Unit 10: Dynamic linear models *

Set up

1st and 2nd order DLMs

Prior/updated/smoothed densities

Sequential and smoothed inference

MCMC in DLMs

- Individual sampling
- Blocking sampling: FFBS algorithm
- Joint sampling

Stochastic volatility model

*Based on Dani Gamerman and Hedibert Lopes' (2007) *Markov Chain Monte Carlo: Stochastic: Simulation for Bayesian Inference*, Chapman&Hall/CRC.

$\Rightarrow$ Large class of models with time-varying parameters.

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

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

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

First order DLM

The simplest time series model is the first order model:

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

and β_t is scalar.

$\Rightarrow$ First order Taylor series approximation of a smooth function representing the time trend of the series.

$\Rightarrow$ Stock control, production planning and financial data analysis.

$\Rightarrow$ Observational and system variances may evolve in time, offering great scope for modelling the variability of the system.

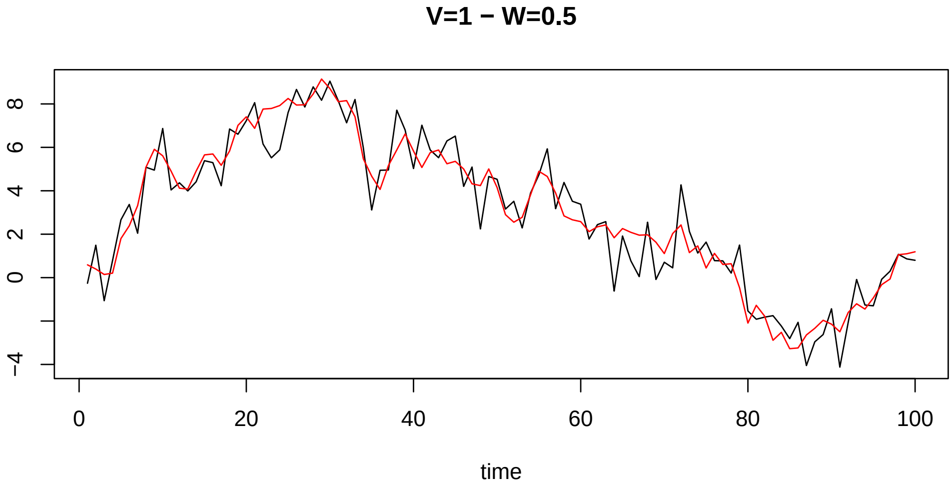
Second order DLM

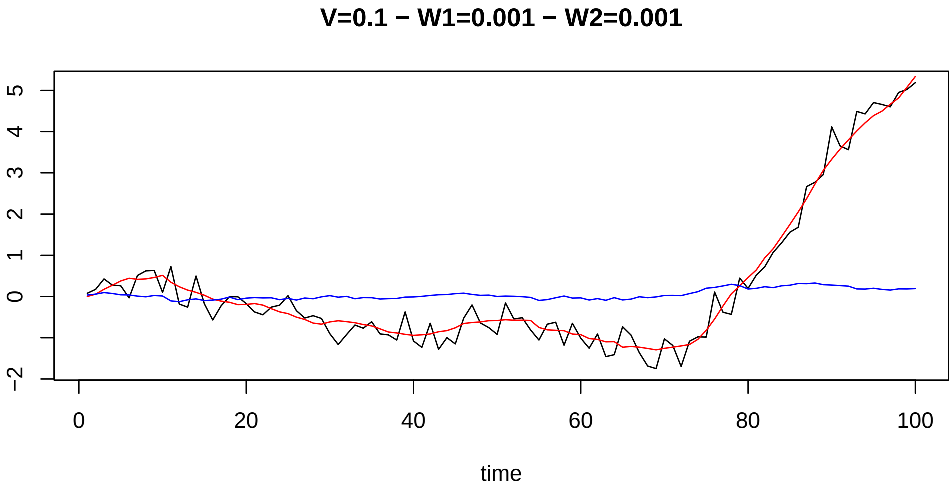
The linear growth model is slightly more elaborate by incorporation of an extra time-varying parameter β_2 representing the growth of the level of the series:

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

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

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





1st and 2nd order dynamic linear models.

prior/updated/smoothed densities

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

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

$\Rightarrow$ $\beta_t | \beta_{t-1} \sim N(G_t \beta_{t-1}, W)$

$\Rightarrow$ Posterior at time $t-1$:

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

$\Rightarrow$ Prior at time t:

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

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

$\Rightarrow$ predictive at time t:

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

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

$\Rightarrow$ Posterior at time t:

