $\tau^2 = 10$ and $x_0 = 0.1$.



Prior hyperparameters

- $x_0 \sim N(m_0, C_0)$

$$m_0 = 0.0 \quad \text{and} \quad C_0 = 10$$

- $\theta|\tau^2 \sim N(\theta_0, \tau^2 V_0)$

$$\theta_0 = (0.5, 25, 8)' \quad \text{and} \quad V_0 = \text{diag}(0.0025, 0.1, 0.04)$$

- $\tau^2 \sim IG(\nu_0/2, \nu_0\tau_0^2/2)$

$$\nu_0 = 6 \quad \text{and} \quad \tau_0^2 = 20/3$$

  such that $E(\tau^2) = \sqrt{V(\tau^2)} = 10.$
- $\sigma^2 \sim IG(n_0/2, n_0\sigma_0^2)$

$$n_0 = 6 \quad \text{and} \quad \sigma_0^2 = 2/3$$

  such that $E(\sigma^2) = \sqrt{V(\sigma^2)} = 1.$

MCMC setup

- Metropolis-Hastings tuning parameter

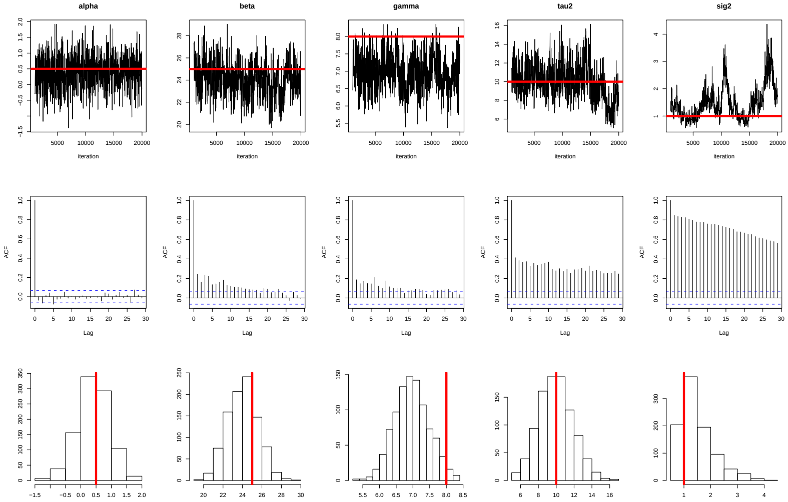
$$v_x^2 = (0.1)^2$$

- Burn-in period, step and MCMC sample size

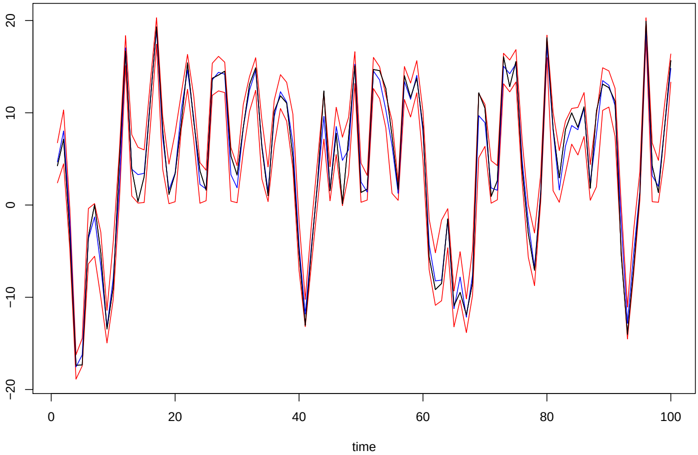
$$M_0 = 1,000 \quad L = 20 \quad M = 950 \;\Rightarrow\; 20,000 \text{ draws}$$

- Initial values
  - $\theta = (0.5, 25, 8)'$
  - $\tau^2 = 10$
  - $\sigma^2 = 1$
  - $x_{0:n} = x_{0:n}^{\text{true}}$

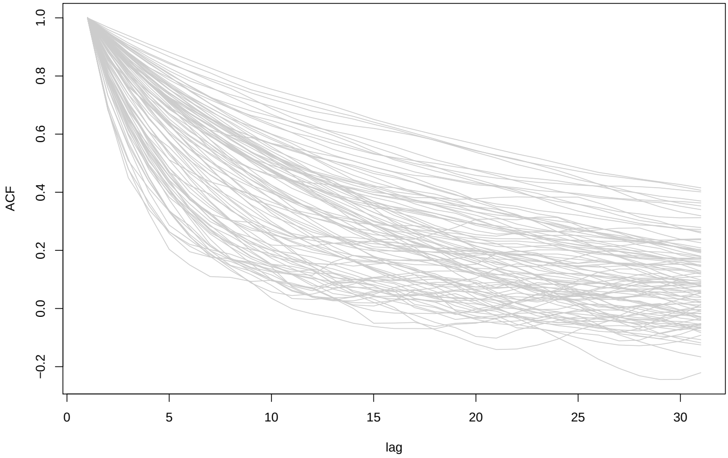
Parameters



States

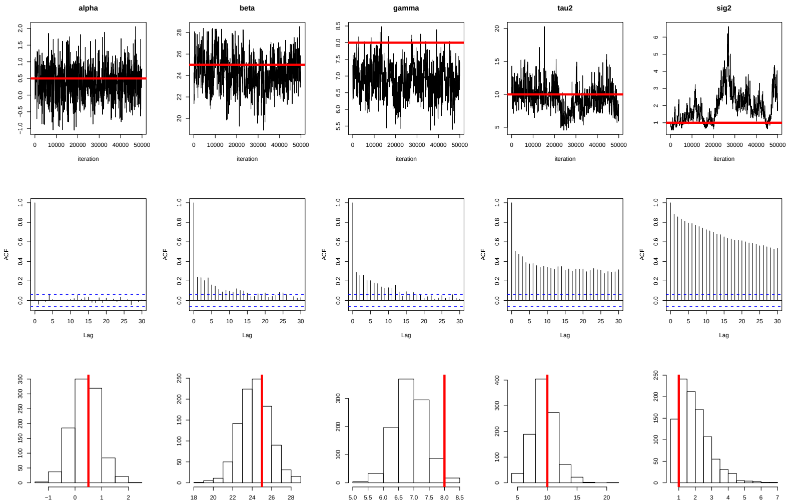


States

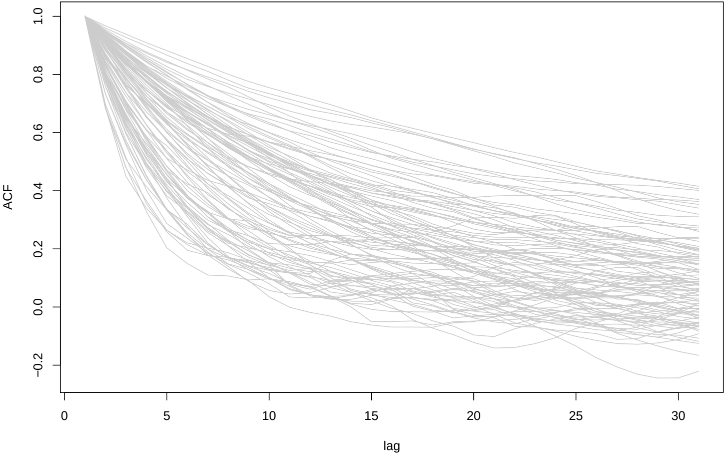


Parameters

$$M_0 = 100,000 \quad L = 50 \quad M = 1000 \quad \Rightarrow \quad 150,000 \text{ draws}$$



States



3rd class: 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

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

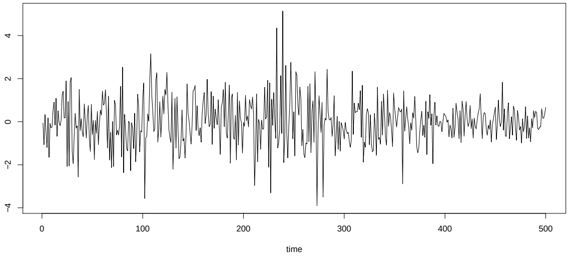
4. New state:

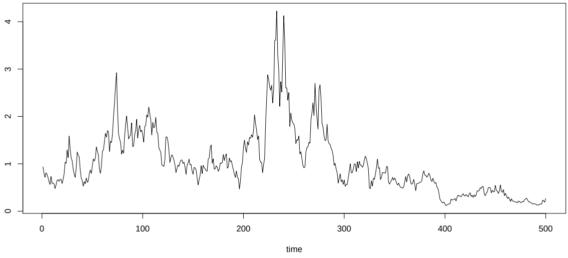
$$h_t^{(j+1)} = \begin{cases} h_t^* & \text{w. p. } \alpha \\ h_t^{(j)} & \text{w. p. } 1-\alpha \end{cases}$$

Example i. Simulated data

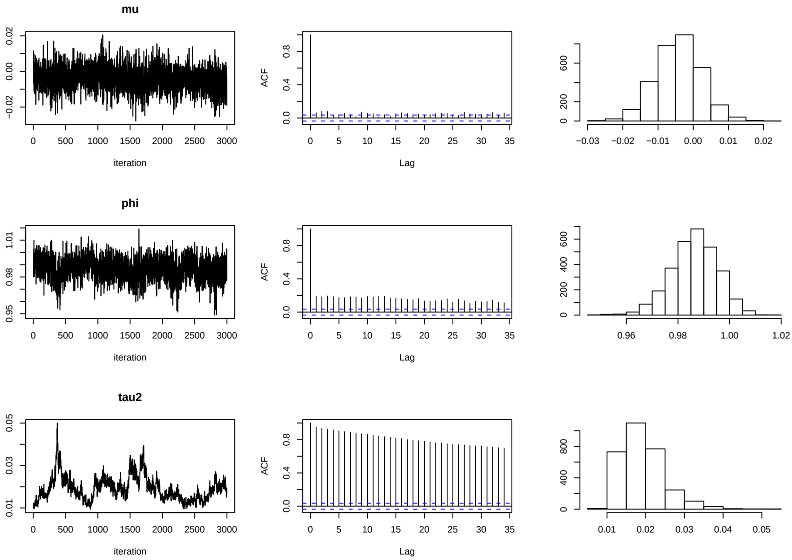
- Simulation setup
  - $n = 500$
  - $h_0 = 0.0$
  - $\mu = -0.00645$
  - $\phi = 0.99$
  - $\tau^2 = 0.15^2$
- Prior distribution
  - $\mu \sim N(0, 100)$
  - $\phi \sim N(0, 100)$
  - $\tau^2 \sim IG(10/2, 0.28125/2)$
  - $h_0 \sim N(0, 100)$
- MCMC setup
  - $M_0 = 1,000$
  - $M = 1,000$

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

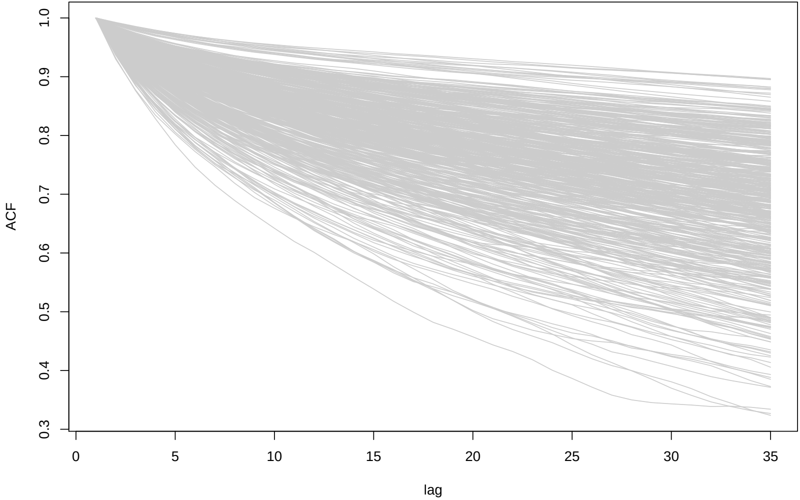




Parameters

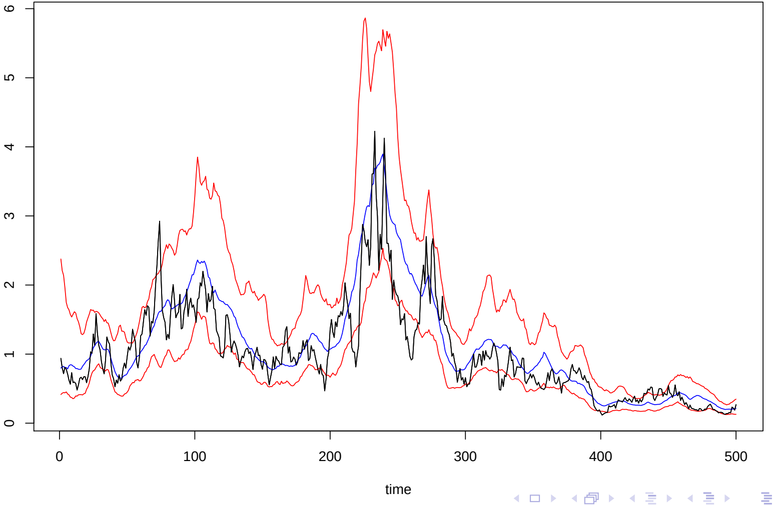


Autocorrelation of h_t



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

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

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

Metropolis-Hastings algorithm

For $t = 1, \ldots, n$

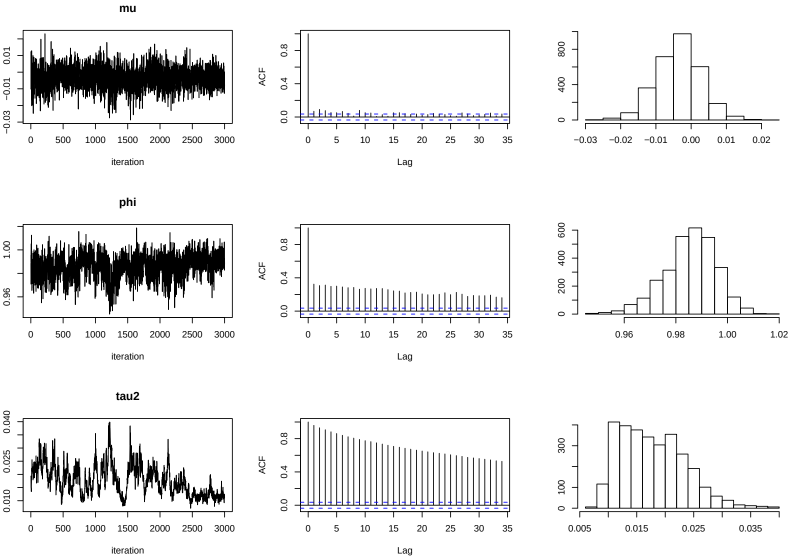
1. Current state: $h_t^{(j)}$
2. Sample h_t^* from $N(\tilde{\mu}_t, \nu_t^2)$
3. Compute the acceptance probability

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

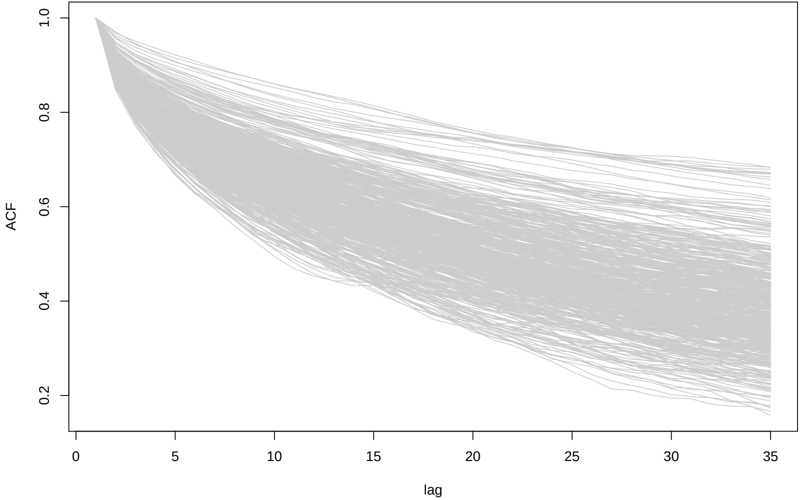
4. New state:

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

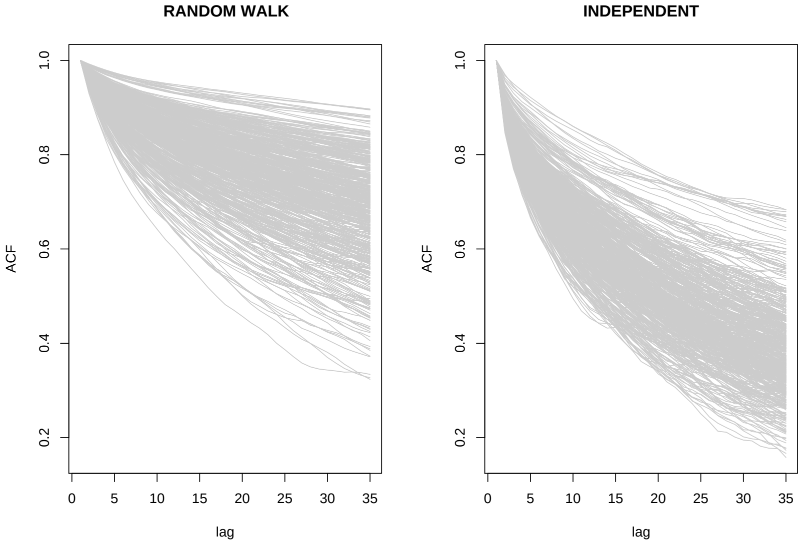
Parameters


mu
phi
tau2
iteration
Lag
ACF
0 500 1000 1500 2000 2500 3000
0 5 10 15 20 25 30 35
−0.03 −0.01 0.01
0.0 0.4 0.8
0 400 800
−0.03 −0.02 −0.01 0.00 0.01 0.02
0.96 1.00
0 200 400 600
0.96 0.98 1.00 1.02
0.010 0.025 0.040
0 100 300
0.005 0.015 0.025 0.035

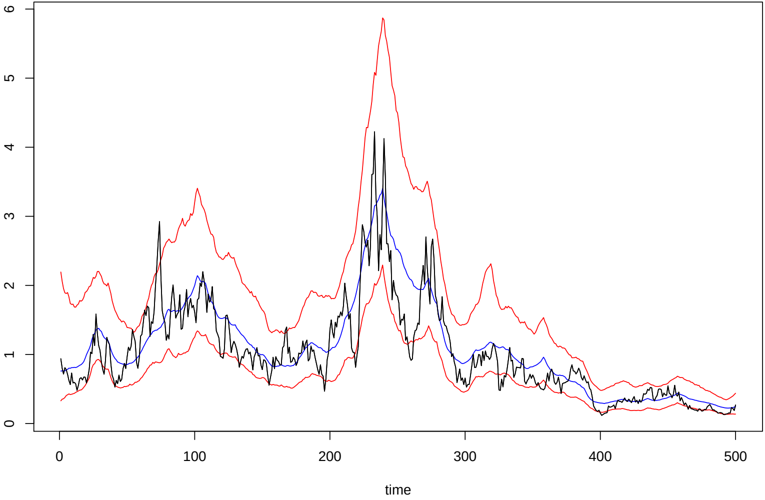
Autocorrelation of h_t



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

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

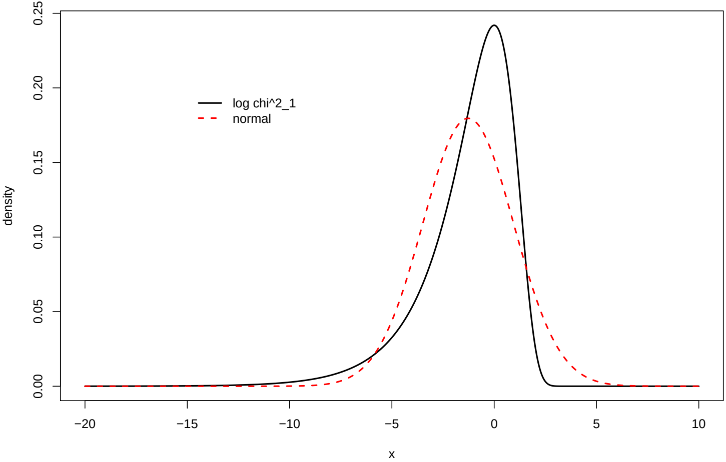
is a simple DLM where v_t and η_t are $N(0, 1)$.

