Lecture 4: Dynamic models

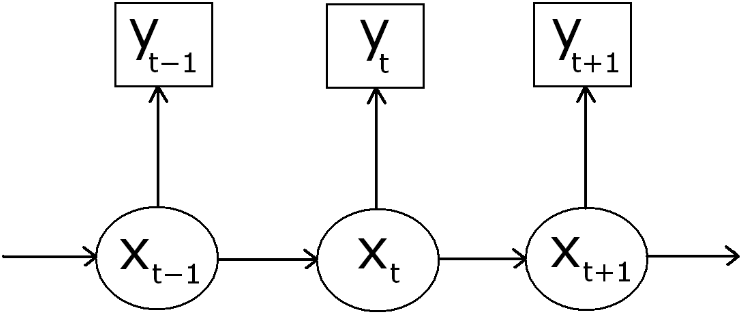
Hedibert Freitas Lopes

The University of Chicago Booth School of Business
5807 South Woodlawn Avenue, Chicago, IL 60637
http://faculty.chicagobooth.edu/hedibert.lopes

hlopes@ChicagoBooth.edu

Outline

Dynamic models (DMs)



1st order DLM

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

Observation equation:

$$y_{t+1}|x_{t+1}, \theta \sim N(x_{t+1}, \sigma^2)$$

System equation:

$$x_{t+1}|x_t, \theta \sim N(x_t, \tau^2)$$

where

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

and

$$\theta = (\sigma^2, \tau^2)$$

fixed (for now).

n-variate normal

It is worth noticing that the model can be rewritten as

$$\begin{aligned} y|x,\theta &\sim N(x,\sigma_2 I_n) \\ x|x_0,\theta &\sim N(x_0 1_n,\tau^2\Omega) \\ x_0 &\sim N(m_0,C_0) \end{aligned}$$

where

$$\Omega = \begin{pmatrix} 1 & 1 & 1 & 1 & \dots & 1 & 1 & 1 \\ 1 & 2 & 2 & 2 & \dots & 2 & 2 & 2 \\ 1 & 2 & 3 & 3 & \dots & 3 & 3 & 3 \\ 1 & 2 & 3 & 4 & \dots & 4 & 4 & 4 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 1 & 2 & 3 & 4 & \dots & n-2 & n-2 & n-2 \\ 1 & 2 & 3 & 4 & \dots & n-1 & n-1 & n-1 \\ 1 & 2 & 3 & 4 & \dots & n-2 & n-1 & n \end{pmatrix}$$

Therefore, the prior of x given θ is

$$x|\theta \sim N(m_0 1_n; C_0 1_n 1_n' + \tau^2 \Omega),$$

while its full conditional posterior distribution is

$$x|y,\theta \sim N(m_1, C_1)$$

where

$$C_1^{-1} = (C_0 1_n 1_n' + \tau^2 \Omega)^{-1} + \sigma^{-2} I_n$$

and

$$C_1^{-1} m_1 = (C_0 1_n 1_n' + \tau^2 \Omega)^{-1} m_0 1_n + \sigma^{-2} y$$

The Kalman filter

Let $y^t = (y_1, \ldots, y_t)$. The previous joint posterior posterior for x given y (omitting θ for now) can be constructed as

$$p(x|y^n) = p(x_1|y^n, x_2) \prod_{t=1}^{n} p(x_t|y^n, x_{t+1}),$$

which is obtained from

$$p(x^n|y^n)$$

and noticing that given y^t and x_{t+1},

- x_t and x_{t+h} are independent, and
- x_t and y_t are independent,

for all integer $h > 1$.

Therefore, we first need to derive the above joint and this is done forward via the well-known Kalman filter recursions.

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

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

$$R_{t+1} = C_t + \tau^2$$

- Marginal likelihood: $(y_{t+1}|y^t) \sim N(m_t, Q_{t+1})$

$$Q_{t+1} = R_{t+1} + \sigma^2$$

- Posterior at $t+1$: $(x_{t+1}|y^{t+1}) \sim N(m_{t+1}, C_{t+1})$

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

where $A_{t+1} = R_{t+1}/Q_{t+1}$.