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

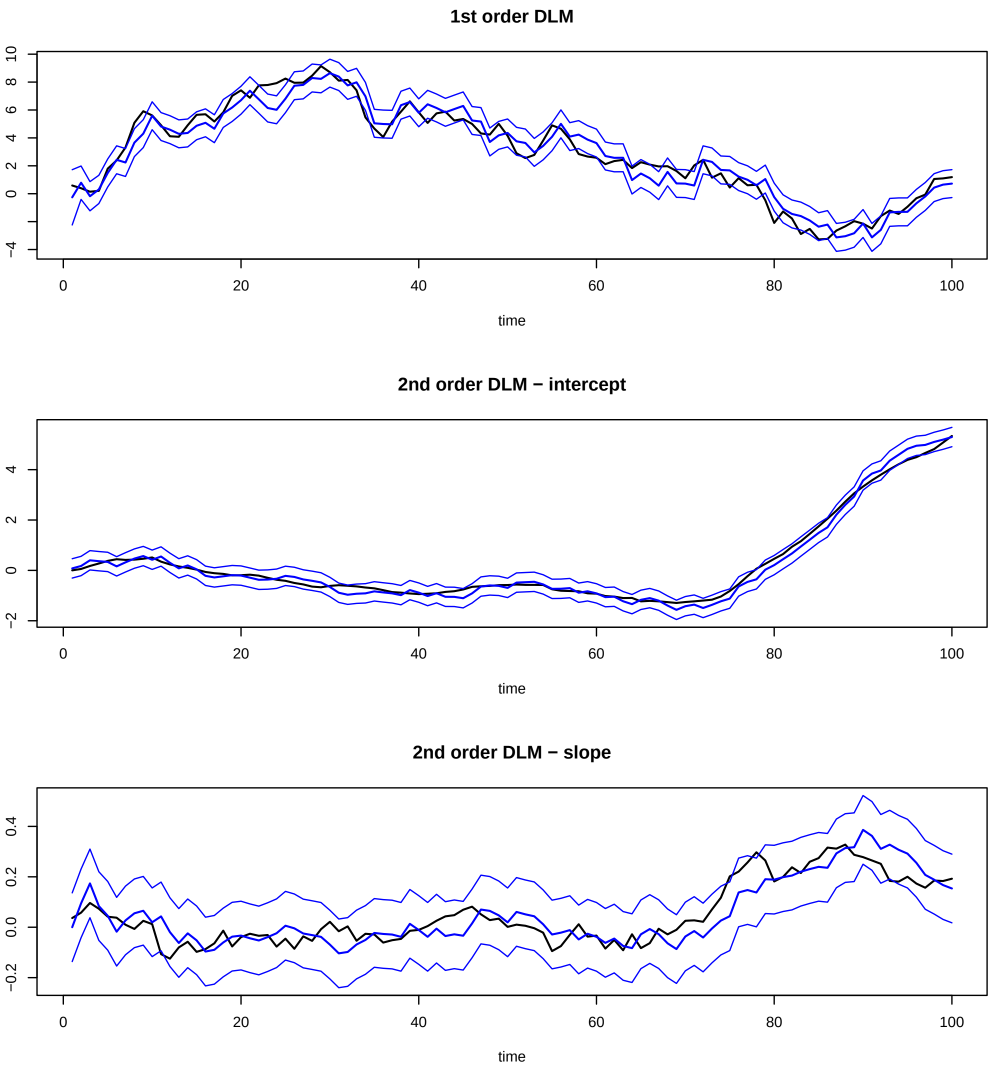
The resulting posterior distribution is

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

with

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

$\Rightarrow$ By induction, these distributions are valid for all times.



Fitted 1st and 2nd order DLMs.