Sampling from

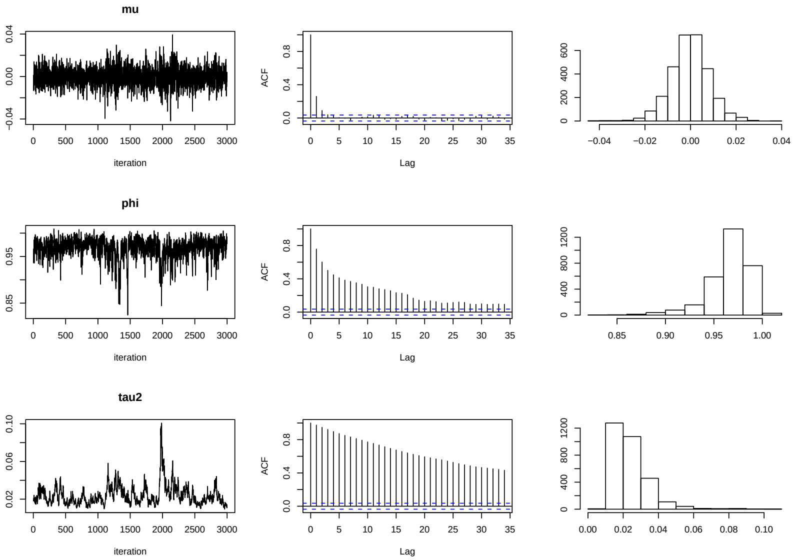
$$p(h^n | \theta, \tau^2, \sigma^2, z^n)$$

can be performed by the FFBS algorithm.

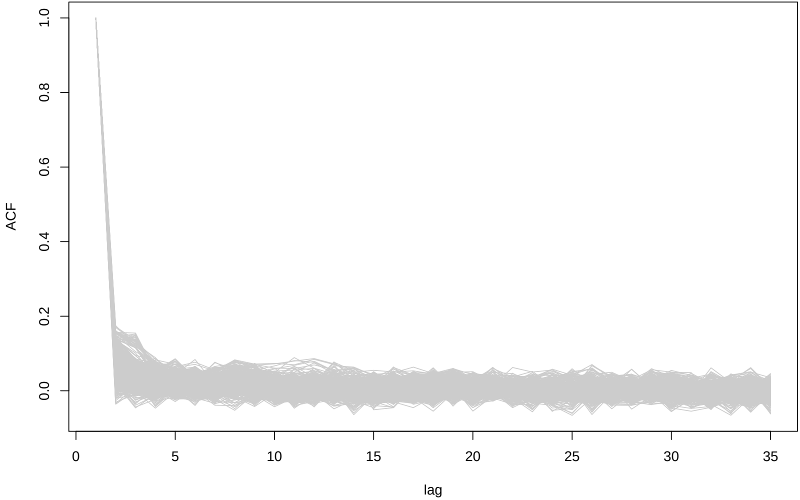
$\log \chi_1^2$ and $N(-1.27, \pi^2/2)$



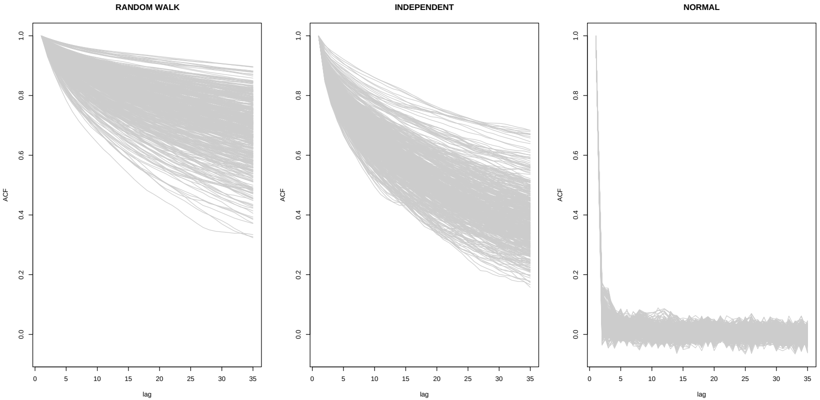
Parameters



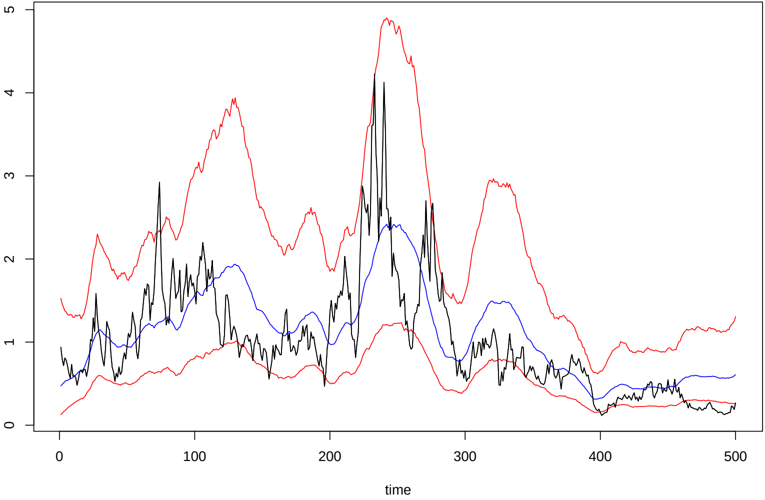
Autocorrelation of h_t



Autocorrelations of h_t for the three schemes



Volatilities



Sampling h^n - mixtures of normals and FFBS

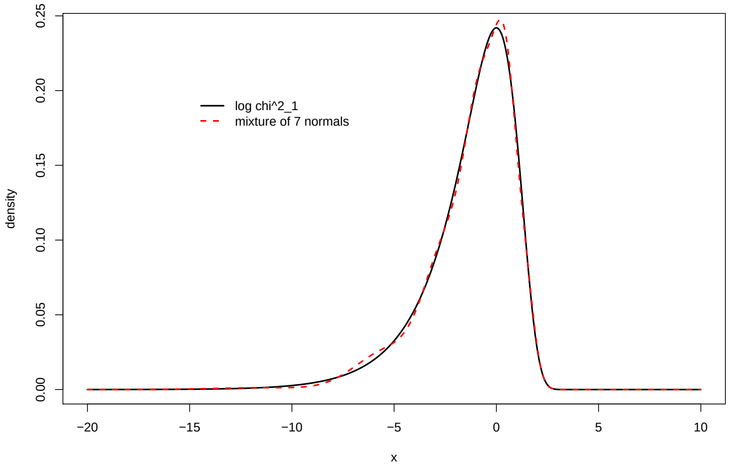
The log χ_1^2 distribution can be approximated by

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

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

$\log \chi^2_1$ 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 $\log y_t^2$,

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

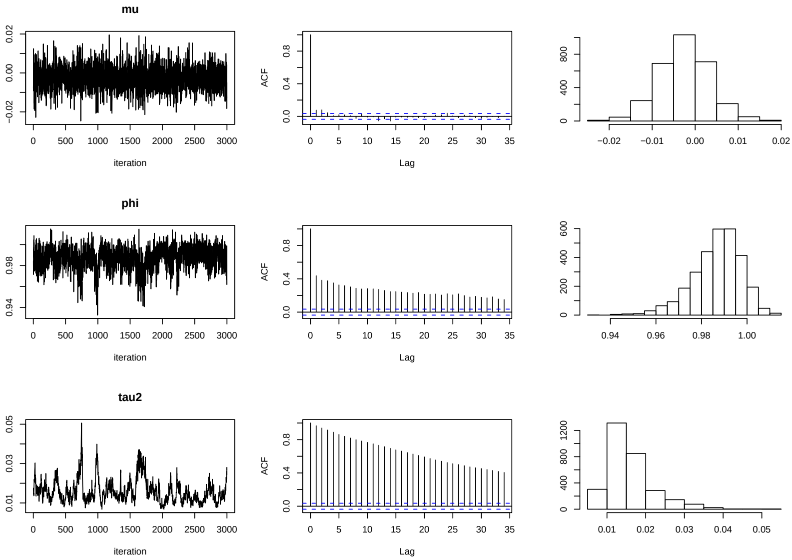
which can be rewritten as a normal DLM:

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

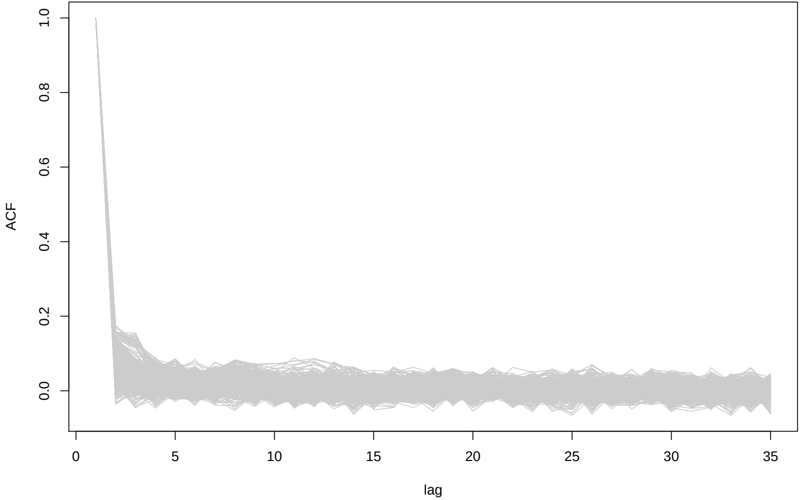
where μ_{z_t} and $\omega_{z_t}^2$ are provided in the previous table.

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

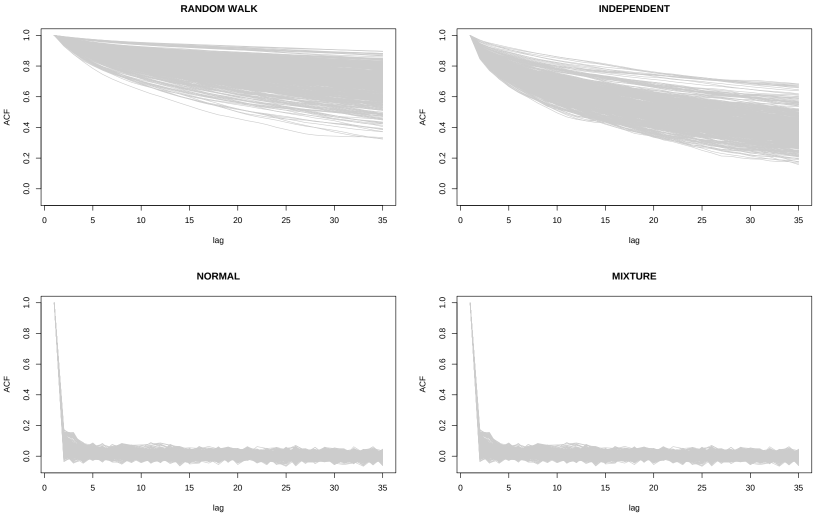
Parameters


mu
phi
tau2
iteration
ACF
Lag
0 500 1000 1500 2000 2500 3000
0 5 10 15 20 25 30 35
−0.02 0.00 0.02
0.0 0.4 0.8
0 400 800
−0.02 −0.01 0.00 0.01 0.02
0.94 0.98
0 200 400 600
0.94 0.96 0.98 1.00
0.01 0.03 0.05
0 400 800 1200
0.01 0.02 0.03 0.04 0.05

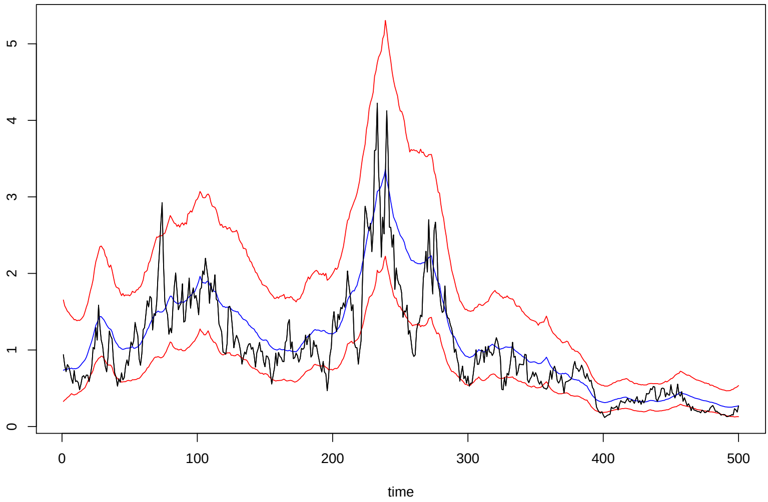
Autocorrelation of h_t



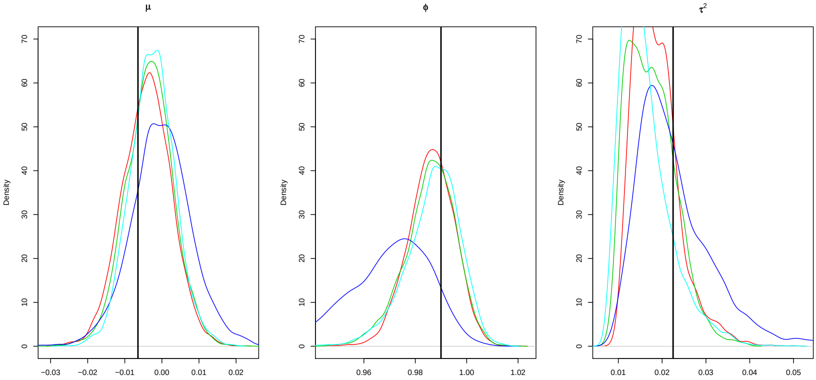
Autocorrelations of h_t for the four schemes



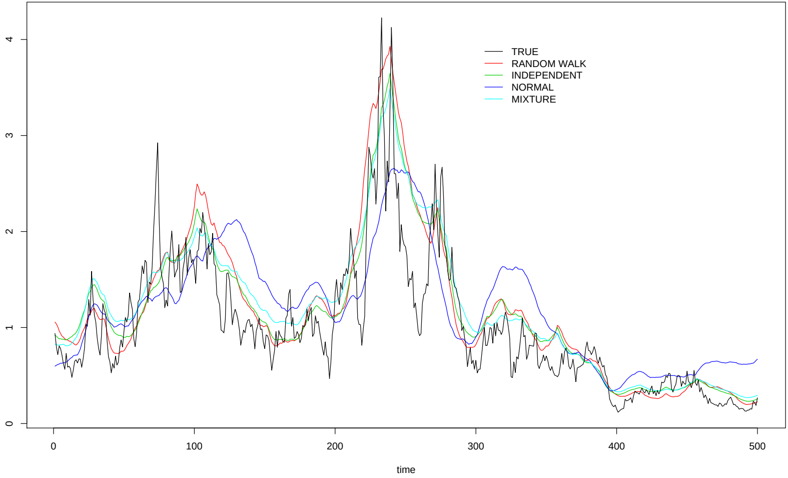
Volatilities



Comparing the four schemes: parameters



Comparing the four schemes: volatilities



4th class: More on stochastic volatility models

Lopes and Salazar (2006)[2]

We extend the SV-AR(1) where

$$y_t \sim N(0, \exp\{h_t\})$$

to accommodate a smooth regime shift, i.e.

$$h_t \sim N(\alpha_{1t} + F(\gamma, \kappa, h_{t-d})\alpha_{2t}, \sigma^2)$$

where

$$\begin{aligned} \alpha_{it} &= \mu_i + \phi_i h_{t-1} + \delta_i h_{t-2} \qquad i = 1, 2 \\ F(\gamma, \kappa, h_{t-d}) &= \frac{1}{1 + exp\left(\gamma(\kappa - h_{t-d})\right)} \end{aligned}$$

such that $\gamma > 0$ drives smoothness and c is a threshold.