The Kalman smoother

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

For $t < n$,

$$\begin{aligned} x_t | y^n &\sim N(m_t^n, C_t^n) \\ x_t | x_{t+1}, y^n &\sim N(a_t^n, R_t^n) \end{aligned}$$

where

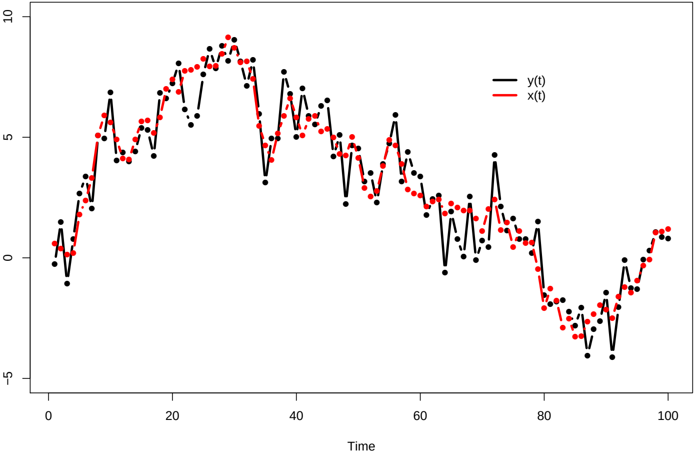
$$\begin{aligned} m_t^n &= (1 - B_t) m_t + B_t m_{t+1}^n \\ C_t^n &= (1 - B_t) C_t + B_t^2 C_{t+1}^n \\ a_t^n &= (1 - B_t) m_t + B_t x_{t+1} \\ R_t^n &= B_t \tau^2 \end{aligned}$$

and

$$B_t = C_t / (C_t + \tau^2).$$

Example

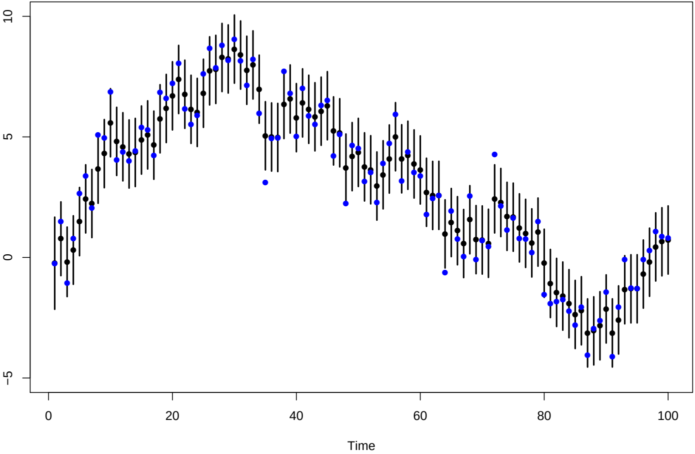
$n = 100$, $\sigma^2 = 1.0$
$\tau^2 = 0.5$ and $x_0 = 0$.