Smoothing

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

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

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

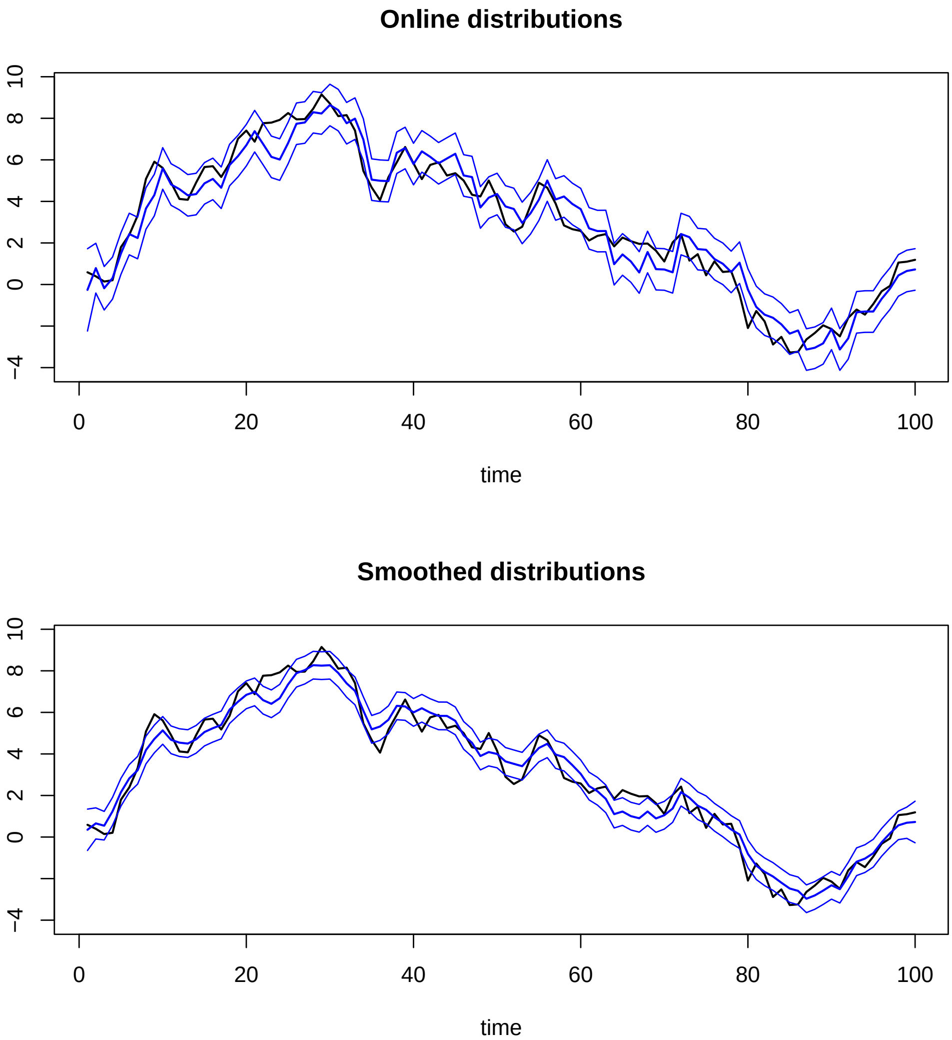
$$
\begin{aligned}
\pi(\beta_t,\ldots,\beta_n|y^n) &= p(\beta_n|y^n) \prod_{k=t}^{n-1} p(\beta_k|\beta_{k+1},y^t) \\
\pi(\beta_t,\beta_{t+1}|y^n) &= p(\beta_{t+1}|y^n) p(\beta_t|\beta_{t+1},y^t)
\end{aligned}
$$

for $t=1,\ldots,n-1$. It can be shown that

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

where

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



First order DLM: online distributions, $p(\beta_t|y^t)$, and smoothed distributions, $p(\beta_t|y^n)$.

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

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

$\Rightarrow$ It is easy to see that

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

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

$\Rightarrow$ The endpoint parameters:

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

The FFBS algorithm
Sampling from $\pi(\beta|y^n)$

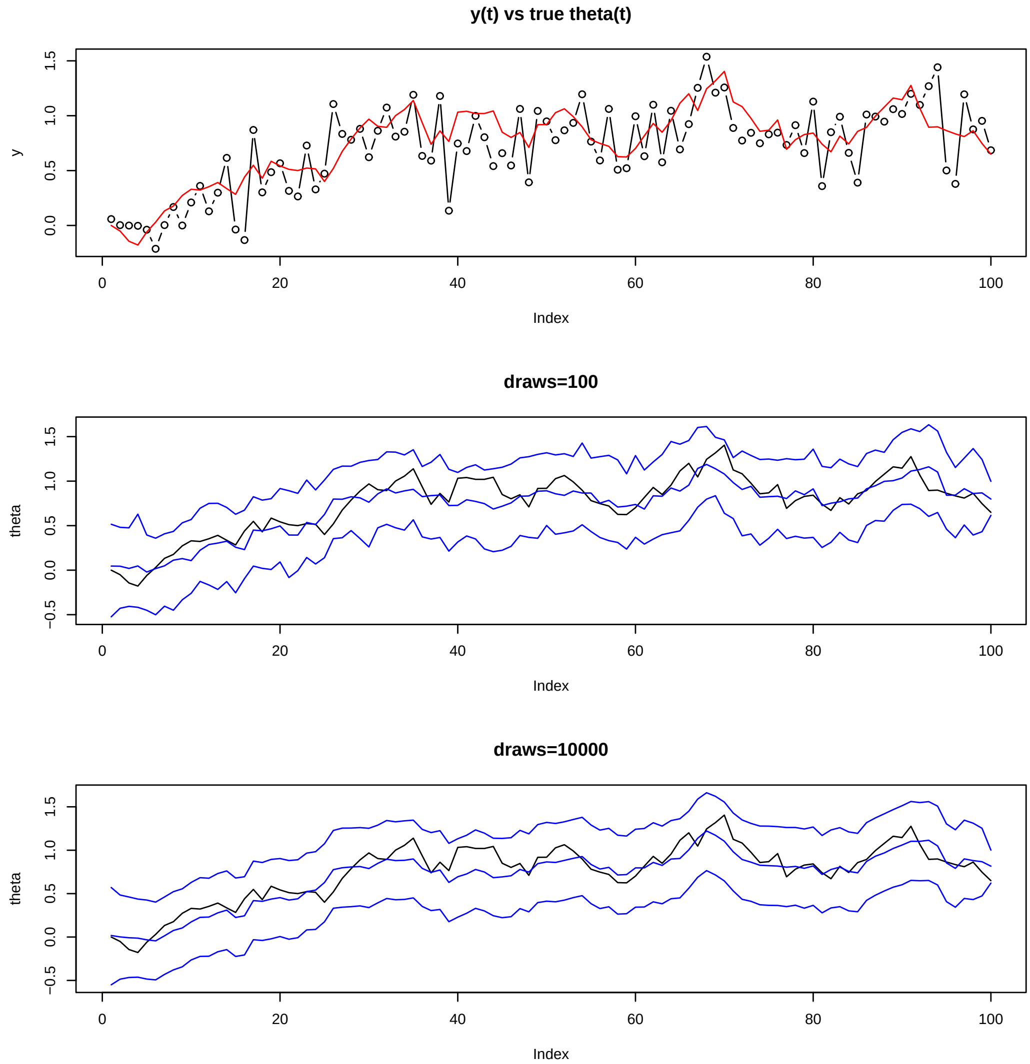
For $t = 1, \ldots, n-1$, it can be shown that $(\beta_t|\beta_{t+1}, V, W, y^t)$ is normally distributed with mean $(G_t' W^{-1} G_t + C_t^{-1})^{-1}(G_t' W^{-1}\beta_{t+1} + C_t^{-1} m_t)$ and variance $(G_t' W^{-1} G_t + C_t^{-1})^{-1}$.