[2] 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*, Volume 20, 229-242.

Modeling S&P500 returns

Data from Jan 7th, 1986 to Dec 31st, 1997 (3127 observations)

Models	AIC	BIC	DIC
AR(1)	12795	31697	7223.1
AR(2)	12624	31532	7149.2
LSTAR(1,d=1)	**12240**	**31165**	**7101.1**
LSTAR(1,d=2)	12244	31170	7150.3
LSTAR(2,d=1)	12569	31507	7102.4
LSTAR(2,d=2)	12732	31670	7159.4

	Models					
Parameter	$AR(1)$	$AR(2)$	LSTAR(1,1)	LSTAR(1,1)	LSTAR(2,1)	LSTAR(2,1)
	Posterior mean (standard deviation)					
μ_1	-0.060	-0.066	0.292	-0.354	-4.842	-6.081
	(0.184)	(0.241)	(0.579)	(0.126)	(0.802)	(1.282)
ϕ_1	0.904	0.184	0.306	0.572	-0.713	-0.940
	(0.185)	(0.242)	(0.263)	(0.135)	(0.306)	(0.699)
δ_1	-	0.715	-	-	-1.018	-1.099
		(0.248)			(0.118)	(0.336)
μ_2	-	-	-0.685	0.133	4.783	6.036
			(0.593)	(0.092)	(0.801)	(1.283)
ϕ_2	-	-	0.794	0.237	0.913	1.091
			(0.257)	(0.086)	(0.314)	(0.706)
δ_2	-	-	-	-	1.748	1.892
					(0.114)	(0.356)
γ	-	-	118.18	163.54	132.60	189.51
			(16.924)	(23.912)	(10.147)	(0.000)
κ	-	-	-1.589	0.022	-2.060	-2.125
			(0.022)	(0.280)	(0.046)	(0.000)
σ^2	0.135	0.234	0.316	0.552	0.214	0.166
	(0.020)	(0.044)	(0.066)	(0.218)	(0.035)	(0.026)

Carvalho and Lopes (2007)[4]

We extend the SV-AR(1) to accommodate a Markovian regime shift[3], i.e.

$$h_t \sim N(\mu_{s_t} + \phi h_{t-1}, \sigma^2)$$

and

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

for

$$\alpha_{s_t} = \gamma_1 + \sum_{j=1}^{k} \gamma_j I_{jt}$$

where $I_{jt} = 1$ if $s_t \geq j$ and zero otherwise, $\gamma_1 \in \Re$ and $\gamma_i > 0$ for $i > 1$.

---

[3]So, Lam and Li (1998) A stochastic volatility model with Markov switching. *JBES*, 16, 244-253.

[4]Simulation-based sequential analysis of Markov switching stochastic volatility models, *Computational Statistics and Data Analysis*, 51, 4526-4542.

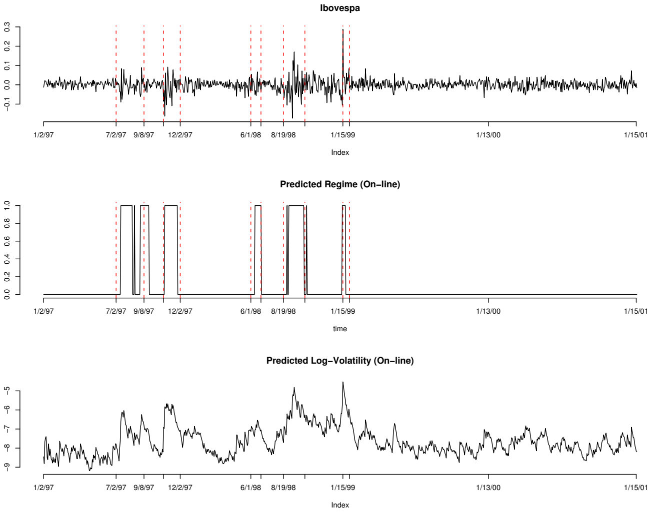
Modeling IBOVESPA returns

We analyzed IBOVESPA returns from 01/02/1997 to 01/16/2001 (1000 observations) based on a 2-regime model.

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.

Model	95% credible interval	$E(\phi\|D_T)$
SV	(0.9325;0.9873)	0.9525
MSSV	(0.8481;0.8903)	0.8707

Also, $E(p_{11}|D_T) = 0.993$ and $E(p_{11}|D_T) = 0.964$.


Ibovespa
0.3
0.2
0.1
0.0
-0.1
1/2/97
7/2/97
9/8/97
12/2/97
6/1/98
8/19/98
1/15/99
1/13/00
1/15/01
Index
Predicted Regime (On-line)
1.0
0.8
0.6
0.4
0.2
0.0
time
Predicted Log-Volatility (On-line)
-5
-6
-7
-8
-9
Index

Abanto, Migon and Lopes (2009)[5]

We use a modied mixture model with Markov switching volatility specfication to analyze the relationship between stock return volatility and trading volume, i.e.

$$
\begin{aligned}
y_t|h_t &\sim t_\nu(0, \exp\{h_t\}) \\
v_t|h_t &\sim \mathit{Poisson}(m_0 + m_1 \exp\{h_t\}) \\
h_t &\sim N(\mu + \gamma s_t + \phi h_{t-1}, \tau^2)
\end{aligned}
$$

with $s_t = 0$ or $s_t = 1$, $\mu \in R$ and $\gamma < 0$.

[5] Bayesian modeling of financial returns: a relationship between volatility and trading volume. *Applied Stochastic Models in Business and Industry*. Available online since June 8th 2009.

Lopes and Polson (2010)[6]

The *stochastic volatility with correlated jumps* (SVCJ) model of Eraker, Johannes and Polson (2003) can be written as

$$
\begin{aligned}
y_{t+1} &= y_t + \mu\Delta + \sqrt{v_t\Delta}\epsilon^y_{t+1} + J^y_{t+1} \\
v_{t+1} &= v_t + \kappa(\theta - v_t) + \sigma_v\sqrt{v_t\Delta}\epsilon^v_{t+1} + J^v_{t+1}
\end{aligned}
$$

where both ϵ^y_{t+1} and ϵ^v_{t+1} follow $N(0,1)$ with $\mathrm{corr}(\epsilon^y_{t+1}, \epsilon^v_{t+1}) = \rho$; and jump components

$$
\begin{aligned}
J^y_{t+1} &= \xi^y_{t+1} N_{t+1} \qquad J^v_{t+1} = \xi^v_{t+1} N_{t+1} \\
\xi^v_{t+1} &\sim Exp(\mu_v) \\
\xi^y_{t+1} | \xi^v_{t+1} &\sim N(\mu_y + \rho_J \xi^v_{t+1}, \sigma^2_y) \\
Pr(N_{t+1} = 1) &= \lambda\Delta
\end{aligned}
$$

Usually, $\Delta = 1$.

[6] Extracting SP500 and NASDAQ volatility: The credit crisis of 2007-2008. *Handbook of Applied Bayesian Analysis*, (to appear).

Credit crisis of 2007

SV model: $\mu = J^y_{t+1} = J^v_{t+1} = 0$ and $\sqrt{v_t\Delta} = 1$ in the evolution equation.

SP500	Mean	StDev	2.5%	97.5%
$\kappa\theta$	-0.0031	0.0029	-0.0092	0.0022
$1-\kappa$	0.9949	0.0036	0.9868	1.0011
σ_v^2	0.0076	0.0026	0.0041	0.0144

SVJ model: $\mu = J^v_{t+1} = \xi^v_{t+1} = 0$ and $\sqrt{v_t\Delta} = 1$ in the evolution equation.

SP500	Mean	StDev	2.5%	97.5%
$\kappa\theta$	-0.0117	0.0070	-0.0262	0.0014
$1-\kappa$	0.9730	0.0084	0.9551	0.9886
σ_v^2	0.0432	0.0082	0.0302	0.0613
λ	0.0025	0.0017	0.0003	0.0066
μ_y	-2.7254	0.1025	-2.9273	-2.5230
σ_y^2	0.3809	0.2211	0.1445	0.9381

Lopes and Migon (2002) and Lopes and Carvalho (2007)[8]

The FSV model of Pitt and Shephard (1999) and Aguilar and West (2000)[7] can be written as

$$
\begin{array}{rcll}
y_t|f_t & \sim & N(\beta f_t, \Sigma_t) & \Sigma_t = \mathrm{diag}(\sigma^2_{1t}, \ldots, \sigma^2_{pt}) \\
f_t & \sim & N(0, H_t) & H_t = \mathrm{diag}(\sigma^2_{p+1,t}, \ldots, \sigma^2_{p+k,t}) \\
\log(\sigma^2_{it}) = \eta_{it} & \sim & N(\alpha_i + \gamma_i \eta_{i,t-1}, \xi^2_i) & i = 1, \ldots, p \\
\log(\sigma^2_{jt}) = \lambda_{jt} & \sim & N(\mu_j + \phi_j \lambda_{j,t-1}, \tau^2_j) & j = 1, \ldots, k \\
\beta_{ijt} & \sim & N(\zeta_{ij} + \Theta_{ij}\beta_{ij,t-1}, \omega^2_{ij}) &
\end{array}
$$

for $i = 2, \ldots, p$ and $j = 1, \ldots, \min(i-1, k)$.

---

[7] AW: Bayesian dynamic factor models and variance matrix discounting for portfolio allocation. *JBES*, 18, 338357. PS: Time varying covariances: a factor stochastic volatility approach (with discussion). *Bayesian Statistics, Volume 6*, 547-570.

[8] LM:Comovements and contagion in emergent markets: stock indexes volatilities. *Case Studies in Bayesian Statistics*, Volume VI, 285-300. LC: Factor stochastic volatility with time varying loadings and Markov switching regimes. *Journal of Statistical Planning and Inference*, 137, 3082-3091.

Daily exchange rate

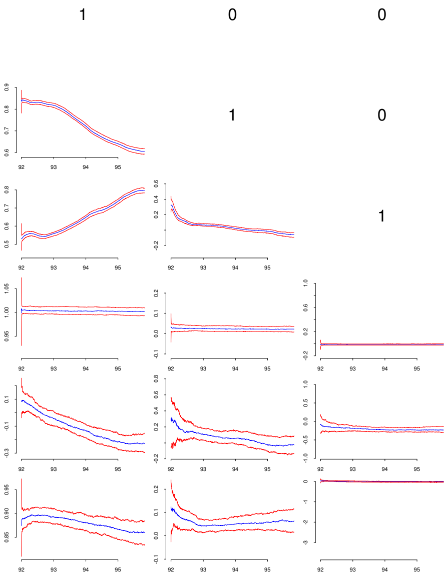
Returns on weekday closing spot prices for six currencies relative to the US dollar.

The data span the period from 1/1/1992 to 10/31/1995 inclusive.

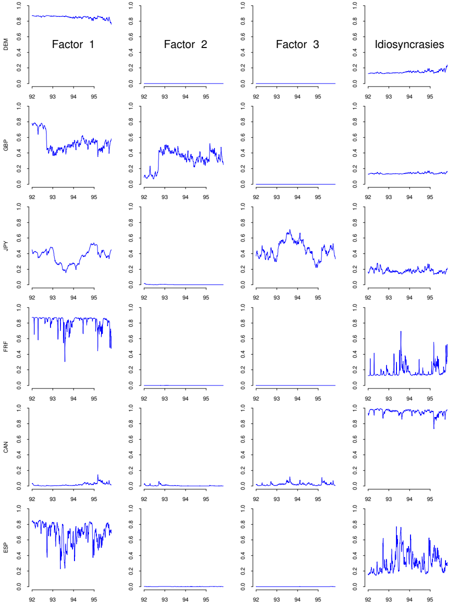
- German Mark(DEM)
- British Pound(GBP)
- Japanese Yen(JPY)
- French Franc(FRF)
- Canadian Dollar(CAD)
- Spanish Peseta(ESP)

A $k = 3$ factor stochastic volatility model with time-varying loadings was implemented with relatively vague priors for all model parameters.

Time-varying factor loadings



Time-varying variance decomposition



More on covariance estimation via factor analysis

West (2003) Bayesian Factor Regression Models in the "Large p, Small n" Paradigm. *Bayesian Statistics 7*, 733-742.

Lopes and West (2004) Bayesian model assessment in factor analysis. *Statistica Sinica*, 14, 41-67.

Carvalho, Chang, Lucas, Wang, Nevins and West (2008) High-dimensional Sparse Factor Modelling: Applications in Gene Expression Genomics. *Journal of the American Statistical Association*, 103, 1438-56.

Lopes, Salazar and Gamerman (2008) Spatial dynamic factor models, *Bayesian Analysis*, 3, 759-92.

Frühwirth-Schnatter and Lopes (2009) Parsimonious BFA when the number of factors is unknown. Technical Report. The University of Chicago Booth School of Business.

Dynamic covariance models

Chib, Nardari and Shephard (2006) Analysis of high dimensional multivariate stochastic volatility models. *Journal of Econometrics*, 134, 341371.

Han (2006) Asset Allocation with a High Dimensional Latent Factor Stochastic Volatility Model. *Review of Financial Studies*, 19, 237-271.

Nardari and Scruggs (2007) Bayesian Analysis of Linear Factor Models with Latent Factors, Multivariate Stochastic Volatility, and APT Pricing Restrictions. *Journal of Financial and Quantitative Analysis*, 42, 857-892.

Carvalho and West (2007) Dynamic Matrix-Variate Graphical Models. *Bayesian Analysis*, 2, 69-98.

Carvalho, Massam and West (2007) Simulation of Hyper Inverse-Wishart Distributions in Graphical Models. *Biometrika*, 94, 647-659.

Rajaratnam, Massam and Carvalho (2008) Flexible Covariance Estimation in Graphical Gaussian Models. *Annals of Statistics*, 36, 2818-2849 .

Scott and Carvalho (2008) Feature-Inclusion Stochastic Search for Gaussian Graphical Models. *Journal of Computational and Graphical Statistics*, 17, 790-808.

Wang, Reeson and Carvalho, Carlos (2009) Dynamic Financial Index Models: Modeling Conditional Dependencies via Graphs. Technical Report, Department of Decision Sciences, Duke University.

5th class: Particle filters with known static parameters

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

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

and

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

respectively.

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

The Bayesian boostrap filter

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

SIR algorithm: The goal is to draw particles from the target density $p(x)$ when the only available draws are from a proposal or candidate density $q(x)$. In fact, both $p(\cdot)$ or $q(\cdot)$ need to be known only up to normalizing constants.

1. Sample $x_1^*, \ldots, x_N^*$ from $q(\cdot)$
2. Compute (unnormalized) weights $\omega_i = p(x_i^*)/q(x_i^*)$
3. Resample $x_1, \ldots, x_M$ from the set $\{x_1^*, \ldots, x_N^*\}$ such that

$$Pr(x_i = x_j^*) \propto \omega_j \quad \forall i$$

Using the prior as proposal

In many Bayesian problems, the target it the posterior distribution, i.e.

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

where $\pi(x)$ and $l(x)$ are prior and likelihood, respectively.

Then a natural (but not necessarily good, actually usually bad!) proposal density is the prior, i.e.

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

This choice leads to the likelihoods as un-normalized weights, i.e.

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

Example: 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: 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: 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

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

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

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

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: 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 $i = 1, \ldots, N$.

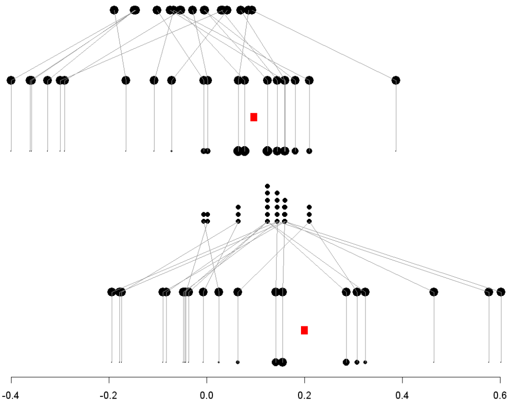
Resampling:

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)

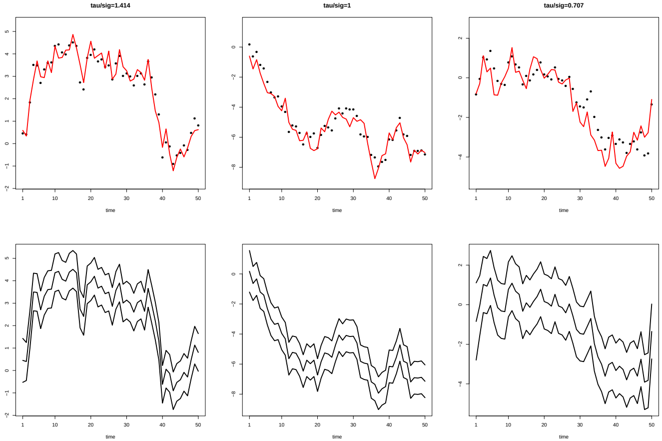
$$
\begin{array}{rcl}
\{x_t^1, \ldots, x_t^N\} & \sim & p(x_t | y^t) \\
\tilde{x}_{t+1}^i & \sim & p(x_{t+1} | x_t^i) \\
\omega_{t+1}^i & \propto & p(y_{t+1} | \tilde{x}_{t+1}^i) \\
\{x_{t+1}^1, \ldots, x_{t+1}^N\} & \sim & p(x_{t+1} | y^{t+1}) \\
\tilde{x}_{t+2}^i & \sim & p(x_{t+2} | x_{t+1}^i) \\
\omega_{t+2}^i & \propto & p(y_{t+2} | \tilde{x}_{t+2}^i)
\end{array}
$$



Uniform weights is the goal!