$p(x_t|y^t)$ via Kalman filter

$m_0 = 0.0$ and $C_0 = 10.0$

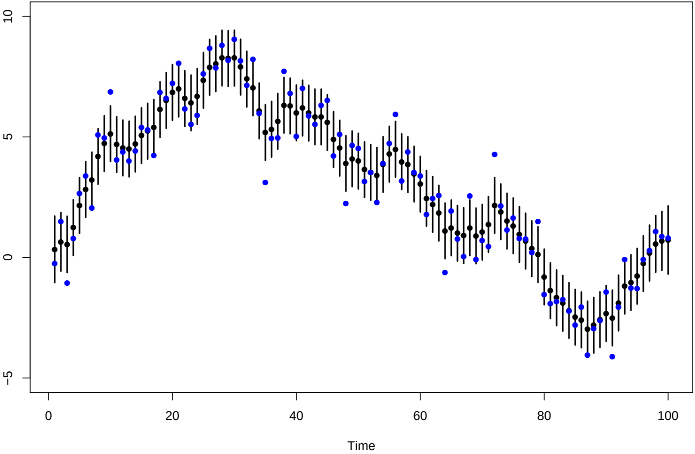
given τ^2 and σ^2



$p(x_t|y^n)$ via Kalman smoother

$m_0 = 0.0$ and $C_0 = 10.0$

given τ^2 and σ^2



Integrating out states x^n

We showed earlier that

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

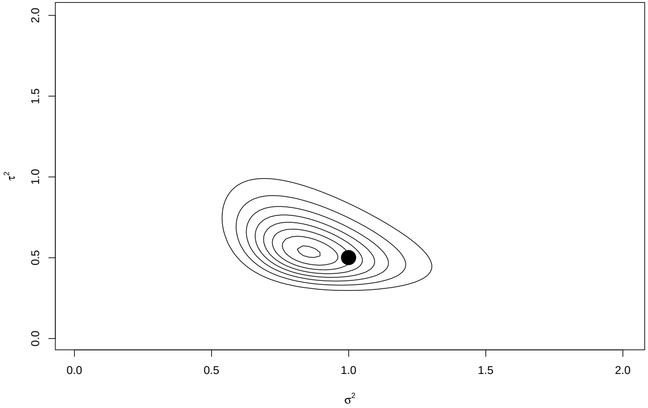
where both m_{t-1} and Q_t were presented before and are functions of $\theta = (\sigma^2, \tau^2)$, y^{t-1}, m_0 and C_0.

Therefore, by Bayes' rule,

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

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

$\sigma^2 \sim IG(\nu_0/2, \nu_0\sigma_0^2/2)$, where $\nu_0 = 5$ and $\sigma_0^2 = 1$.
$\tau^2 \sim IG(n_0/2, n_0\tau_0^2/2)$, where $n_0 = 5$ and $\tau_0^2 = 0.5$



MCMC scheme

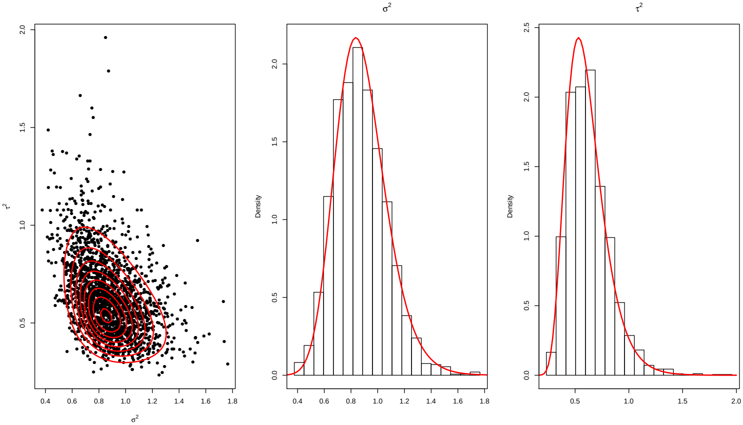
- Sample θ from $p(\theta|y^n, x^n)$

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

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

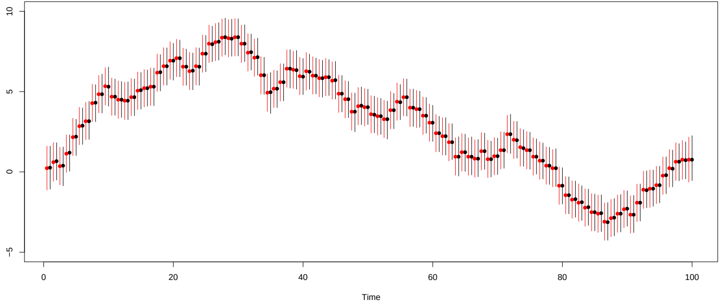
$$p(x^n|y^n, \theta) = \prod_{t=1}^{n} f_N(x_t|a_t^n, R_t^n)$$

Example: $p(x_t|y^n)$



Example: Comparison

$p(x_t|y^n)$ versus $p(x_t|y^n, \tilde{\sigma}^2 = 0.87, \tilde{\tau}^2 = 0.63)$.



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 (DLMs)

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;
- $\beta_1 \sim N(a, R)$.

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

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

Full conditionals

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

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

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

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 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.

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 $n \longrightarrow \infty$,

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

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.

Schemes

Scheme I: Sampling $\beta_1, \ldots, \beta_n$, V, 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.

		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 n_{eff} based on V.

Nonlinear DM

Let y_t, for $t = 1, \ldots, n$, be generated by

$$
\begin{aligned}
y_t &= \frac{x_t^2}{20} + \epsilon_t & \epsilon_t &\sim N(0, \sigma^2) \\
x_t &= \alpha x_{t-1} + \beta \frac{x_{t-1}}{1 + x_{t-1}^2} + \gamma \cos(1.2(t-1)) + u_t & u_t &\sim N(0, \tau^2)
\end{aligned}
$$

where $x_0 \sim N(m_0, C_0)$ and $\theta = (\alpha, \beta, \gamma)'$.