So, a scheme for sampling from the full conditional of the block β is given by sampling β_n from $N(m_n, C_n)$ (*Forward filtering*) and then sampling β_t from $(\beta_t|\beta_{t+1}, V, W, y^t)$, for $t = n-1, n-2, \ldots, 2, 1$ (*Backward sampling*).

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

Carter and Kohn (1994), Frühwirth-Schnatter (1994) and Shephard (1994) show that convergence becomes orders of magnitude faster than sampling each β_t at a time.

FFBS in action



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

$\Rightarrow$ Assume that

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

$\Rightarrow$ 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

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

and f_N, f_G and f_W are normal, gamma and Wishart densities, respectively.

Gibbs sampler for (β, V, W)

- 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

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

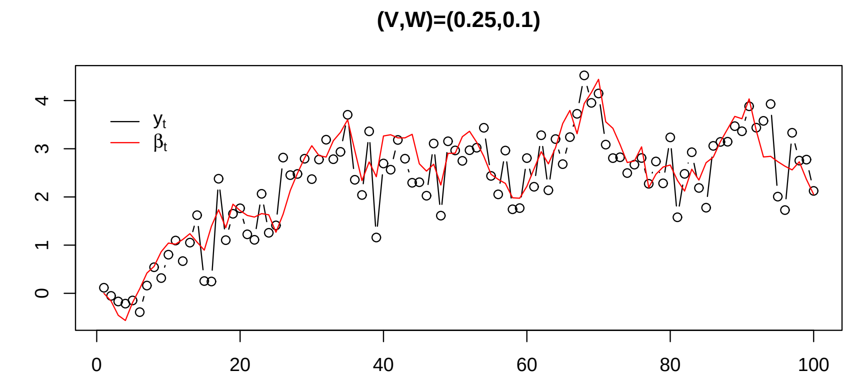
  by the FFBS algorithm.

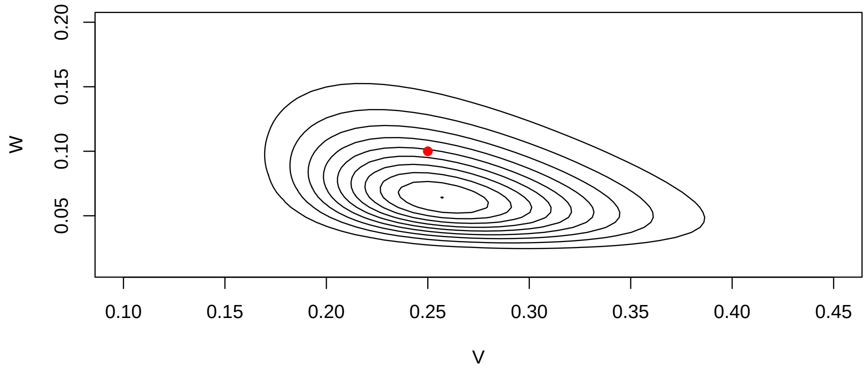
Likelihood for (V, W)

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)

It is easy to see that (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;

- 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 i.

Comparing MCMC schemes

Reis, Salazar and Gamerman (2006)

Model:

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

with $V = 1$ and (n, W) equal to $(100, .01)$, $(100, .5)$, $(1000, .01)$ and $(1000, .5)$ and 100 replications per combination.

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

Posterior inference for each one of the 400 synthetic data was based on 20000 iterations of the MCMC algorithms.