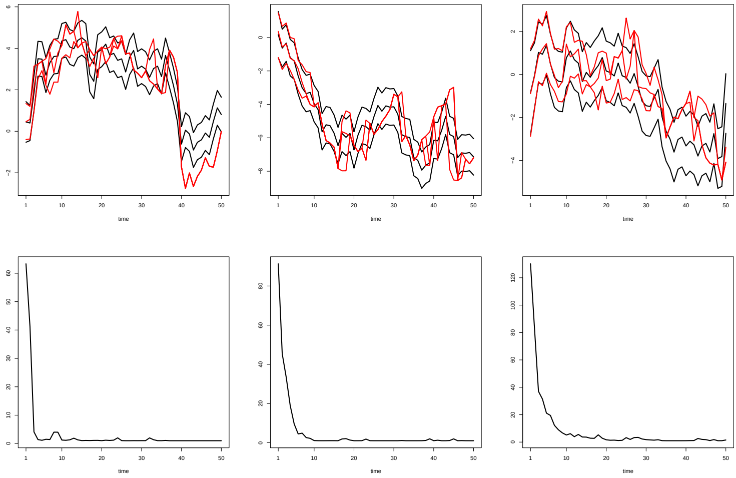
Example: Simulated data

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



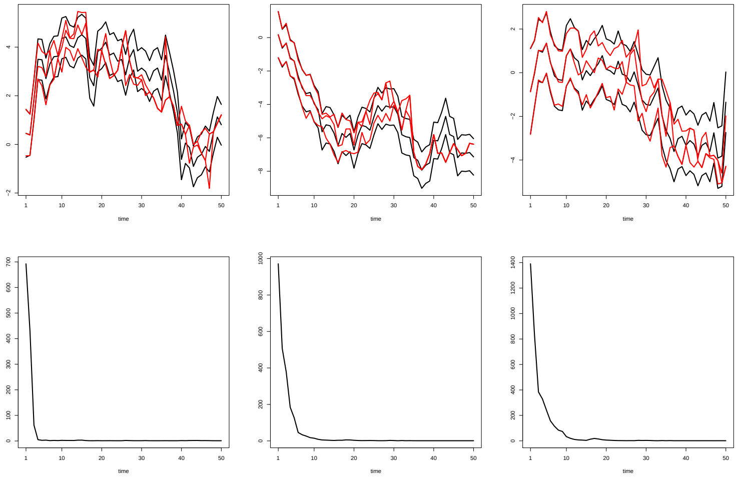
Top: y_t and x_t; bottom: m_t and $m_t \pm 2\sqrt{C_t}$.
Left: $\tau/\sigma = 1.414$; center: $\tau/\sigma = 1.000$; right: $\tau/\sigma = 0.707$.

Example: SIS, $N = 1,000$



Top: States; Bottom: N_{eff}.
Left: $\tau/\sigma = 1.414$; center: $\tau/\sigma = 1.000$; right: $\tau/\sigma = 0.707$.

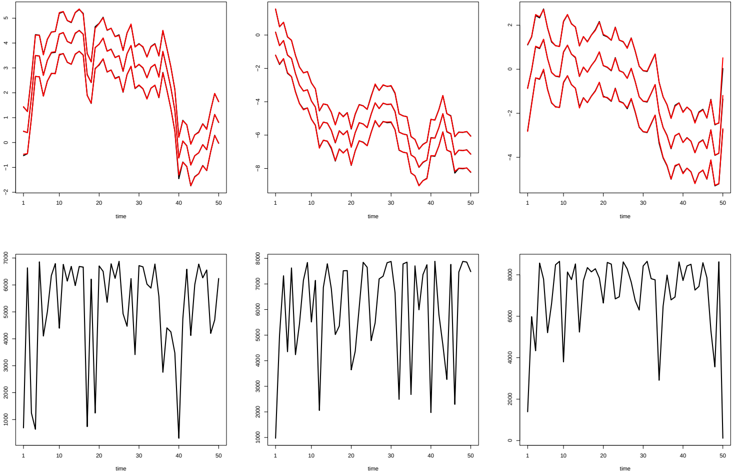
Example: SIS, $N = 10,000$



Top: States; Bottom: N_{eff}.

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

Example: SISR, $N = 10,000$



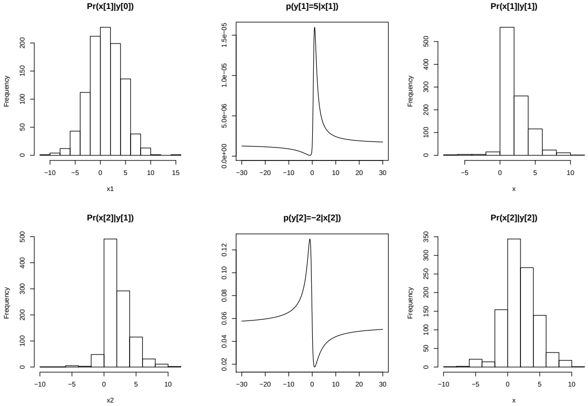
Top: States; Bottom: N_{eff}.
Left: $\tau/\sigma = 1.414$; center: $\tau/\sigma = 1.000$; right: $\tau/\sigma = 0.707$.

Example: Boostrap filter step by step

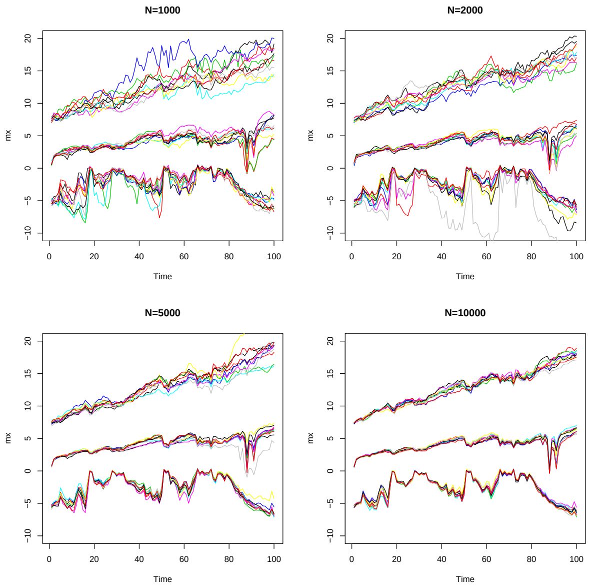
Model

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

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



Example: Boostrap filter Monte Carlo error



Auxiliary particle filter (APF)

Recall the two main steps in any dynamic model:

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

- $\{(x_{t-1}, \omega_{t-1})^{(i)}\}_{i=1}^N$ 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

- $\left\{(x_{t-1}, \omega_{t-1})^{(i)}\right\}_{i=1}^{N}$ 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

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

- $\left\{(x_t, \omega_t)^{(i)}\right\}_{i=1}^{N}$ 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 $t = n-1, \ldots, 1$

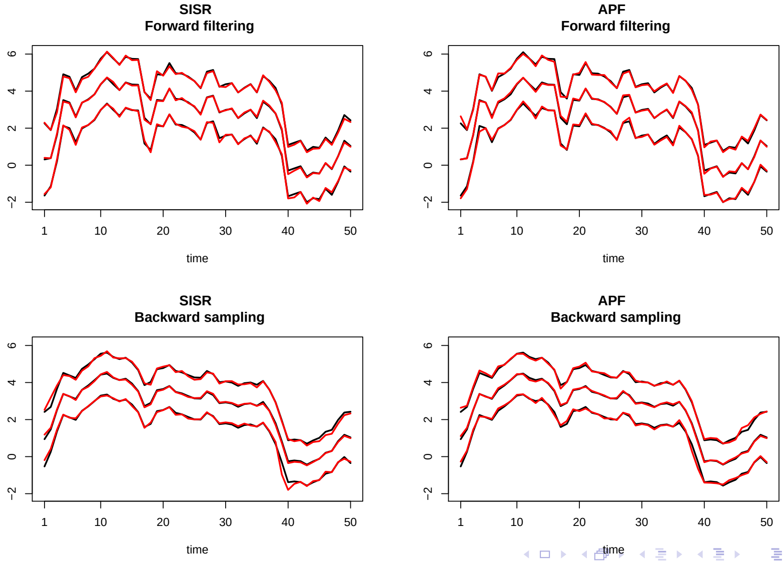
  Sample $\tilde{x}_t$ from $\{x_t^{(i)}\}_{i=1}^N$ with weights $\{\tilde{\omega}_t^{(i)}\}_{i=1}^N$

$$\tilde{\omega}_t^{(i)} \propto \omega_t^{(i)} p\left(\tilde{x}_{t+1} | x_t^{(i)}\right) \qquad i = 1, \ldots, N$$

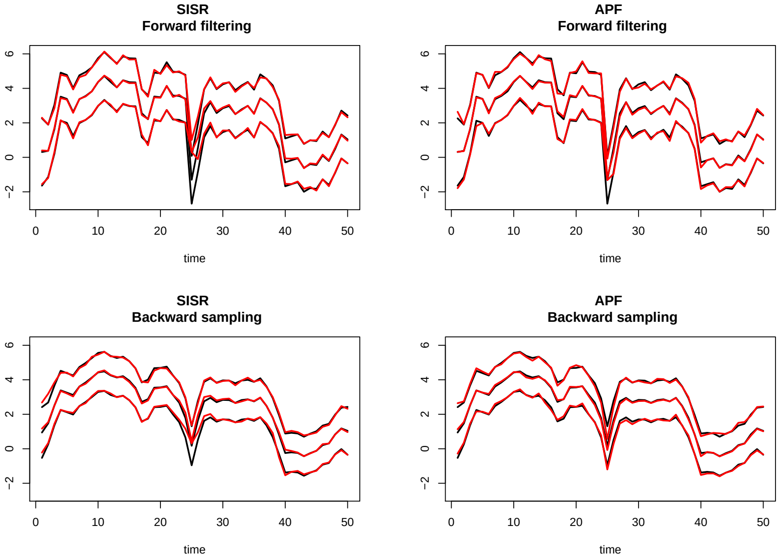
- Then $\{\tilde{x}_1^{(j)}, \ldots, \tilde{x}_n^{(j)}\}$ is a draw from $p(x^n | y^n)$.

Example: smoothing

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



Example: outlier in y_t



Two main types of filters

Sample-resample filters

1. Sample $\tilde{x}_{t+1}^{(j)}$ from $q(x_{t+1}|x_t^{(j)}, y_{t+1})$;
2. Resample $x_{t+1}^{(i)}$ from $\{\tilde{x}_{t+1}^{(j)}\}_{j=1}^N$ with weights

$$\omega_{t+1}^{(j)} \propto \frac{p(y_{t+1}|\tilde{x}_{t+1}^{(j)})p(\tilde{x}_{t+1}^{(j)}|x_t^{(j)})}{q(\tilde{x}_{t+1}^{(j)}|x_t^{(j)}, y_{t+1})}.$$

Resample-sample filters

1. Resample $\tilde{x}_t^{(i)}$ from $\{x_t^{(j)}\}_{j=1}^N$ with weights $q(x_t^{(j)}|y_{t+1})$;
2. Sample $x_{t+1}^{(i)}$ from $q(x_{t+1}|\tilde{x}_t^{(i)}, y_{t+1})$;
3. New weights

$$\omega_{t+1}^{(i)} = \frac{p(y_{t+1}|x_{t+1}^{(i)})p(x_{t+1}^{(i)}|\tilde{x}_t^{(i)})}{q(\tilde{x}_t^{(i)}|y_{t+1})q(x_{t+1}^{(i)}|\tilde{x}_t^{(i)}, y_{t+1})}.$$

Gordon, Salmond and Smith's sample-resample filter

In the Bayesian 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

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

Optimal sample-resample 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

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

This is a **perfectly adapted** filter (Pitt and Shephard, 1999).

Pitt and Shephard's resample-sample 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 *blind* proposal.

The weights are then equal to

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

Optimal resample-sample 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.

Example: BBF, APF and fully adapted versions

Simulation: $M = 20$ data sets with $n = 100$ observations each, $\tau^2 = 0.013$ and $\sigma^2 \in \{5, 0.13, 0.013, 0.0065, 0.0013\}$ from

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

with $x_0 = 0$.

Particle filters: $R = 20$ replications of $N = 1000$ particles.

Prior set up: $x_0 \sim N(0, 10)$.

Example: Log relative mean square error

Let q_α^t be such that

$$Pr(x_t < q_t^\alpha | y^t) = \alpha$$

and $\alpha = (0.05, 0.25, 0.5, 0.75, 0.95)$.

Then, the average MSE for filter f, time t and quantile 100α-percentile is

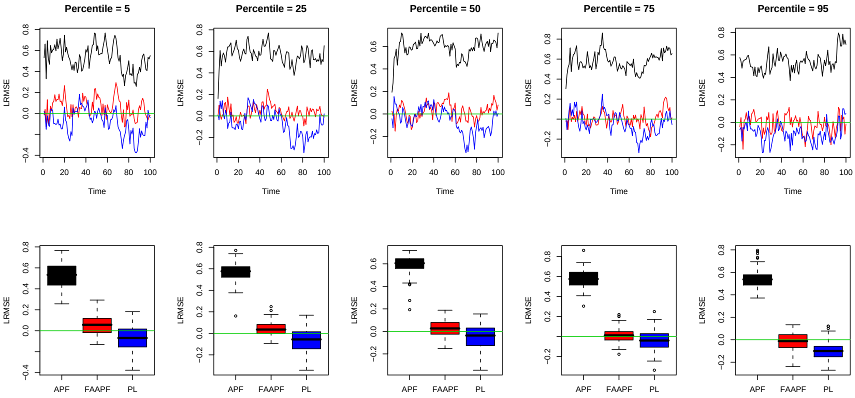
$$MSE_{t,f}^\alpha = \frac{1}{MR} \sum_{i=1}^{M} \sum_{j=1}^{R} (\hat{q}_{tij,f}^\alpha - q_{it}^\alpha)^2$$

We compare filters f_1 and f_2 based on log relative MSE, i.e.

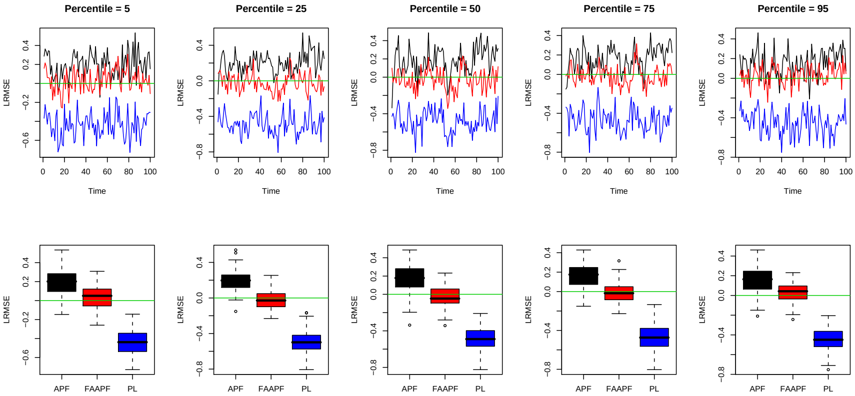
$$LRMLE_{t,f_1,f_2}^\alpha = \log MSE_{t,f_1}^\alpha - MSE_{t,f_2}^\alpha$$

Example: $\sigma^2 = 5.0$ and $\tau/\sigma = 0.05$

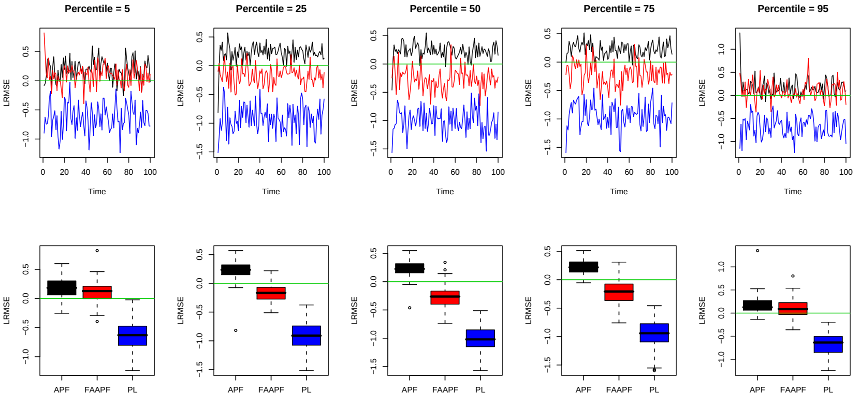
(APF,BBF), (FABBF,BBF) and (FAAPF,BBF)



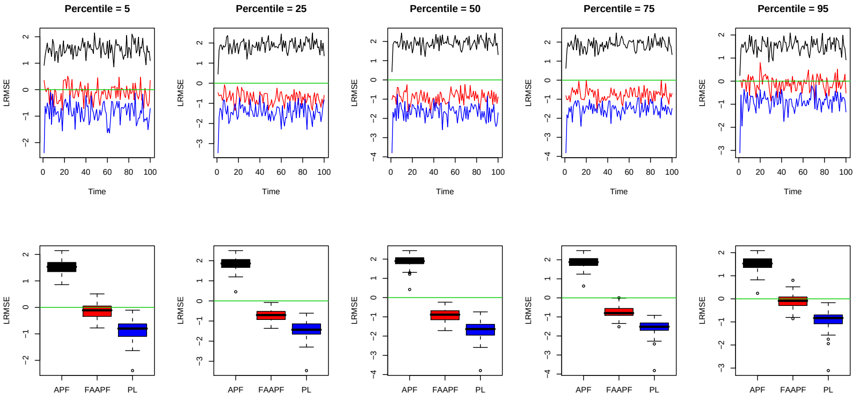
Example: $\sigma^2 = 0.13$ and $\tau/\sigma = 0.32$



Example: $\sigma^2 = 0.013$ and $\tau/\sigma = 1.00$



Example: $\sigma^2 = 0.0013$ and $\tau/\sigma = 3.16$



How about $p(\theta|y^n)$?