Prior distribution

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

Sampling $(\sigma^2, \theta, \tau^2 | x_0, x^n, y^n)$

It follows that

$$(\sigma^2 | y^n, x^n) \sim IG(n_1/2, n_1\sigma_1^2/2)$$

where $n_1 = n_0 + n$ and

$$n_1\sigma_1^2 = n_0\sigma_0^2 + \sum_{t=1}^{n}(y_t - x_t^2/20)^2.$$

Also

$$(\theta, \tau^2 | x_{0:n}) \sim N(\theta_1, \tau^2 V_1) IG(\nu_1/2, \nu_1\tau_1^2/2)$$

where $\nu_1 = \nu_0 + n$,

$$\begin{aligned}
V_1^{-1} &= V_0^{-1} + Z'Z \\
V_1^{-1}\theta_1 &= V_0^{-1}\theta_0 + Z'x_{1:n} \\
\nu_1\tau_1^2 &= \nu_0\tau_0^2 + (y - Z\theta_1)'(y - Z\theta_1) + (\theta_1 - \theta_0)'V_0^{-1}(\theta_1 - \theta_0) \\
Z &= (G_{x_0}, \ldots, G_{x_{n-1}})' \\
G_{x_t} &= (x_{t-1}, x_{t-1}/(1 + x_{t-1}^2), \cos(1.2(t-1)))'.
\end{aligned}$$

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

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

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

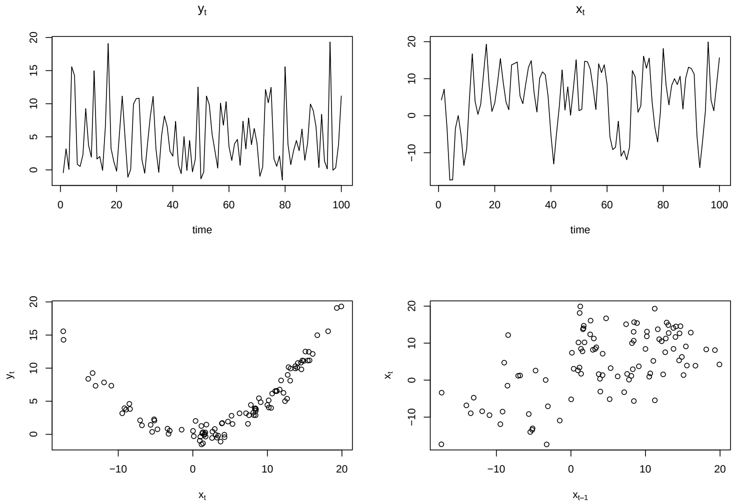
$$\alpha = \min\left\{1, \frac{p(x_t^*|x_{-t}, y_t, \psi)}{p(x_t^{(j)}|x_{-t}, y_t, \psi)}\right\}$$