Three different MCMC schemes were implemented for each one of the 400 synthetic data.

Scheme I: Sampling $\beta_1, \ldots, \beta_n, V$ and W individually from their full conditionals;

Scheme II: Sampling β jointly (FFBS algorithm) and V and W from their full 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.

Example ii.

Stochastic volatility model

$$
\begin{aligned}
y_t &= e^{h_t/2}\varepsilon_t \\
h_t &= \mu + \phi h_{t-1} + \sigma\eta_t \\
h_1 &\sim N(a_1, R_1)
\end{aligned}
$$

with $\theta = (\mu, \phi)'$, ε_t and η_t uncorrelated $N(0,1)$ shocks and σ the volatility of the log-volatility. If $|\phi| < 1$ then h_t is a stationary process.

- Prior distributions of θ and σ

$$
\begin{aligned}
\theta|\sigma^2 &\sim N(\theta_0, \sigma^2 A^{-1}) \\
\sigma^2 &\sim IG(\nu_0/2, \nu_0 s_0^2/2)
\end{aligned}
$$

- Sampling θ and σ:

$$
\begin{aligned}
\sigma^2|h, y &\sim IG(\nu_1/2, \nu_1 s_1^2/2) \\
\theta|\sigma^2, h, y &\sim N(\tilde{\theta}, \sigma^2(X'X + A)^{-1})
\end{aligned}
$$

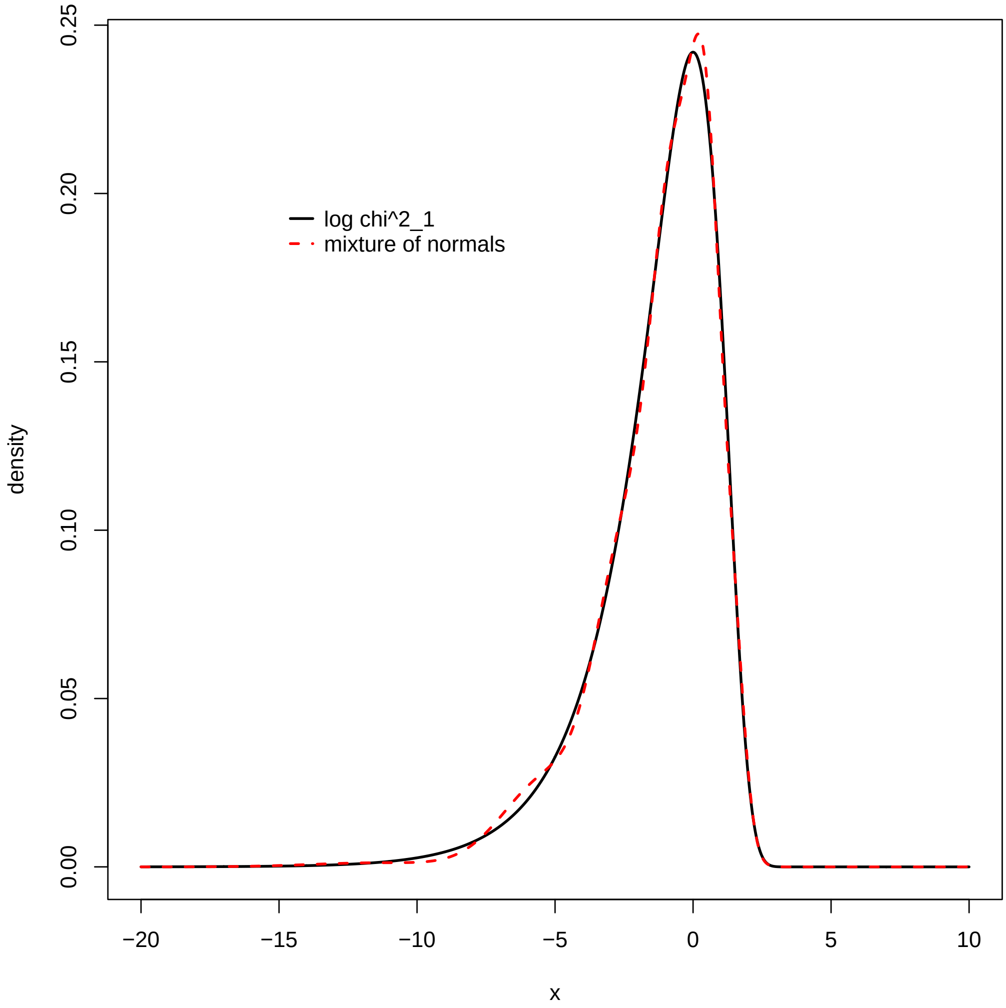
where $\tilde{h} = (h_1, \ldots, h_{n-1})'$, $\bar{h} = (h_2, \ldots, h_n)'$, $X = (1_{n-1}, \tilde{h})$, $\tilde{\theta} = (X'X + A)^{-1}(X'\bar{h} + A\theta_0)$, $\nu_1 = \nu_0 + (n-1)$ $\nu_1 s_1^2 = \nu_0 s_0^2 + (\tilde{\theta} - \theta_0)' A (\tilde{\theta} - \theta_0)$.