Two-step strategy: On the first step, approximate $p(\theta|y^n)$ by

$$p^N(\theta|y^n) = \frac{p^N(y^n|\theta)p(\theta)}{p(y^n)} \propto p^N(y^n|\theta)p(\theta)$$

where $p^N(y^n|\theta)$ is a SMC approximation to $p(y^n|\theta)$. Then, on the 2nd step, sample θ via a MCMC scheme or a SIR scheme[9].

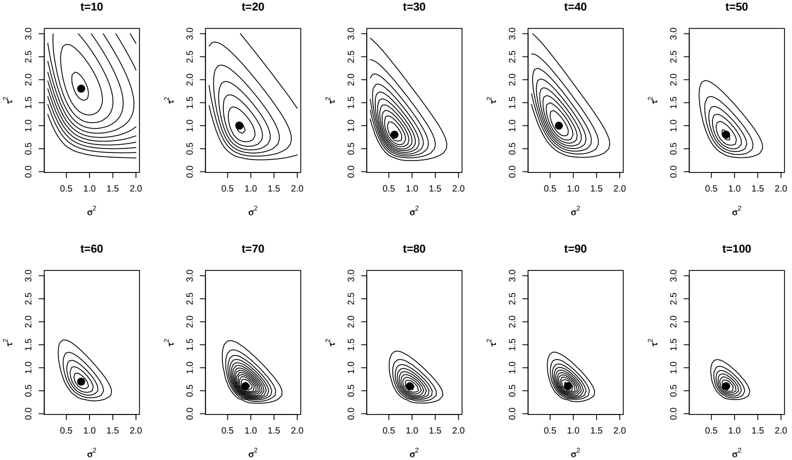
Problem 1: SMC looses its appealing sequential nature.

Problem 2: Overall sampling scheme is sensitive to $p^N(y|\theta)$.

[9] See Fernándes-Villaverde and Rubio-Ramírez (2007) "Estimating Macroeconomic Models: A Likelihood Approach", DeJong, Dharmarajan, Liesenfeld, Moura and Richard (2009) "Efficient Likelihood Evaluation of State-Space Representations" for applications of this two-step strategy to DSGE and related models.

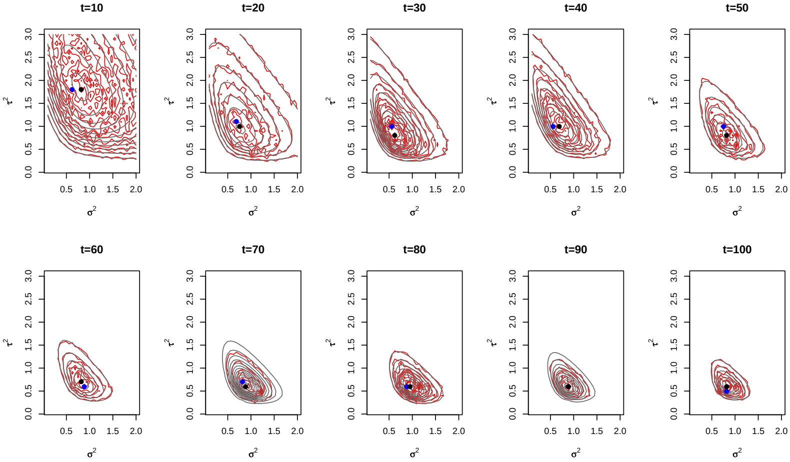
Example: Exact integrated likelihood $p(y^n|\sigma^2, \tau^2)$

$n = 100, x_0 = 0, \sigma^2 = 1, \tau^2 = 0.5$ and $x_0 \sim N(0.0, 100)$
30×30 grid: $\sigma^2 = (0.1, \ldots, 2)$ and $\tau^2 = (0.1, \ldots, 3)$



Example: Approximated $p^N(y^n|\sigma^2, \tau^2)$

Based on $N = 1000$ particles



Another (perhaps more natural) idea

Sequentially learning x_t and θ.

$$\begin{array}{rcl}
\text{Posterior at t} & : & p(x_t|\theta, y^t)p(\theta|y^t) \\
& \Downarrow & \\
\text{Prior ate t+1} & : & p(x_{t+1}|\theta, y^t)p(\theta|y^t) \\
& \Downarrow & \\
\text{Posterior at t+1} & : & p(x_{t+1}|\theta, y^{t+1})p(\theta|y^{t+1})
\end{array}$$

Advantages:

Sequential updates of $p(\theta|y^t)$, $p(x_t|y^t)$ and $p(\theta, x_t|y^t)$
Sequential h-steps ahead forecast $p(y_{t+h}|y^t)$
Sequential approximations for $p(y_t|y^{t-1})$
Sequential Bayes factors

$$B_{12t} = \frac{\prod_{j=1}^{t} p(y_j|y^{j-1}, M_1)}{\prod_{j=1}^{t} p(y_j|y^{j-1}, M_2)}$$

Resample-sample filters with parameter learning

The objective is to go from

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

to

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

Algorithm

- Resample $(\tilde{x}_t, \tilde{\theta}_t)^{(i)}$ from $\{(x_t, \theta_t)^{(j)}\}_{j=1}^N$ with weights $q_1((x_t, \theta_t)^{(j)} | y_{t+1})$, for $j = 1, \ldots, N$.
- 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): APF + learning θ

They approximate $p(\theta|y^t)$ by

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

where $\bar{\theta}_t$ and V_t approximate the mean and variance of θ, given y^t. This leads to

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

and weights

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

where the same decomposition is used for q_2.

Resampling step

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

where

$$\begin{aligned} g(x_t) &= E(x_{t+1} | x_t, m(\theta_t)) \\ m(\theta_t) &= a\theta_t + (1-a)\bar{\theta}_t \end{aligned}$$

The weights are then

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

Choice of 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

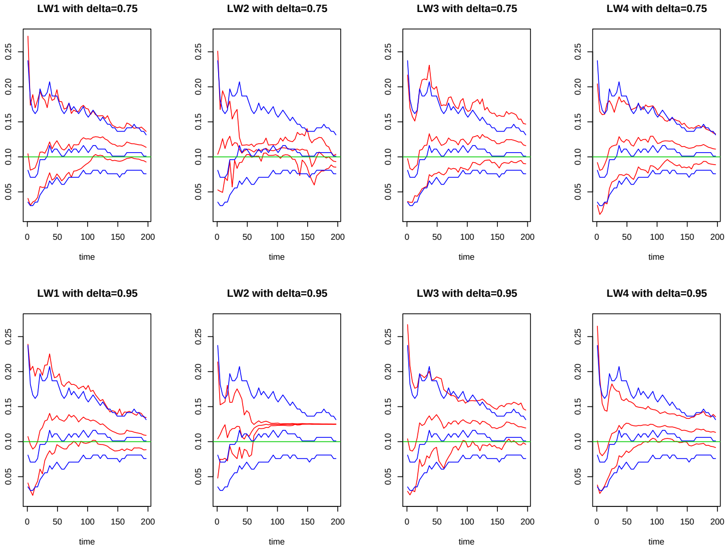
For particles $\{(x_t, \theta_t, \omega_t)^{(j)}\}_{j=1}^N$ summarizing $p(x_t, \theta | y^t)$, estimates $\bar{\theta}_t = \sum_{i=1}^N \omega_t^{(i)} \theta_t^{(i)}$ and $V_t = \sum_{i=1}^N \omega_t^{(i)} (\theta_t^{(i)} - \bar{\theta}_t)(\theta_t^{(i)} - \bar{\theta}_t)'$, and given shrinkage parameter 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 $\theta_{t+1}^{(i)} \sim N(m(\tilde{\theta}_t^{(i)}), h^2 V_t)$.
  - 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 iv. $N = 2000, x_0 = 25, \sigma^2 = .1, \tau^2 = .05$

LW1:$\log \sigma^2$, LW2:σ^2, LW3:LW1 + o.p. LW4:LW2 + o.p.[10]



[10] o.p.=optimal propagation

6th and 7th classes: Particle Learning (PL)

Particle Learning (PL)

1. General framework of sequential parameter learning;
2. Practical alternative to MCMC[11];
3. Sequential Monte assessment;
4. Resample-sample is key[12];
5. *Essential state vector* generalizes sufficient statistics[13];
6. Connection to Rao-Blackwellization[14];
7. PL is not sequential importance sampling;
8. Smoothing offline.

[11] Chen and Liu (1999).
[12] Pitt and Shephard (1999).
[13] Storvik (2002) and Fearnhead (2002)
[14] Kong, Liu and Wong (1994).

The general PL algorithm

- Posterior at t: $\Phi_t \equiv \{(x_t,\theta)^{(i)}\}_{i=1}^N \sim p(x_t,\theta|y^t)$.
- Compute, for $i = 1, \ldots, N$,

$$w_{t+1}^{(i)} \propto p(y_{t+1}|x_t^{(i)},\theta^{(i)})$$

- Resample from Φ_t with weights w_{t+1}: $\tilde{\Phi}_t \equiv \left\{(\tilde{x}_t,\tilde{\theta})^{(i)}\right\}_{i=1}^N$.
- Propagate states

$$x_{t+1}^{(i)} \sim p(x_{t+1}|\tilde{x}_t^{(i)},\tilde{\theta}^{(i)},y_{t+1})$$

- Update sufficient statistics

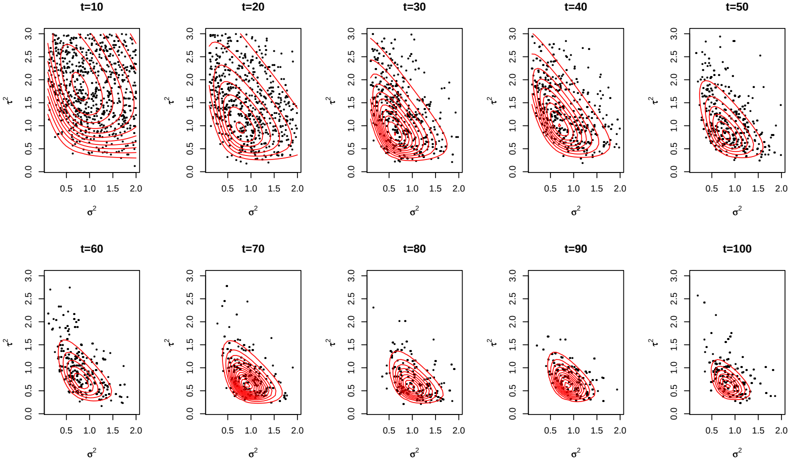
$$s_{t+1}^{(i)} = \mathcal{S}(s_t^{(i)},x_{t+1}^{(i)},y_{t+1})$$

- Sample parameters

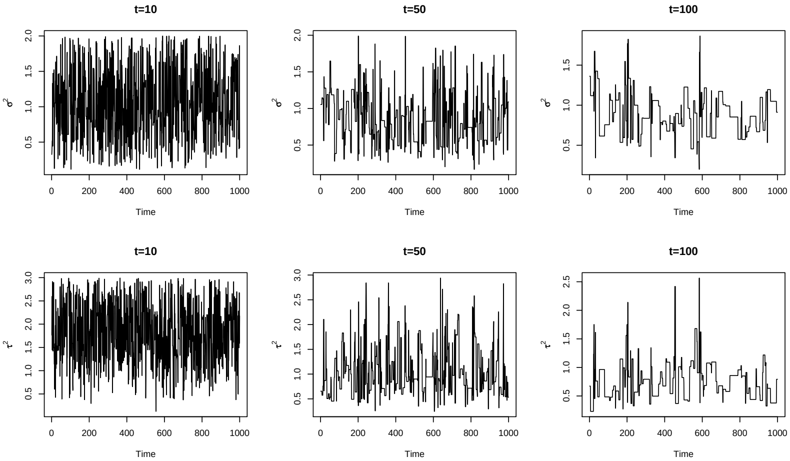
$$\theta^{(i)} \sim p(\theta|s_{t+1}^{(i)})$$

Example i. $p^N(\theta|y^t)$ via PL for x_t and MCMC for (σ^2, τ^2)

Prior: $\sigma^2 \sim U(0.1, 2)$ and $\tau^2 \sim U(0.1, 3)$.
Based on $N = 1000$ particles.

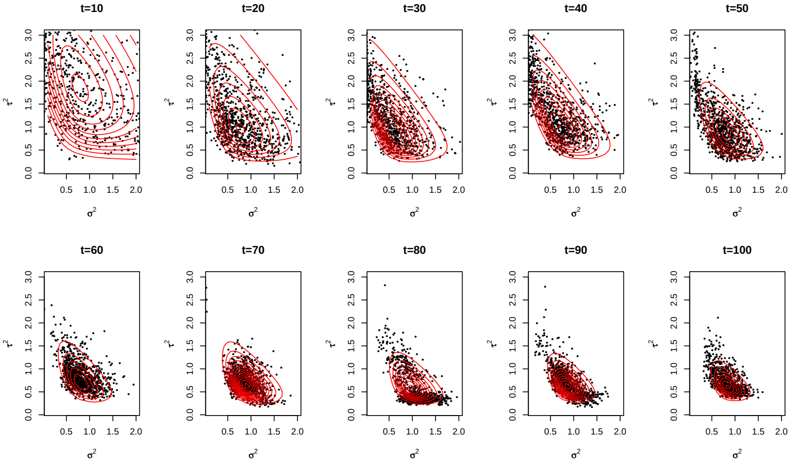


Example i. Trace plots for σ^2 and τ^2



Example i. $p^N(\theta|y^t)$ via PL for (x_t, σ^2, τ^2)

Prior: $\sigma^2 \sim U(0.1, 2)$ and $\tau^2 \sim U(0.1, 3)$.
Based on $N = 1000$ particles.



Comparing PL to alternative filters

Consider a first order DLM for $n = 10$ observations:

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

where $x_0 = 0.0$, $\theta = (\mu, \beta)$, $\mu = 1.0$, $\beta = 0.95$, $\sigma^2 = 1.0$ and $\tau^2 = 0.5$.

We fit the above model with the prior

$$p(x_0, \mu, \beta) = p(x_0)p(\mu)p(\beta)$$

$x_0 \sim N(m_0, C_0)$, $\mu \sim N(\mu_0, V_0)$ and $\beta \sim N(\beta_0, B_0)$, where $m_0 = \mu_0 = \beta_0 = 0.0$ and $C_0 = V_0 = B_0 = 1.0$.

The variances σ^2 and τ^2 are kept fixed.

Exact solution

We already showed that, for any given $t = 1, \dots, n$,

$$p(\theta|y^t) \propto p(\theta) \prod_{t'=1}^{t} f_N(y_{t'}; \mu + \beta m_{t'-1}, \beta^2 C_{t'-1} + \tau^2 + \sigma^2)$$

where, for $t = 1, \dots, n$,

$$\begin{aligned} m_t &= (1 - A_t)\mu + (1 - A_t)\beta m_{t'-1} + A_t y_t \\ C_t &= A_t \sigma^2 \end{aligned}$$

and

$$A_t = \frac{\beta^2 C_{t-1} + \tau^2}{\beta^2 C_{t-1} + \tau^2 + \sigma^2}.$$

Liu and West (LW)

- Posterior at t: $\left\{(x_t, \theta, w_t)^{(i)}\right\}_{i=1}^N \sim p(x_t, \theta | y^t)$
- Weights:

$$w_{t+1}^{(i)} \propto w_t^{(i)} f_N(y_{t+1}; \mu^{(i)} + \beta^{(i)} x_t^{(i)}, \sigma^2)$$

- Resample: $\{(\tilde{x}_t, \tilde{\theta})^{(i)}\}_{i=1}^N$
- Propagate θ: $\theta^{(i)} \sim N(a\tilde{\theta}^{(i)} + (1-a)\bar{\theta}_t, (1-a^2)V_t)$
- Propagate x_{t+1}: $x_{t+1}^{(i)} \sim f_N(x_{t+1}; \tilde{\mu}^{(i)} + \tilde{\beta}^{(i)} \tilde{x}_t^{(i)}, \tau^2)$
- New weights:

$$w_{t+1}^{(i)} \propto \frac{f_N(y_{t+1}; x_{t+1}^{(i)}, \sigma^2)}{f_N(y_{t+1}; \tilde{\mu}^{(i)} + \tilde{\beta}^{(i)} \tilde{x}_t^{(i)}, \sigma^2)}$$

Particle Learning (PL)

- Posterior at t: $\left\{(x_t,\theta)^{(i)}\right\}_{i=1}^{N} \sim p(x_t,\theta|y^t)$

- Weights: $w_{t+1}^{(i)} \propto f_N(y_{t+1};\mu^{(i)}+\beta^{(i)}x_t^{(i)},\sigma^2+\tau^2)$
- Resample: $\{(\tilde{x}_t,\tilde{\theta})^{(i)}\}_{i=1}^{N}$

- Propagate states: $x_{t+1}^{(i)} \sim f_N(x_{t+1};\hat{x}_{t+1}^{(i)},A\sigma^2)$

  $\hat{x}_{t+1}^{(i)} = Ay_{t+1}+(1-A)(\tilde{\mu}^{(i)}+\tilde{\beta}^{(i)}\tilde{x}_t^{(i)})$ and $A=\tau^2/(\sigma^2+\tau^2)$

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

$$
\begin{aligned}
s_{1,t+1} &= s_{1t}+1\\
s_{2,t+1} &= s_{2t}+x_{t-1}\\
s_{3,t+1} &= s_{3t}+x_{t-1}^2\\
s_{4,t+1} &= s_{4t}+x_t^2\\
s_{5,t+1} &= s_{5t}+x_{t-1}x_t
\end{aligned}
$$