4. New state:

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

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

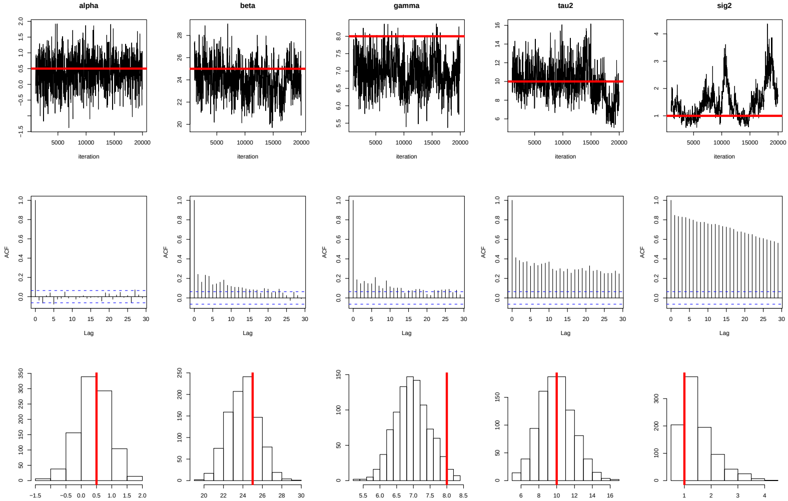
$$v_x^2 = (0.1)^2$$

- Burn-in period, step and MCMC sample size

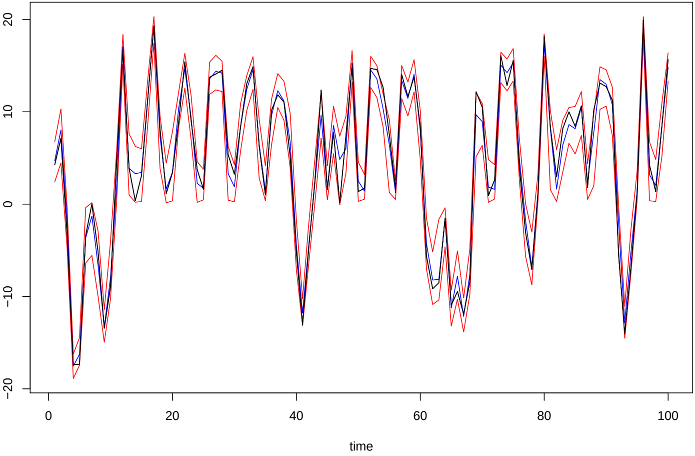
$$M_0 = 1,000 \quad L = 20 \quad M = 950 \;\Rightarrow\; 20,000 \text{ draws}$$

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

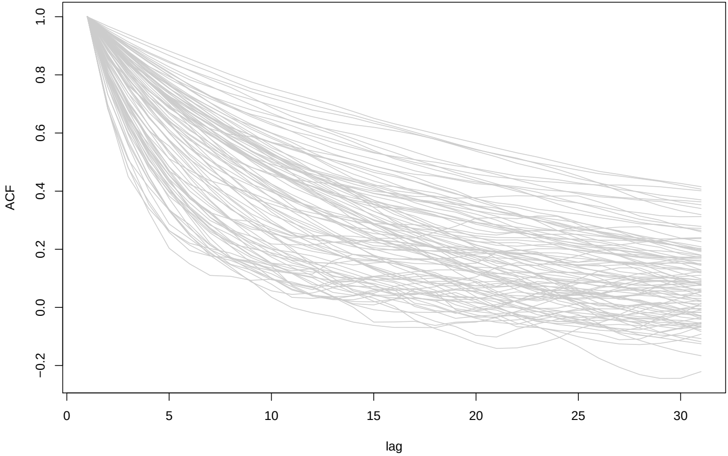
Parameters



States

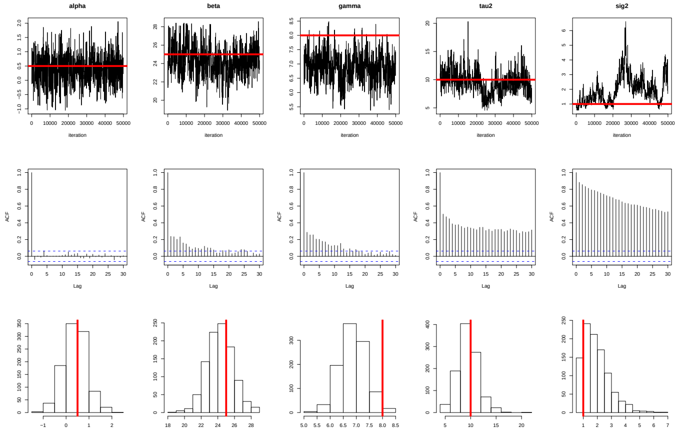


States

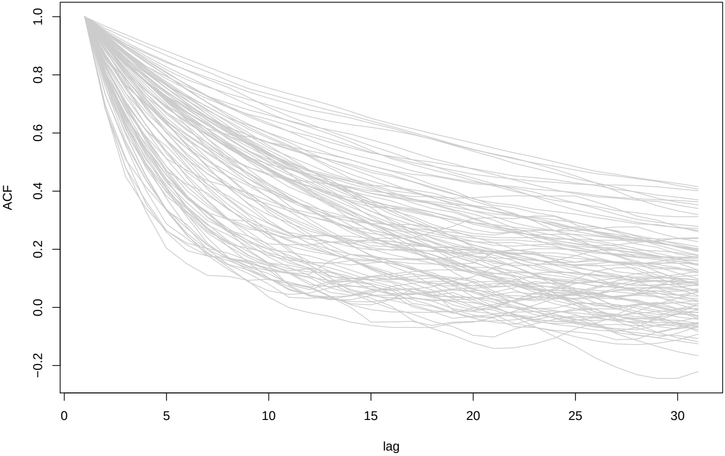


Parameters

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



States



Stochastic volatility model

The canonical stochastic volatility model (SVM), 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 $l \neq 0$.