- Sampling $h_1, \ldots, h_n$

Kim, Shephard and Chib (1998) proposed sampling h in block (a standard FFBS algorithm) by approximating $\log \chi_1^2$ by the following mixture of 7 normal densities, ie. $\sum_{i=1}^7 \pi_i N(\mu_i, \tau_i^2)$.

i	π_i	μ_i	τ_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

$E\left(\log \chi_1^2\right) = -1.2704$ and $V\left(\log \chi_1^2\right) = 4.9349$.



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

$\Rightarrow$ 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 $i = 1, \ldots, 7$.

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

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

which can be rewritten as a normal dynamic linear model (West and Harrison, 1997):

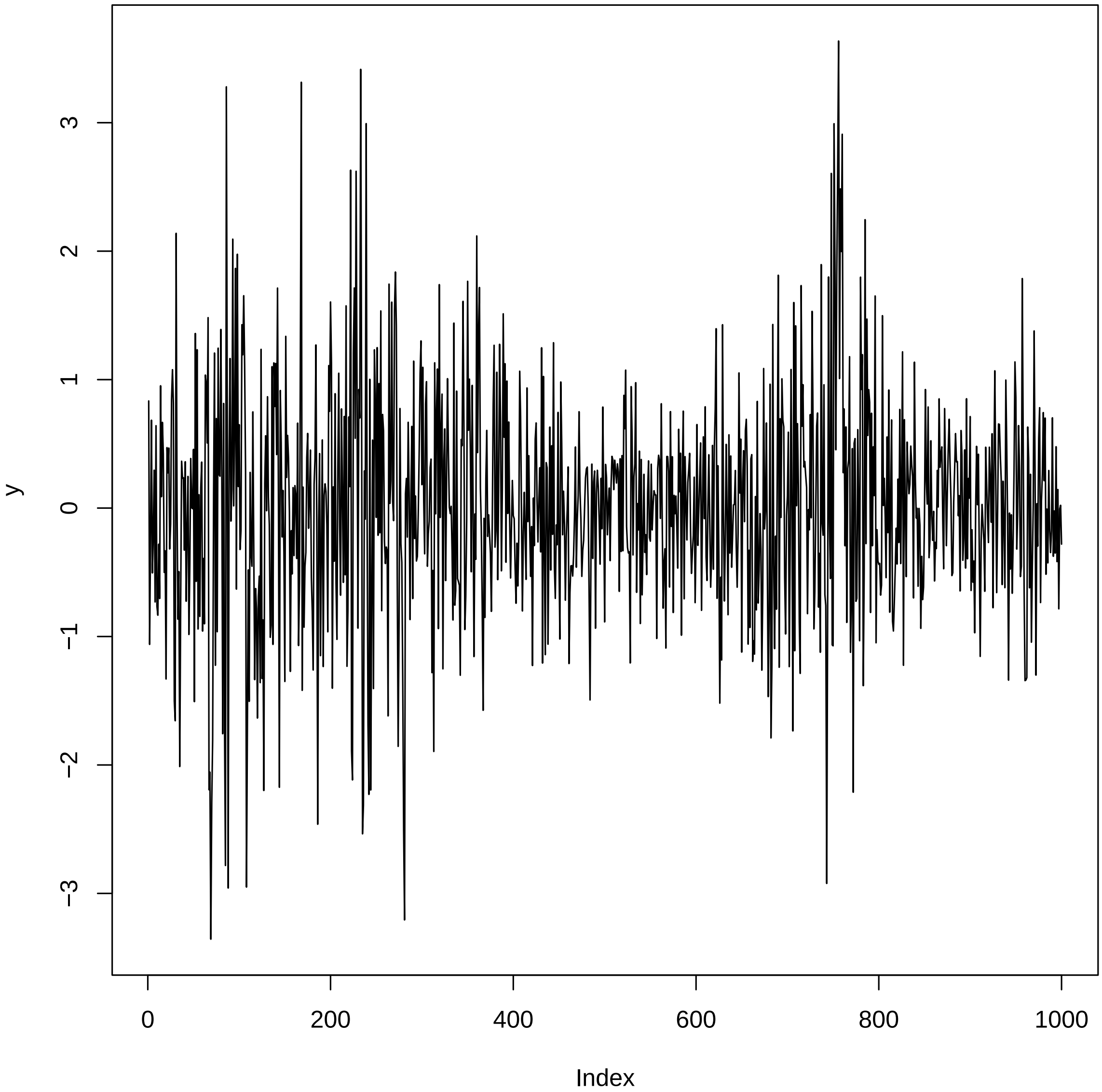
$$\begin{aligned} \tilde{y}_t &= h_t + v_t & v_t \sim N(0, \tau_{z_t}^2) \\ h_t &= \mu + \phi h_{t-1} + w_t & w_t \sim N(0, \sigma_\eta^2) \end{aligned}$$

where $\tilde{y}_t = \log y_t^2 - \mu_{z_t}$ μ_{z_t} and $\tau_{z_t}^2$ are provided in the previous table.