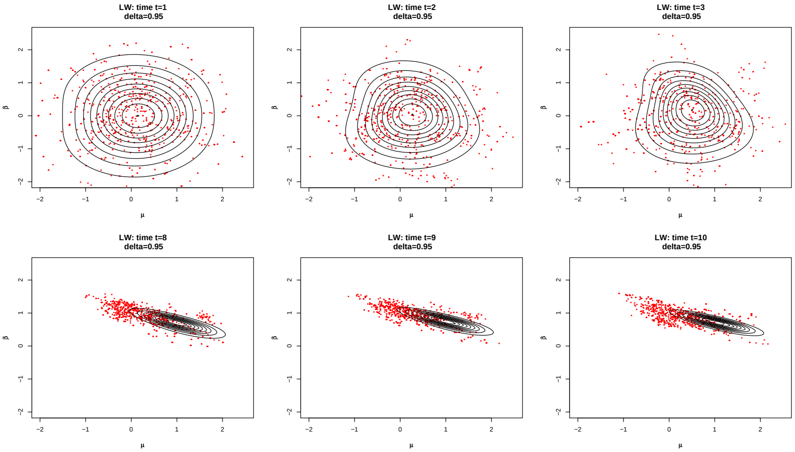
- Sample $\theta \sim p(\theta|s_{t+1})$

LWFA and APFSS

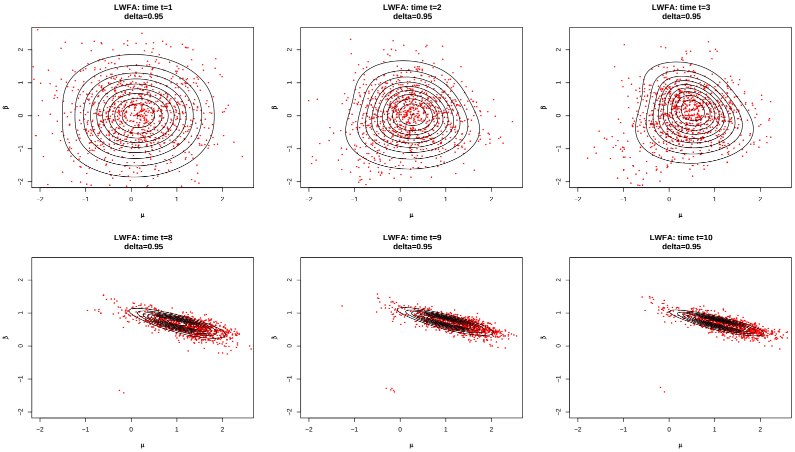
LWFA is LW fully adapted, i.e. with $f_N(y_{t+1}; \mu + \beta x_t, \sigma^2)$ replaced by $f_N(y_{t+1}; \mu + \beta x_t, \sigma^2 + \tau^2)$ and $f_N(x_{t+1}; \mu + \beta x_t, \tau^2)$ replaced by $f_N(x_{t+1}; Ay_{t+1} + (1 - A)(\mu + \beta x_t), A\sigma^2)$

APFSS is APF (blind resample and propagate) with recursive sufficient statistics for θ.

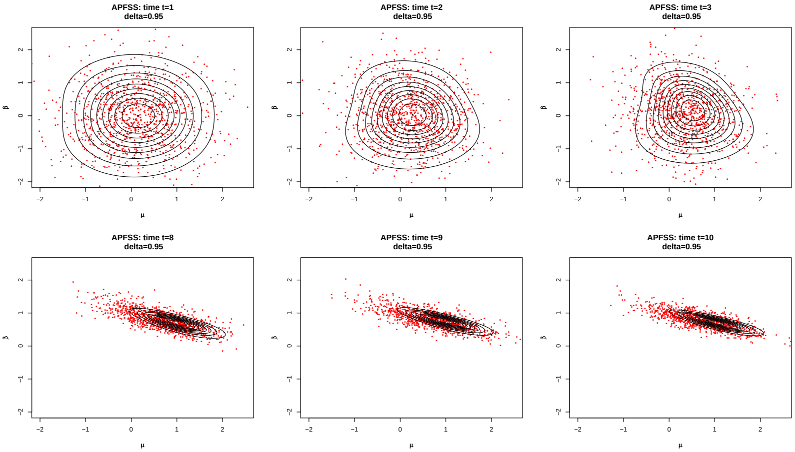
Liu and West (LW)



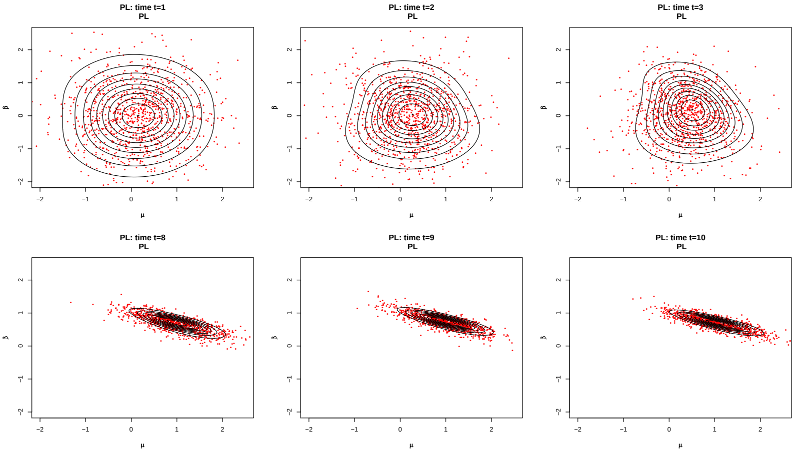
Liu and West + full adaptation (LWFA)



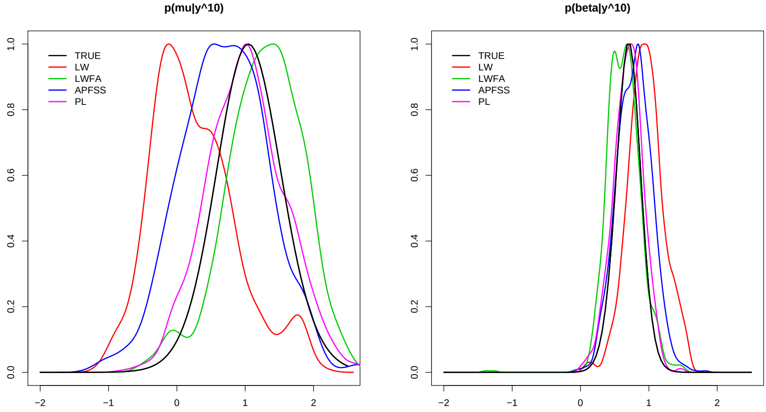
APF + sufficient statistics (APFSS)



Particle Learning (PL)



Comparison


p(mu|y^10)
TRUE
LW
LWFA
APFSS
PL
1.0
0.8
0.6
0.4
0.2
0.0
−2
−1
0
1
2
p(beta|y^10)
TRUE
LW
LWFA
APFSS
PL
1.0
0.8
0.6
0.4
0.2
0.0
−2
−1
0
1
2

SMC smoothers

SMC smoothers are alternatives to MCMC in state-space models.

Godsill, Doucet and West (2004) "Monte Carlo smoothing for non-linear time series" introduced an $O(TN^2)$ algorithm that relies on

- Forward particles, and
- Backward re-weighting via evolution equation.

See also Briers, Doucet and Maskell's (2009) "Smoothing Algorithms for State-Space Models" for other $O(TN^2)$ smoothers.

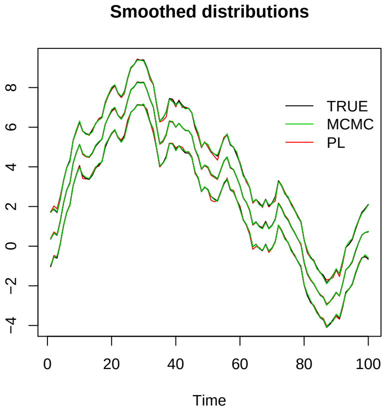
An $O(TN)$ smoothing algorithm is introduced by Fearnhead, Wyncoll and Tawn's (2008) "A sequential smoothing algorithm with linear computational cost".

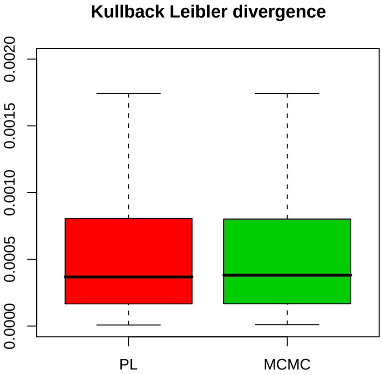
Example ii. PL and MCMC (filtering and smoothing)

$n = 100$, $\sigma^2 = 1, \tau^2 = 0.5$ and $x_0 = 0$ and $x_0 \sim N(0, 100)$.

$N = 1000$ PL particles

$M = 1000$ MCMC draws, after discarding the first $M_0 = 1000$.





Example ii. Computing time (in seconds)

$N = M_0 = M = 500$

n	PL	MCMC
100	9.3	4.7
200	18.8	9.1
500	47.7	23.4
1000	93.9	46.1

$n = 100$ and $N = M_0 = M$

N	PL	MCMC
500	9.3	4.7
1000	32.8	9.6
2000	127.7	21.7

MacBook 2.4GHz Intel Duo processor, 4GB 1067MHz memory.

Example iii. Sample-resample or PL?

Three time series of length $T = 1000$ were simulated from

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

with $x_0 = 0$ and (σ^2, τ^2) in $\{(0.1, 0.01), (0.01, 0.01), (0.01, 0.1)\}$. Throughout σ^2 is kept fixed.

The independent prior distributions for x_0 and τ^2 are $x_0 \sim N(m_0, V_0)$ and $\tau^2 \sim IG(a, b)$, for $a = 10$, $b = (a+1)\tau_0^2$, $m_0 = 0$ and $V_0 = 1$, where τ_0^2 is the true value of τ^2 for a given study.

We also include BBF in the comparison, for completion.

In all filters τ^2 is sampled offline from $p(\tau^2 | S_t)$ where S_t is the vector of conditional sufficient statistics.

Example iii. Mean absolute error

The three filters are rerun $R = 100$ times, all with the same seed within run, for each one of the three simulated data sets. Five different number of particles N were considered: 250, 500, 1000, 2000 and 5000.

Mean absolute errors (MAE) taken over the 100 replications are constructed by comparing percentiles of the true sequential distributions $p(x_t|y^t)$ and $p(\tau^2|y^t)$ to percentiles of the estimated sequential distributions $p_N(x_t|y^t)$ and $p_N(\tau^2|y^t)$.

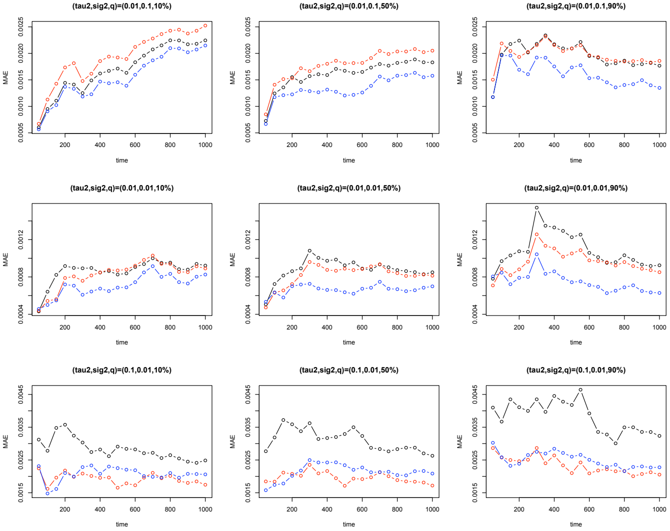
For $\alpha = 0.1, 0.5, 0.9$, true and estimated values of $q^x_{t,\alpha}$ and $q^{\tau^2}_{t,\alpha}$ were computed, for $Pr(x_t < q^x_{t,\alpha}|y^t) = Pr(\tau^2 < q^{\tau^2}_{t,\alpha}|y^t) = \alpha$.

For a in $\{x, \tau^2\}$ and α in $\{0.01, 0.50, 0.99\}$,

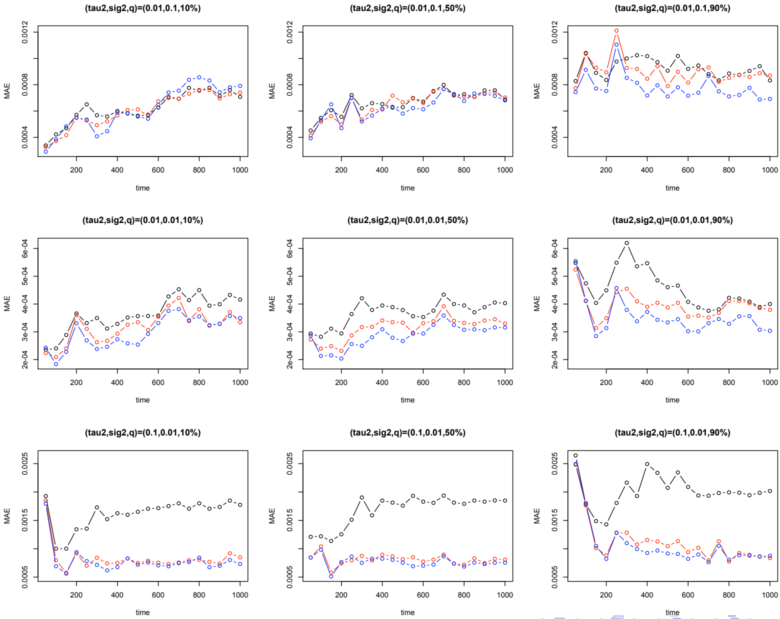
$$MAE^a_{t,\alpha} = \frac{1}{R}\sum_{r=1}^{R} |q^a_{t,\alpha} - \hat{q}^a_{t,\alpha,r}|$$

Example iii. $M = 500$ and learning τ^2.

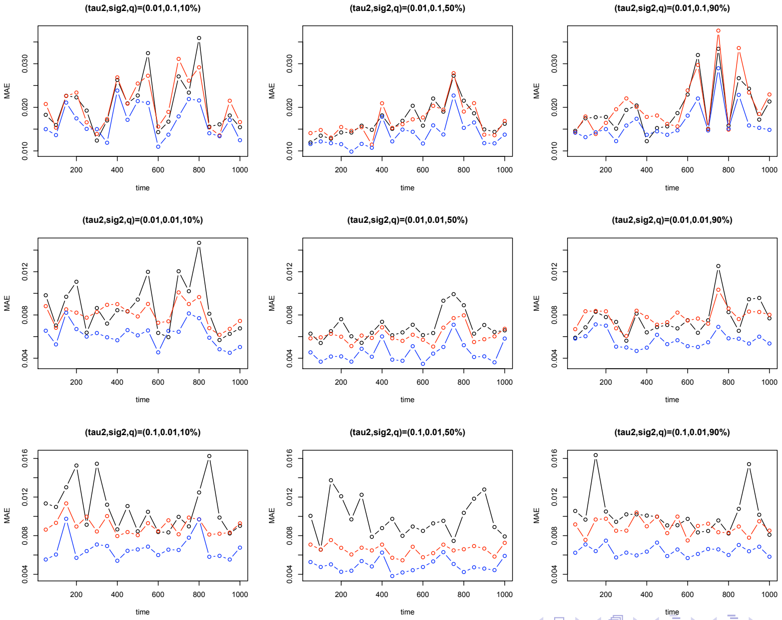
BBF,sample-resample,PL.



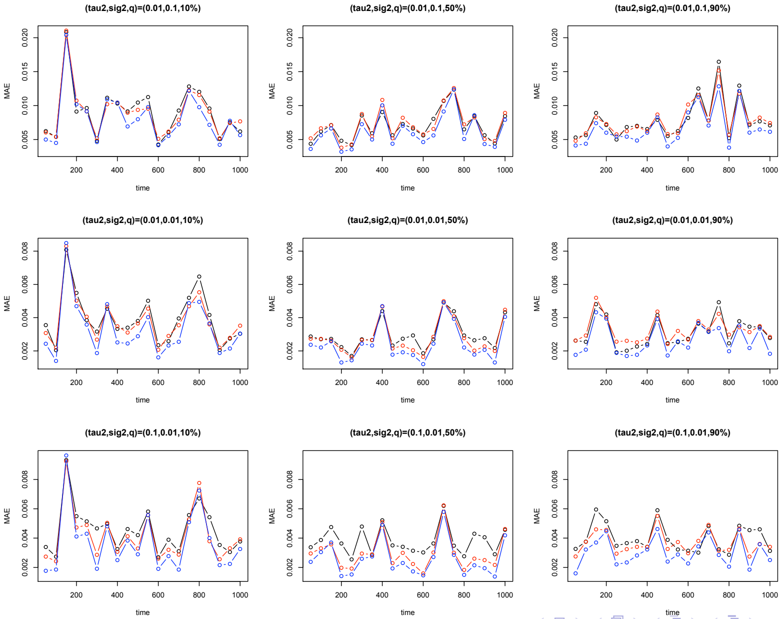
Example iii. $M = 5000$ and learning τ^2.



Example iii. $M = 500$ and learning x_t.



Example iii. $M = 5000$ and learning x_t.