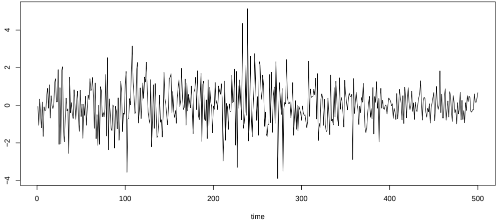
τ^2: volatility of the log-volatility.

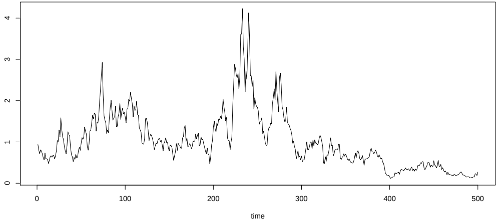
$|\phi| < 1$ then h_t is a stationary process.

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

Simulated data

$n = 500$, $h_0 = 0.0$ and $(\mu, \phi, \tau^2) = (-0.00645, 0.99, 0.0225)$





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.

Simulated data

Simulation setup

- $h_0 = 0.0$
- $\mu = -0.00645$
- $\phi = 0.99$
- $\tau^2 = 0.0225$

Prior distribution

- $h_0 \sim N(0, 100)$
- $\mu \sim N(0, 100)$
- $\phi \sim N(0, 100)$
- $\tau^2 \sim IG(5, 0.14)$ (Mode=0.0234; 95% c.i.=(0.014; 0.086))

Posterior inference

The SVM 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

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

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

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

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 RWM

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

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

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

RWM with tuning v_h^2 $(t = 1, \ldots, n)$:

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:

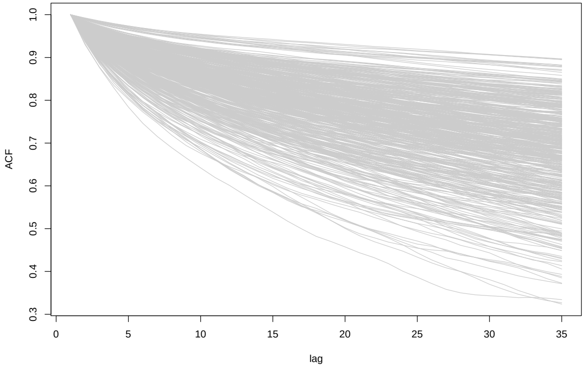
$$h_t^{(j+1)} = \begin{cases} h_t^* & \text{w. p. } \alpha \\ h_t^{(j)} & \text{w. p. } 1-\alpha \end{cases}$$

Simulated data

MCMC setup

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

Autocorrelation of h_ts



Sampling h_t via IMH

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

IMH algorithm

For $t = 1, \ldots, n$

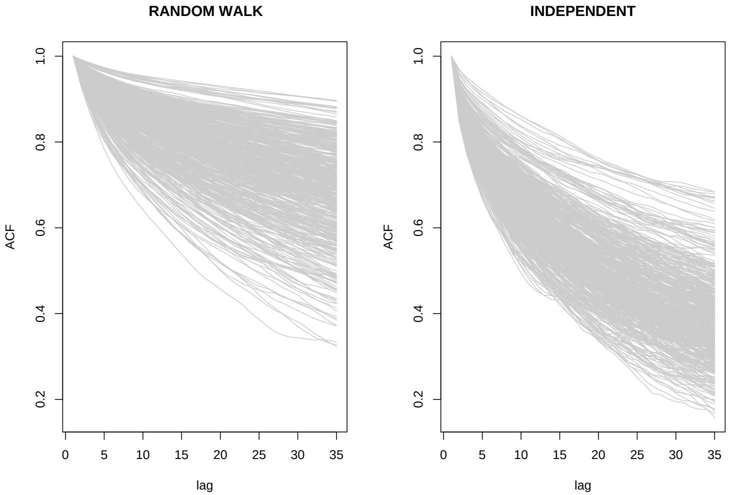
1. Current state: $h_t^{(j)}$
2. Sample h_t^* from $N(\tilde{\mu}_t, \nu_t^2)$
3. Compute the acceptance probability