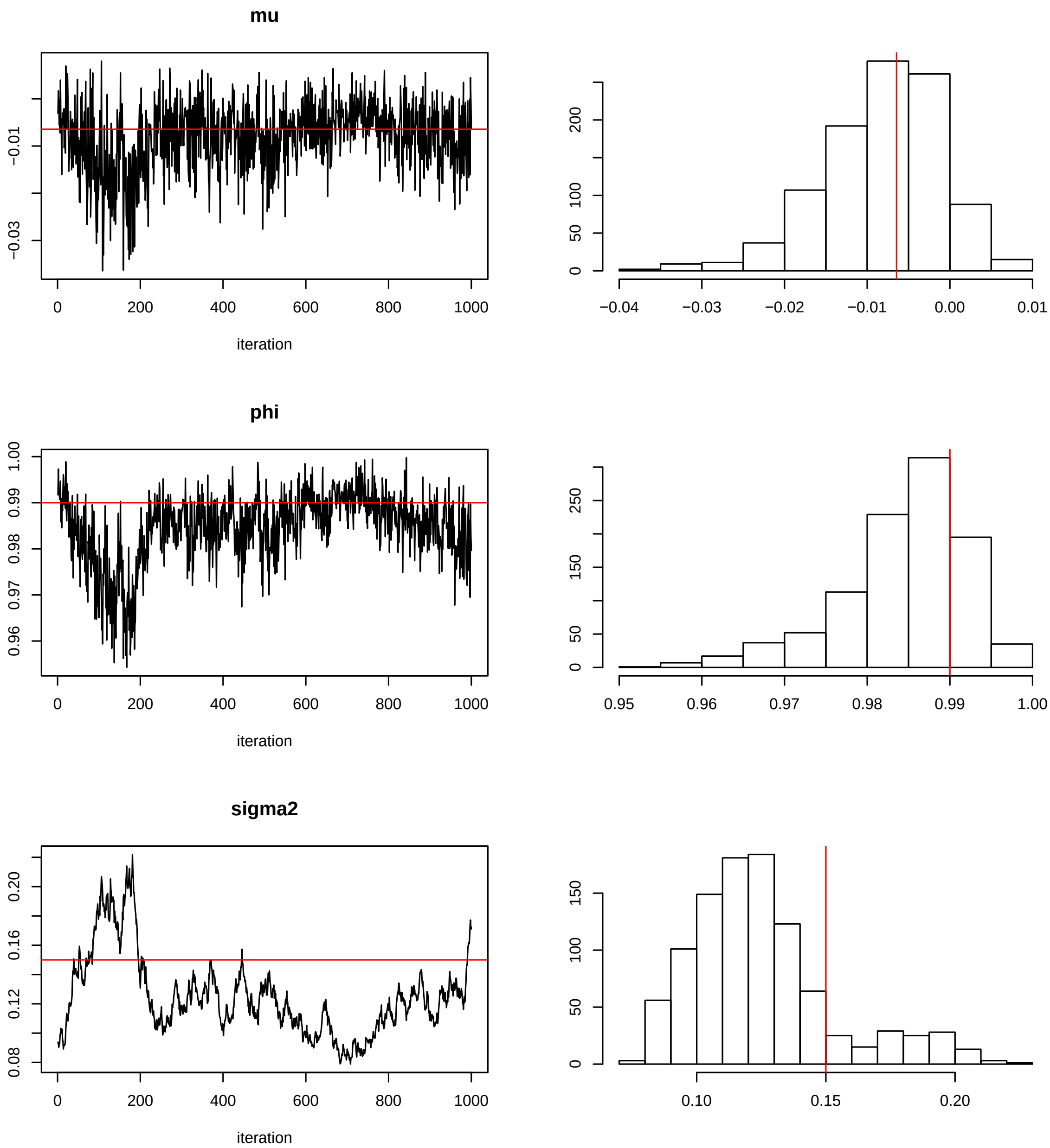
$\Rightarrow$ $(h_1, \ldots, h_n)$ can be jointly sampled by using the the FFBS algorithm.