Example iv. Computing sequential Bayes factors

A time series y_t is simulated from a *AR(1) plus noise* model:

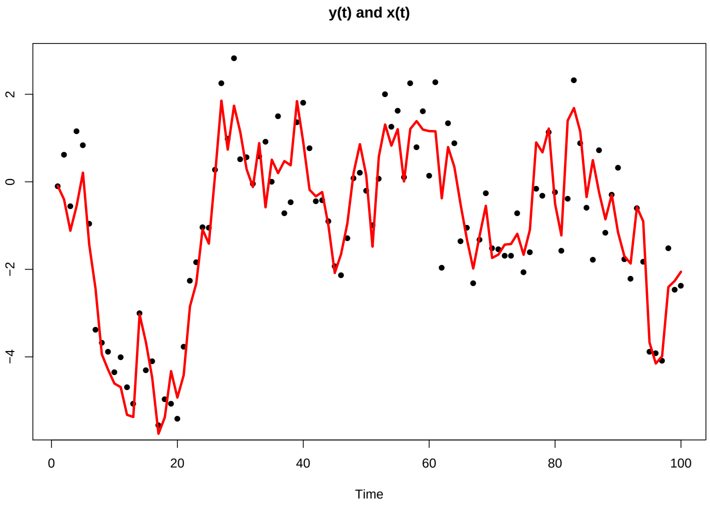
$$\begin{aligned}(y_{t+1}|x_{t+1},\theta) &\sim N(x_{t+1},\sigma^2)\\(x_{t+1}|x_t,\theta) &\sim N(\beta x_t,\tau^2)\end{aligned}$$

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

We set $T = 100$, $x_0 = 0$, $\theta = (\beta,\sigma^2,\tau^2) = (0.9, 1.0, 0.5)$.

σ^2 and τ^2 are kept known and the independent prior distributions for β and x_0 are both $N(0,1)$.

Example iv. Simulated data



Example iv. PL pure filter versus PL

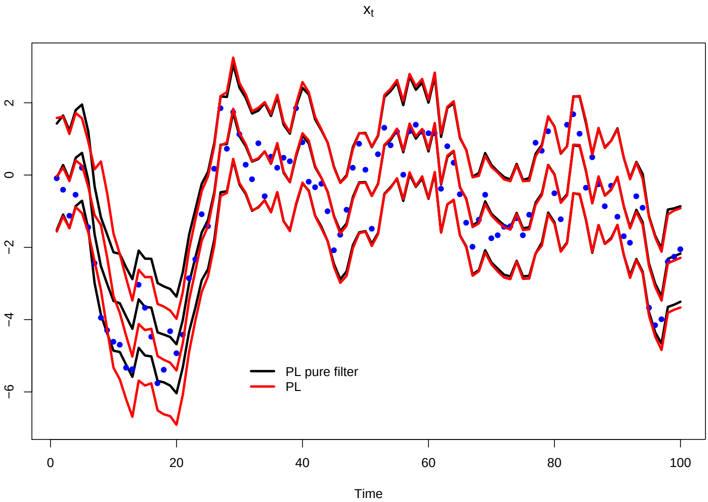
We run two filters:

- PL pure filter - our particle learning algorithm for learning x_t and keeping β fixed;
- PL - our particle learning algorithm for learning x_t and β sequentially.

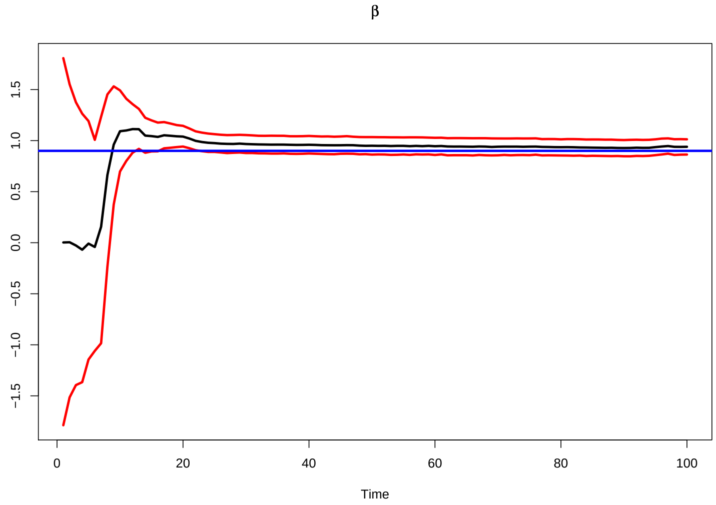
The filters are based on $N = 10,000$ particles.

Example iv. PL pure filter versus PL

β was fixed at the true value.

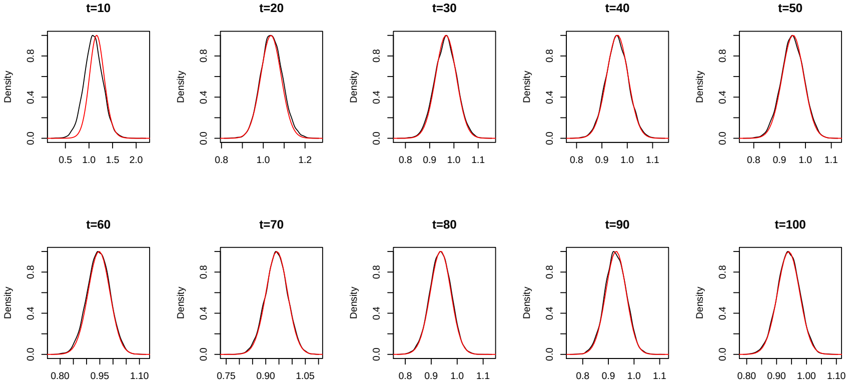


Example iv. PL - learning β

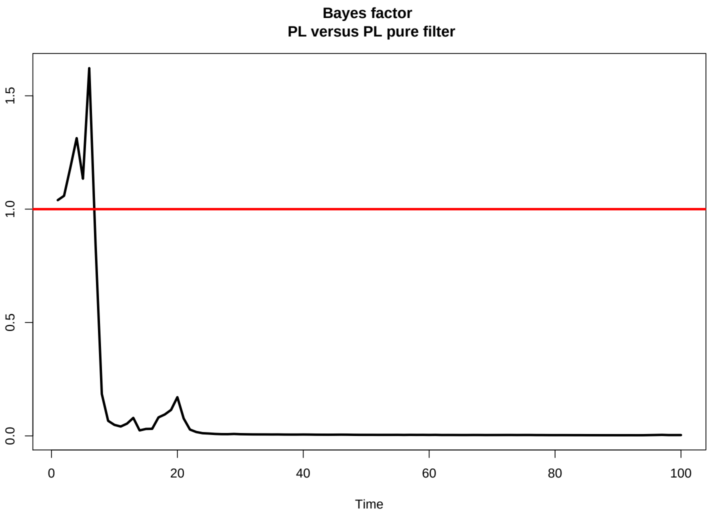


Example iv. PL - learning β

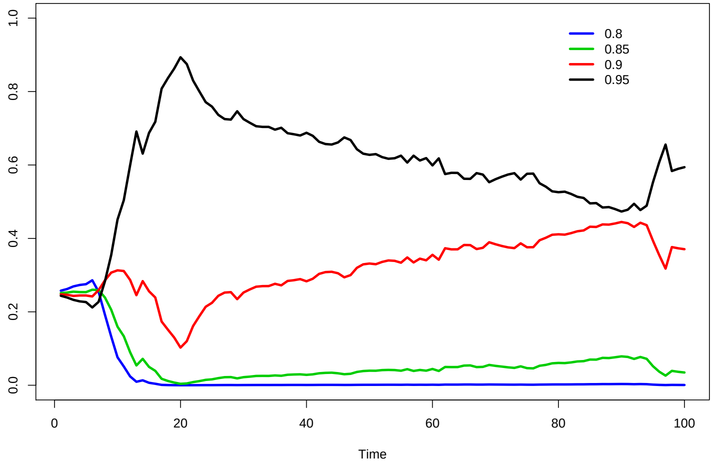
Comparing $p^N(\beta|y^t)$ with true $p(\beta|y^t)$.



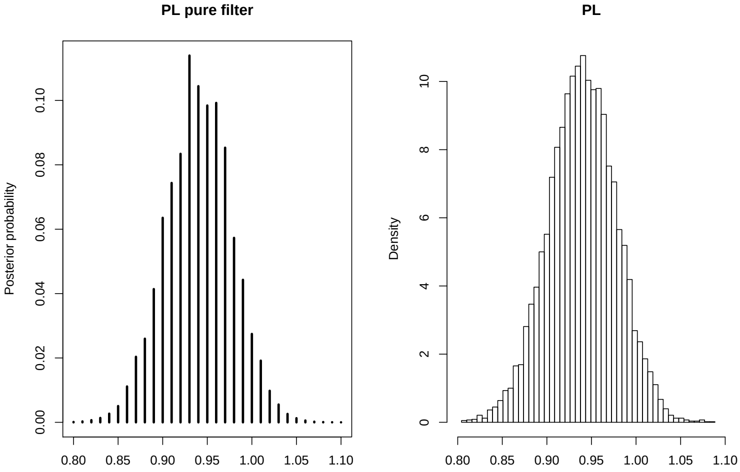
Example iv. Sequential Bayes factor



Example iv. Posterior model probabilities: 4 models



Example iv. Posterior model probabilities: 31 models



PL in Conditional Dynamic Linear Models (CDLM)

The model is

$$y_{t+1} = F_{\lambda_{t+1}} x_{t+1} + \epsilon_{t+1} \text{ where } \epsilon_{t+1} \sim \mathcal{N}\left(0, V_{\lambda_{t+1}}\right)$$
$$x_{t+1} = G_{\lambda_{t+1}} x_t + \epsilon^x_{t+1} \text{ where } \epsilon^x_{t+1} \sim \mathcal{N}\left(0, W_{\lambda_{t+1}}\right)$$

The error distribution

$$p\left(\epsilon_{t+1}\right) = \int \mathcal{N}\left(0, V_{\lambda_{t+1}}\right) p\left(\lambda_{t+1}\right) d\lambda_{t+1}$$

The augmented latent state is

$$\lambda_{t+1} \sim p(\lambda_{t+1}|\lambda_t)$$

PL extends Liu and Chen's (2000) "Mixture of Kalman Filters".

Algorithm

Step 1 (Re-sample): Generate an index $k(i) \sim Multi\left(w^{(i)}\right)$ where

$$w^{(i)} \propto p\left(y_{t+1} | (s_t^x, \theta)^{(i)}\right)$$

Step 2 (Propagate): States

$$\lambda_{t+1} \sim p\left(\lambda_{t+1} | \left(\lambda_t, \theta\right)^{k(i)}, y_{t+1}\right)$$
$$x_{t+1} \sim p(x_{t+1} | \left(x_t, \theta\right)^{k(i)}, \lambda_{t+1}, y_{t+1})$$

Step 3 (Propagate): Sufficient Statistics

$$s_{t+1}^x = \mathcal{K}\left(s_t^x, \theta, \lambda_{t+1}, y_{t+1}\right)$$
$$s_{t+1} = \mathcal{S}\left(s_t, x_{t+1}, \lambda_{t+1}, y_{t+1}\right)$$

Example A. Dynamic factor with switching loadings[15]

For $t = 1, \ldots, T$, the model is defined as follows:

- Observation equation

$$y_t | z_t, \theta \sim N(\gamma_t x_t, \sigma^2 I_2)$$

- State equations

$$\begin{aligned} x_t | x_{t-1}, \theta &\sim N(x_{t-1}, \sigma_x^2) \\ \lambda_t | \lambda_{t-1}, \theta &\sim Ber((1-p)^{1-\lambda_{t-1}} q^{\lambda_{t-1}}) \end{aligned}$$

where $z_t = (x_t, \lambda_t)'$.

Factor loadings: $\gamma_t = (1, \beta_{\lambda_t})'$.

Parameters: $\theta = (\beta_1, \beta_2, \sigma^2, \sigma_x^2, p, q)'$.

[15] Lopes and Carvalho (2007) and Lopes, Salazar and Gamerman (2008)

Example A. Conditionally conjugate prior

$$
\begin{aligned}
(\beta_i|\sigma^2) &\sim N\left(b_{i0}, \sigma^2 B_{i0}\right) \qquad \text{for } i = 1, 2, \\
\sigma^2 &\sim IG\left(\frac{\nu_{00}}{2}, \frac{d_{00}}{2}\right) \\
\sigma_x^2 &\sim IG\left(\frac{\nu_{10}}{2}, \frac{d_{10}}{2}\right) \\
p &\sim Beta(p_1, p_2) \\
q &\sim Beta(q_1, q_2) \\
x_0 &\sim N(m_0, C_0)
\end{aligned}
$$

Example A. Particle representation

At time t, particles

$$\left\{ (x_t, \lambda_t, \theta, s_t^x, s_t)^{(i)} \right\}_{i=1}^{N}$$

approximating

$$p\left(x_t, \lambda_t, \theta, s_t^x, s_t | y^t\right)$$

where

- $s_t^x = \mathcal{S}(s_{t-1}^x, \theta)$ are state sufficient statistics
- $s_t = \mathcal{S}(s_{t-1}, x_t, \lambda_t)$ are fixed parameter sufficient statistics

Example A. Re-sampling $(x_t, \lambda_t, \theta, s_t^x, s_t)$

Let us redefine $\beta_i = (1, \beta_i)'$ whenever necessary.

Draw an index $k(i) \sim Multi(\omega^{(i)})$ with weights

$$\omega^{(i)} \propto p(y_{t+1} | (s_t^x, \lambda_t, \theta)^{k(i)})$$

with

$$p(y_{t+1}|m_t, C_t, \lambda_t, \theta) = \sum_{j=1}^{2} f_N \left(y_{t+1}; \beta_j m_t, V_j\right) Pr \left(\lambda_{t+1} = j | \lambda_t, \theta\right)$$

where $V_j = (C_t + \sigma_x^2)\beta_j\beta_j' + \sigma^2 I_2$, m_t and C_t are components of s_t^x and f_N denotes the normal density function.

Example A. Propagating states

Draw auxiliary state λ_{t+1}

$$\lambda_{t+1}^{(i)} \sim p(\lambda_{t+1} | (s_t^x, \lambda_t, \theta)^{k(i)}, y_{t+1})$$

where

$$Pr(\lambda_{t+1} = j | s_t^x, \lambda_t, \theta, y_{t+1}) \propto f_N\left(y_{t+1}; \beta_j m_t, V_j\right) p\left(\lambda_{t+1} = j | \lambda_t, \theta\right).$$

Draw state x_{t+1} conditionally on λ_{t+1}

$$x_{t+1}^{(i)} \sim p(x_{t+1} | \lambda_{t+1}^{(i)}, (s_t^x, \theta)^{k(i)}, y_{t+1})$$

by a simply Kalman filter update.

Example A. Updating sufficient statistics for states, s^x_{t+1}

The Kalman filter recursion yield

$$m_{t+1} = m_t + A_{t+1}(y_{t+1} - \beta_{\lambda_{t+1}} m_t)$$
$$C_{t+1} = C_t + \sigma_x^2 - A_{t+1} Q_{t+1}^{-1} A'_{t+1}$$

where

$$Q_{t+1} = (C_t + \sigma_x^2)\gamma_{t+1}\gamma'_{t+1} + \sigma^2 I_2$$
$$A_{t+1} = (C_t + \sigma_x^2)\gamma'_{t+1} Q_{t+1}^{-1}$$

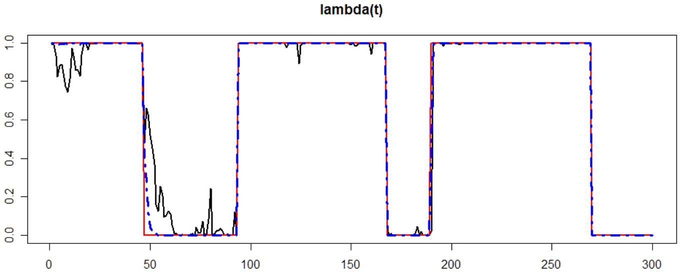
Example A. Updating suff. statistics for parameters, s_{t+1}

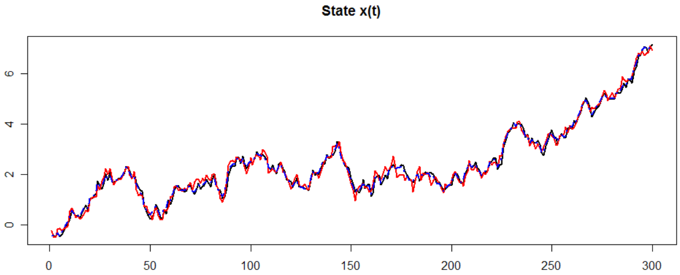
Recall that $s_{t+1} = \mathcal{S}(s_t, x_{t+1}, \lambda_{t+1})$. Then,

$$
\begin{aligned}
(\beta_i|\sigma^2, s_{t+1}) &\sim N\left(b_{i,t+1}, \sigma^2 B_{i,t+1}\right) \qquad \text{for } i = 1,2,\\
(\sigma^2|s_{t+1}) &\sim IG\left(\frac{\nu_{0t}}{2}, \frac{d_{0,t+1}}{2}\right)\\
(\sigma_x^2|s_{t+1}) &\sim IG\left(\frac{\nu_{1t}}{2}, \frac{d_{1,t+1}}{2}\right)\\
(p|s_{t+1}) &\sim Beta(p_{1,t+1}, p_{2,t+1})\\
(q|s_{t+1}) &\sim Beta(q_{1,t+1}, q_{2,t+1})
\end{aligned}
$$

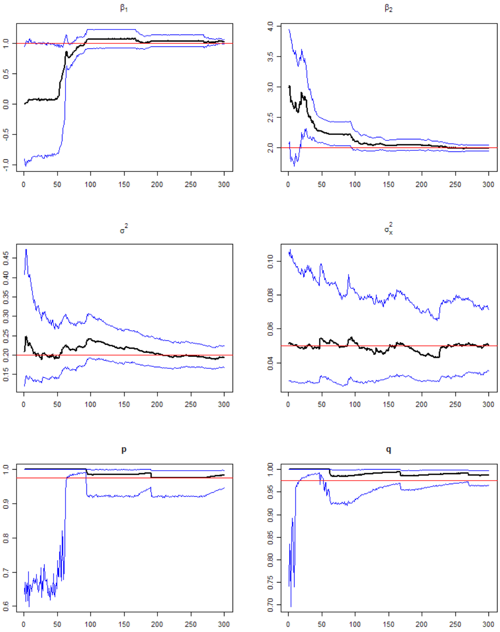
where $\mathbb{I}_{\lambda_{t+1}=i} = \mathbb{I}_i$, $\mathbb{I}_{\lambda_t=i,\lambda_{t+1}=j} = \mathbb{I}_{ij}$, $\nu_{it} = \nu_{i,t-1} + 1$, $B_{i,t+1}^{-1} = B_{it}^{-1} + x_{t+1}^2$, $B_{i,t+1}^{-1} b_{i,t+1} = B_{it}^{-1} b_{it} + x_{t+1} y_{t+1,2} \mathbb{I}_i$, $p_{i,t+1} = p_{it} + \mathbb{I}_{1i}$ (similarly for $q_{i,t+1}$) for $i = 1,2$, $d_{0,t+1} = d_{0,t} + (y_{t+1,1} - x_{t+1})^2 + \sum_{j=1}^2 \left[(y_{t+1,2} - b_{j,t+1} x_{t+1})\, y_{t+1,2} + B_{j,t+1}^{-1} b_{j,t+1}\right] \mathbb{I}_j$, and $d_{1,t+1} = d_{1,t} + (x_{t+1} - x_t)^2$.