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

4. New state:

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

ACF for both schemes



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

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

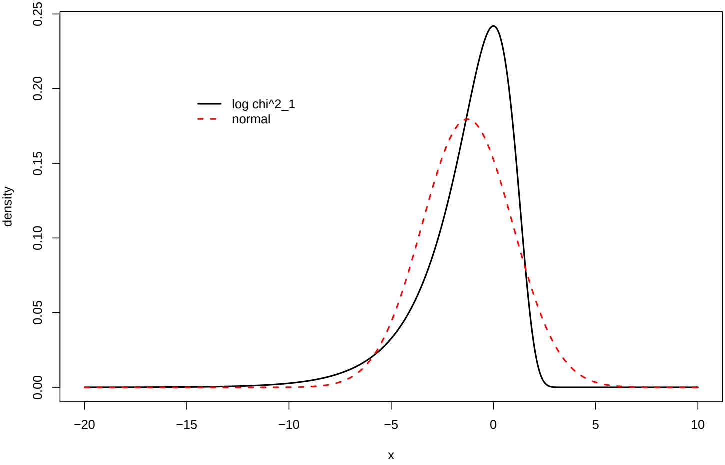
is a simple DLM where v_t and η_t are $N(0, 1)$.

Sampling from

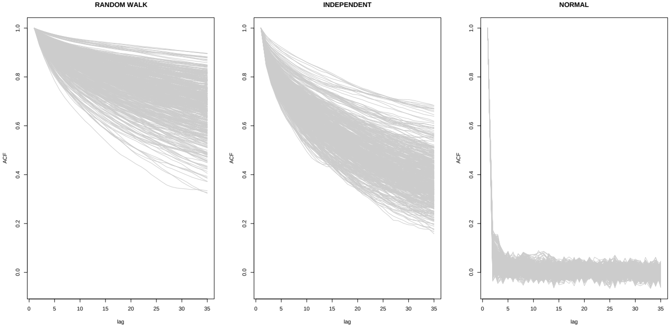
$$p(h^n | \theta, \tau^2, \sigma^2, z^n)$$

can be performed by the FFBS algorithm.

$\log \chi^2_1$ and $N(-1.27, \pi^2/2)$



ACF for the three schemes



Sampling h^n - mixtures of normals and FFBS

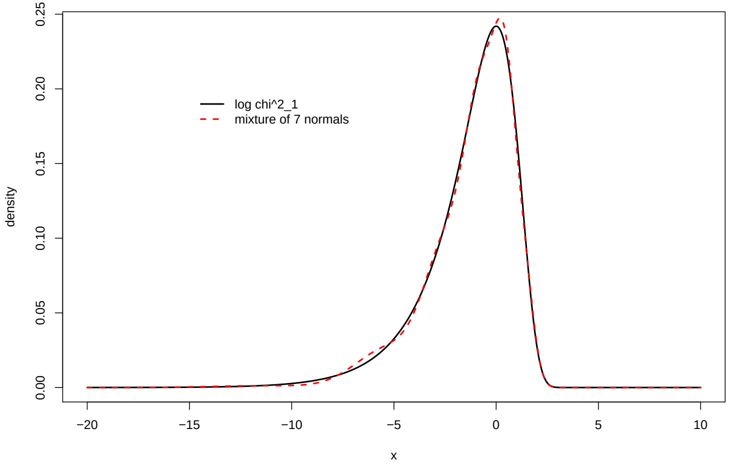
The log χ^2_1 distribution can be approximated by

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

where

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

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



Mixture of normals

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

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

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

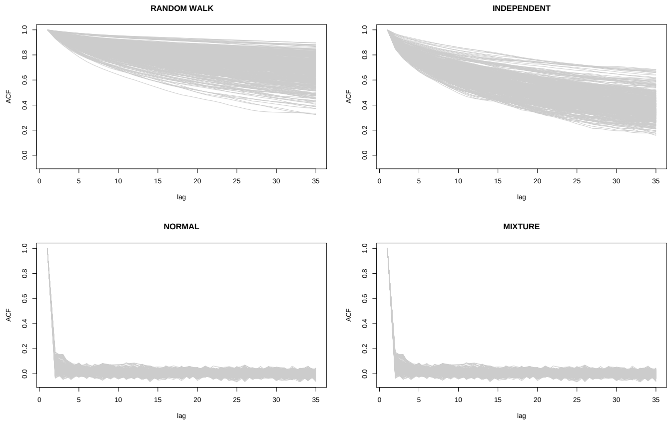
which can be rewritten as a normal DLM:

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

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

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

ACF for the four schemes



Posterior means: volatilities