Example iii.
Simulation

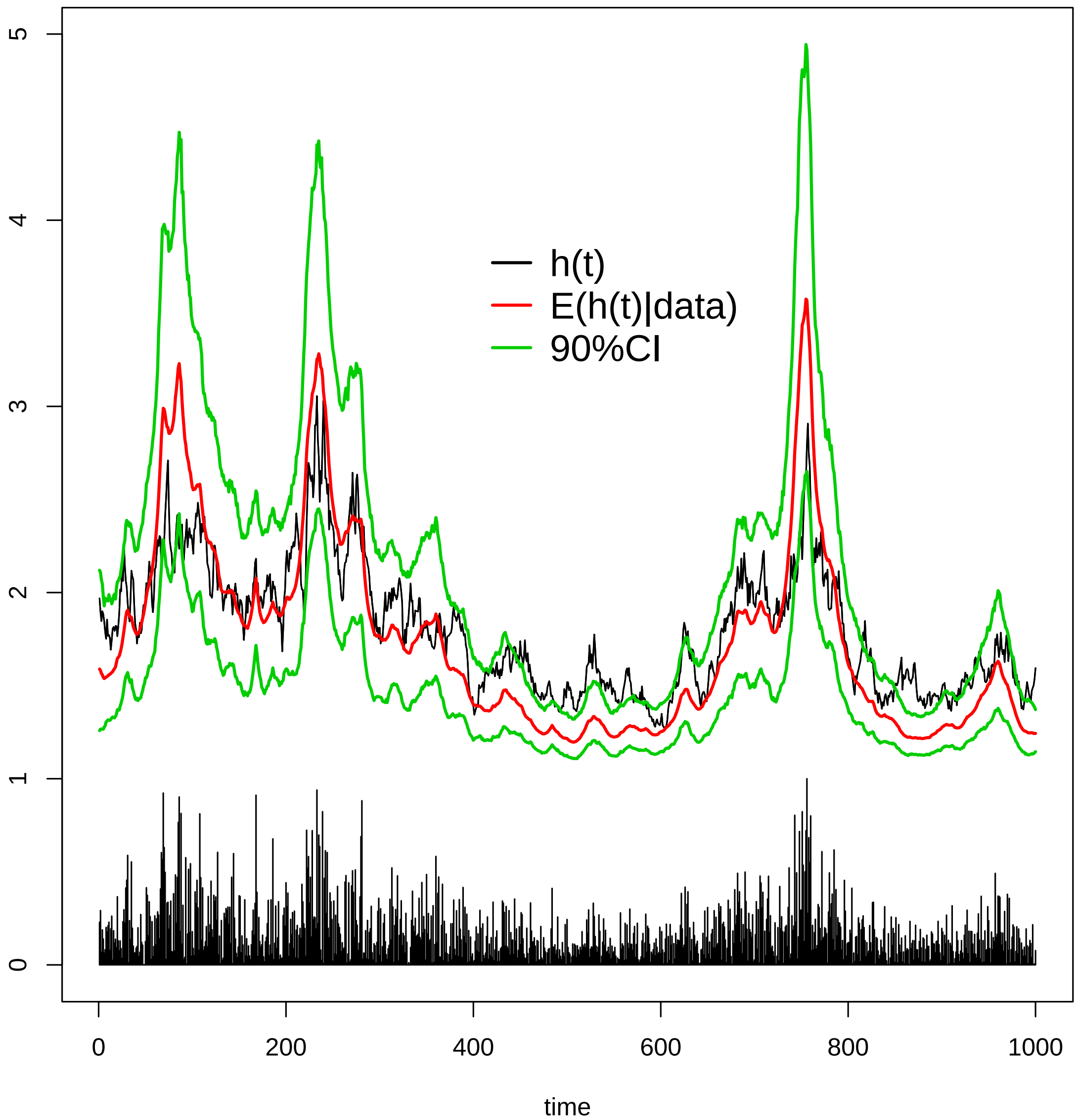
- $n = 1000$, $h_0 = 0.0$,
- $\mu = -0.00645$, $\phi = 0.99$ and $\sigma_\eta^2 = 0.15^2$.
- $\mu \sim N(0, 10^8)$
- $\phi \sim N(0, 10^8)$
- $\sigma_\eta^2 \sim IG(0.00000002/2, 0.00000002/2)$
- $h_1 \sim N(0, 1000)$
- $M_0 = 1000$ and $M = 1000$.



Simulated data.



Chains of μ, ϕ and σ_η^2. Not very good! Should run longer chains.



Posterior summary of the stochastic volatilities, h_t. Vertical bars are the absolute values of y_t.

References

Carter, C. K. and Kohn, R. (1994) On Gibbs sampling for state space models. *Biometrika*, **81**, 541-53.

Frühwirth-Schnatter, S. (1994) Data augmentation and dynamic linear models. *Journal of Time Series Analysis*, **15**, 183-202.

Gamerman, D. and Moreira, A. R. B. (2004) Multivariate spatial regression models. *Journal of Multivariate Analysis*, **91**, 262-81.

Kim, Shephard and Chib (1998) Stochastic volatility: likelihood inference and comparison with ARCH models, *The Review of Economic Studies*, 65, 361–393.

Reis, E. A., Salazar, E. and Gamerman, D. (2006) Comparison of sampling schemes for dynamic linear models. *International Statistical Review* (to appear).

Shephard, N. (1994) Partial non-Gaussian state space. *Biometrika*, **81**, 115-31.