Example A. Filtering and smoothing for states





Example A. Sequential parameter learning



PL in (state) non-linear normal dynamic models

The model now is

$$
\begin{aligned}
y_{t+1} &= F_{\lambda_{t+1}} x_{t+1} + \epsilon_{t+1} \quad \text{where} \quad \epsilon_{t+1} \sim \mathcal{N}\left(0, V_{\lambda_{t+1}}\right) \\
x_{t+1} &= G_{\lambda_{t+1}} Z\left(x_t\right) + \omega_{t+1} \quad \text{where} \quad \omega_{t+1} \sim \mathcal{N}\left(0, W_{\lambda_{t+1}}\right)
\end{aligned}
$$

where ϵ_{t+1} and λ_{t+1} are modeled as before.

Algorithm:

Step 1 (Re-sample): Generate an index $k(i) \sim Multi\left(w^{(i)}\right)$ where

$$
w^{(i)} \propto p\left(y_{t+1} | (x_t, \theta)^{(i)}\right)
$$

Step 2 (Propagate):

$$
\begin{aligned}
\lambda_{t+1} &\sim p\left(\lambda_{t+1} \middle| \left(\lambda_t, \theta\right)^{k(i)}, y_{t+1}\right) \\
x_{t+1} &\sim p\left(x_{t+1} \middle| \left(x_t, , \theta\right)^{k(i)}, \lambda_{t+1}, y_{t+1}\right) \\
s_{t+1} &= \mathcal{S}\left(s_t, x_{t+1}, \lambda_{t+1}, y_{t+1}\right)
\end{aligned}
$$

Example B. Fat-tailed nonlinear model[16]

Let

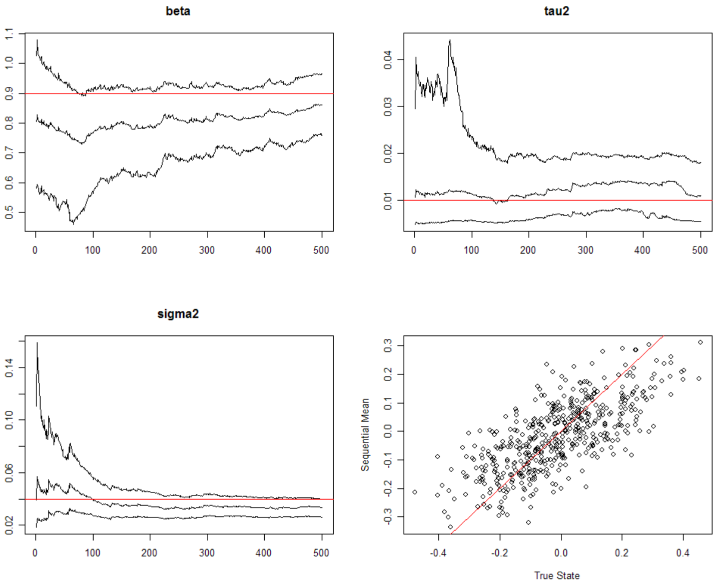
$$y_{t+1} = x_{t+1} + \sigma\sqrt{\lambda_{t+1}}\epsilon_{t+1} \;\text{ where }\; \lambda_{t+1} \sim \mathcal{IG}\left(\frac{\nu}{2}, \frac{\nu}{2}\right)$$
$$x_{t+1} = g(x_t)\beta + \sigma_x u_{t+1} \qquad \text{where }\; g(x_t) = \frac{x_t}{1+x_t^2}$$

where ϵ_{t+1} and u_{t+1} are independent standard normals and ν is known.

The observation error term is non-normal $\sqrt{\lambda_{t+1}}\epsilon_{t+1} \sim t_\nu$.

[16] From deJong et al (2007)

Example B. Sequential inference



Example C. Dynamic multinomial logit model

Let us study the multinomial logit model[17]

$$P\left(y_{t+1}=1|\beta_{t+1}\right)=\frac{e^{F_t\beta_t}}{1+e^{F_t\beta_t}} \text{ and } \beta_{t+1} \quad =\phi\beta_t+\sigma_x\epsilon^{\beta}_{t+1}$$

where $\beta_0 \sim N(0,\sigma^2/(1-\rho^2))$. Scott's (2007) data augmentation structure leads to a mixture Kalman filter model

$$\begin{aligned} y_{t+1} &= \mathbb{I}(z_t \geq 0) \\ z_{t+1} &= Z_t\beta+\epsilon_{t+1} \text{ where } \epsilon_{t+1}\sim -ln\mathcal{E}(1) \end{aligned}$$

Here ϵ_t is an extreme value distribution of type 1 where $\mathcal{E}(1)$ is an exponential of mean one. The key is that it is easily to simulate $p(z_t|\beta,y_t)$ using

$$z_{t+1}=-\ln\left(\frac{\ln U_i}{1+e^{\beta_i\beta}}-\frac{\ln V_i}{e^{\beta_i\beta}}\mathcal{I}_{y_{t+1}=0}\right)$$

[17] Carvalho, Lopes and Polson (2008)

Example C. 10-component mixture of normals

Frunwirth-Schnatter and Schnatter (2007) uses a 10-component mixture of normals:

$$p(\epsilon_t) = e^{-\epsilon_t - e^{-\epsilon_t}} \approx \sum_{j=1}^{10} w_j \mathcal{N}(\mu_j, s_j^2)$$

Hence conditional on an indicator λ_t we can analyze

$$y_t = \mathbb{I}(z_t \geq 0) \quad \text{and} \quad z_t = \mu_{\lambda_t} + Z_t \beta + s_{\lambda_t} \epsilon_t$$

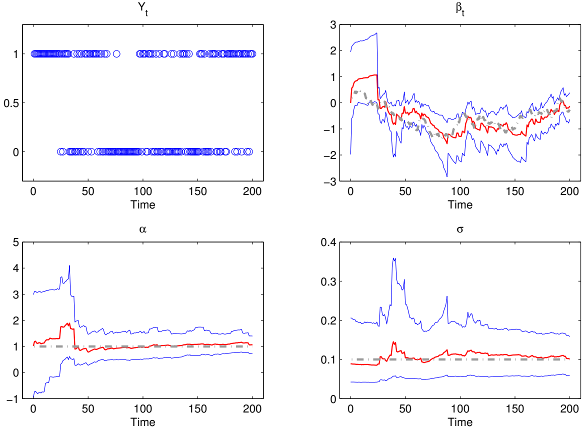
where $\epsilon_t \sim N(0,1)$ and $Pr(\lambda_t = j) = w_j$. Also,

$$\begin{aligned} s_{t+1}^{\beta} &= \mathcal{K}\left(s_t^{\beta}, z_{t+1}, \lambda_{t+1}, \theta, y_{t+1}\right) \\ p(y_{t+1}|s_t^{\beta}, \theta) &= \sum_{\lambda_{t+1}} p(y_{t+1}|s_t^{\beta}, \lambda_{t+1}, \theta) \end{aligned}$$

for re-sampling. Propagation now requires

$$\begin{aligned} \lambda_{t+1} &\sim p\left(\lambda_{t+1}|(s_t^{\beta}, \theta)^{k(i)}, y_{t+1}\right) \\ z_{t+1} &\sim p\left(z_{t+1}|(s_t^{\beta}, \theta)^{k(i)}, \lambda_{t+1}, y_{t+1}\right) \\ \beta_{t+1} &\sim p\left(z_{t+1}|(s_t^{\beta}, \theta)^{k(i)}, \lambda_{t+1}, z_{t+1}\right) \end{aligned}$$

Example C. Simulated exercise



PL based on 30,000 particles.

Example D. Sequential Bayesian Lasso

We develop a sequential version of Bayesian Lasso[18] for a simple problem of signal detection. The model takes the form

$$\begin{aligned} (y_t|\theta_t) &\sim N(\theta_t, 1) \\ p(\theta_t|\tau) &= (2\tau)^{-1} \exp\left(-|\theta_t|/\tau\right) \end{aligned}$$

for $t = 1, \ldots, n$ and $\tau^2 \sim IG(a_0, b_0)$.

Data augmentation: It is easy to see that

$$p(\theta_t|\tau) = \int p(\theta_t|\tau, \lambda_t) p(\lambda_t) d\lambda_t$$

where

$$\begin{aligned} \lambda_t &\sim Exp(2) \\ \theta_t|\tau, \lambda_t &\sim N(0, \tau^2 \lambda_t) \end{aligned}$$

[18] Carlin and Polson (1991) and Hans (2008)

Example D. Data augmentation

The natural set of latent variables is given by the augmentation variable λ_{n+1} and conditional sufficient statistics leading to

$$Z_n = (\lambda_{n+1}, a_n, b_n)$$

The sequence of variables λ_{n+1} are i.i.d. and so can be propagated directly with $p(\lambda_{n+1})$.

The conditional sufficient statistics (a_{n+1}, b_{n+1}) are deterministically determined based on parameters $(\theta_{n+1}, \lambda_{n+1})$ and previous values (a_n, b_n).

Example D. PL algorithm

1. After n observations: $\left\{(Z_n, \tau)^{(i)}\right\}_{i=1}^{N}$.
2. Draw $\lambda_{n+1}^{(i)} \sim Exp(2)$.
3. Resample old particles with weights

$$w_{n+1}^{(i)} \propto p(y_{n+1}; 0, 1+\tau^{2(i)}\lambda_{n+1}^{(i)}).$$

4. Sample $\theta_{n+1}^{(i)} \sim N(m_n^{(i)}, C_n^{(i)})$, where $m_n^{(i)} = C_n^{(i)} y_{n+1}$ and $C_n^{-1} = 1 + \tilde{\tau}^{-2(i)} \tilde{\lambda}_{n+1}^{-1(i)}$.
5. Suff. stats: $a_{n+1}^{(i)} = \tilde{a}_n^{(i)} + 1/2$, $b_{n+1}^{(i)} = \tilde{b}_n^{(i)} + \theta_{n+1}^{2(i)}/(2\tilde{\lambda}_{n+1}^{(i)})$.
6. Sample (offline) $\tau^{2(i)} \sim IG(a_{n+1}, b_{n+1})$.
7. Let $Z_{n+1}^{(i)} = (\lambda_{n+1}^{(i)}, a_{n+1}^{(i)}, b_{n+1}^{(i)})$.
8. After $n+1$ observations: $\left\{(Z_{n+1}, \tau)^{(i)}\right\}_{i=1}^{N}$.

Example D. Sequential Bayes factor

As the Lasso is a model for sparsity we would expect the evidence for it to increase when we observe $y_t = 0$.

We can sequentially estimate $p(y_{n+1} \mid y^n, \text{lasso})$ via

$$p(y_{n+1} \mid y^n, \text{lasso}) = \frac{1}{N} \sum_{i=1}^{N} p(y_{n+1} \mid (\lambda_n, \tau)^{(i)})$$

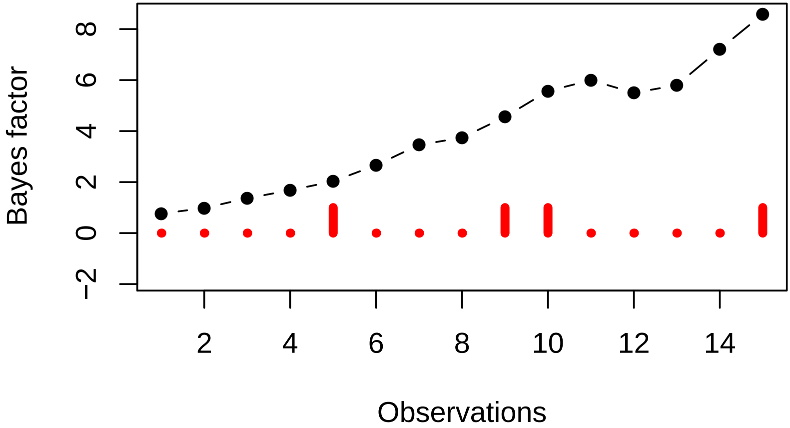
with predictive $p(y_{n+1} \mid \lambda_n, \tau) \sim N(0, \tau^2 \lambda_n + 1)$.

This leads to a sequential Bayes factor

$$BF_{n+1} = \frac{p(y^{n+1} \mid \text{lasso})}{p(y^{n+1} \mid \text{normal})}.$$

Example D. Simulated data

Data based on $\theta = (0,0,0,0,1,0,0,0,1,1,0,0,0,0,1)$ and priors $\tau^2 \sim IG(2,1)$ for the double exponential case and $\tau^2 \sim IG(2,3)$ for the normal case, reflecting the ratio of variances between those two distributions.



Final remarks

PL is a general framework for sequential Bayesian inference in dynamic and static models.

PL is able to deal with filtering and learning and reduce the accumulation of error.

The loose definition of sufficient statistics and the flexibility to freely augment x_t makes PL a competitive alternative to MCMC in highly structured models.

A powerful by-product of PL (and SMC in general) over MCMC schemes, is its ability to sequentially produce model comparison, assessment indicators.

SAMSI effect

Papers with Carvalho, Dukic, Johannes, Polson, Prado, Taddy and Tsay:

- Particle Learning for Sequential Bayesian Computation;
- Particle learning for general mixtures;
- Sequential parameter learning and filtering in structured AR models.
- Tracking flu epidemics using Google trends and particle learning;
- Bayesian analysis of financial time series via particle filters;
- Bayesian statistics with a smile: a resampling-sampling perspective.

Papers with Chen, Lund, Macaro, Petralia and Rios:

- SMC Methods for LMSV Models;
- SMC Estimation of DSGE Models;
- Learning in a Regime Switching Macro-Finance Nelson Siegel Model;
- PL in Markov-switching SV models.

Textbooks with chapters on Particle Learning or SMC:

- Petris, Petrone and Campagnoli (2009) DLMs with R. Springer.
- Prado and West (2010) Time Series: Modeling, Inference and Forecasting. Chapman & Hall/CRC.
- Lopes and McCulloch (2012) A First Course in Bayesian Econometrics. Wiley.

Basic references

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.

Carvalho, Johannes, Lopes and Polson (2008) Particle Learning and Smoothing. Technical Report. The University of Chicago Booth School of Business.

Eraker, Johannes and Polson (2003) The Impact of Jumps in Volatility and Returns. *Journal of Finance*, 58, 1269-1300.

Gamerman and Lopes (2006) *MCMC: Stochastic Simulation for Bayesian Inference*. Baton Rouge: Chapman & Hall/CRC.

Jacquier, Polson and Rossi (1994) Bayesian Analysis of Stochastic Volatility Models. *Journal of Business and Economic Statistics*, 12, 371-89.

Johannes and Polson (2009) MCMC methods for Financial Econometrics. In *Handbook of Financial Econometrics* (Eds Y. Ait-Sahalia and L. Hansen). Oxford: Elsevier, 1-72.
Johannes, Polson and Stroud (2009) Optimal Filtering of Jump Diffusions: Extracting Latent States from Asset Prices. *Review of Financial Studies*, 22, 2559-2599.

Kim, Shephard and Chib (1994) Stochastic Volatility: Likelihood Inference and Comparison with ARCH Models. *Review of Economic Studies*, 65, 361-393.

Liu and West (2001) Combined parameters and state estimation in simulation-based filtering. In *Sequential Monte Carlo Methods in Practice* (Eds. A. Doucet, N. de Freitas and N. Gordon). New York: Springer-Verla, 197-223.

Basic references (cont.)

Gordon, Salmond and Smith (1993) Novel approach to nonlinear/non-Gaussian Bayesian state estimation. *Radar and Signal Processing, IEE Proceedings F 140*, 107-113.

Migon, Gamerman, Lopes and Ferreira (2005) Dynamic models, In *Handbook of Statistics, Volume 25: Bayesian Thinking, Modeling and Computation* (Eds. D. Dey and C. R. Rao), Amsterdam: Elsevier, 553-588.

Petris, Petrone and Campagnoli (2009) *Dynamic Linear Models with R*. New York: Springer.

Pitt and Shephard (1999) Filtering via simulation: auxiliary particle filters. *Journal of the American Statistical Association*, 94, 590-599.

Polson, Lopes and Carvalho (2009) Bayesian Statistics with a Smile: a Resampling-Sampling Perspective. Technical Report. The University of Chicago Booth School of Business.

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

Prado and West (2010) *Time Series: Modelling, Computation and Inference*. Baton Rouge: Chapman & Hall/CRC.

Storvik (2002) Particle filters in state space models with the presence of unknown static parameters. *IEEE Transactions of Signal Processing*, 50, 281-289.

West and Harrison (1997) *Bayesian Forecasting and Dynamic Models (2nd edition)*. New York: Springer-Verlag.