AN EFFICIENT SAMPLING SCHEME FOR DYNAMIC GENERALIZED MODELS

Helio S. Migon, Romy R. Ravines and Alexandra M. Schmidt

Instituto de Matemática. Universidade Federal do Rio de Janeiro. Brazil

Generalized Dynamic Models

- **Dynamic Models**

$$
\begin{aligned}
y_t|\mu_t &\sim p(\mu_t, \boldsymbol{\phi}), & t = 1, \ldots, T. && \text{(1a)}\\
g(\mu_t) &= \boldsymbol{F}_t(\boldsymbol{\psi}_1)'\boldsymbol{\theta}_t & && \text{(1b)}\\
\boldsymbol{\theta}_t &= \boldsymbol{G}_t(\boldsymbol{\psi}_2)\boldsymbol{\theta}_{t-1} + \boldsymbol{w}_t, & \boldsymbol{w}_t \sim N(\boldsymbol{0}, \boldsymbol{W}_t) && \text{(1c)}
\end{aligned}
$$

where $p(\mu_t, \boldsymbol{\phi})$ belongs to the exponential family, $\mu_t = E[y_t]$, $\boldsymbol{\phi}$ denotes other parameters in $p(\cdot)$, and $\boldsymbol{\psi}_1$ and $\boldsymbol{\psi}_2$ are the parameters in $\boldsymbol{F}_t$ and $\boldsymbol{G}_t$.$\boldsymbol{\theta}_t$ **are the state parameters and are related through time via (1c), the system equation**.

- Inference on $\boldsymbol{\theta}$
  - MCMC: Sampling from the posterior of $\boldsymbol{\theta}_t$ can be complicated.
  - Metropolis-Hastings: Gamerman (1998), Geweke and Tanizaki (2001), etc.

- **Generalized Dynamic Linear Models**
  - West, Harrison, and Migon (1985) introduced the Generalized Dynamic Linear Models:

$$
y_t|\eta_t, \phi \sim \exp[\phi\{y_t\eta_t - a(\eta_t)\}]b(y_t, \phi), \quad t = 1, \ldots, T. \qquad \text{(2a)}
$$

$$
\begin{aligned}
\eta_t|D_{t-1} &\sim CP(r_t, s_t) && \text{(2b)}\\
g(\eta_t) &= \boldsymbol{F}_t'\boldsymbol{\theta}_t && \text{(2c)}\\
\boldsymbol{\theta}_t &= \boldsymbol{G}_t\boldsymbol{\theta}_{t-1} + \boldsymbol{w}_t, \quad \boldsymbol{w}_t \sim [0, \boldsymbol{W}_t] && \text{(2d)}\\
\boldsymbol{\theta}_0|D_0 &\sim [\boldsymbol{m}_0, \boldsymbol{C}_0]
\end{aligned}
$$

  where D_t denotes the information at t, $\boldsymbol{m}_t$ e $\boldsymbol{C}_t$ are the first and second moments of $\boldsymbol{\theta}_t$, given D_t.
  - In (2), $\boldsymbol{F}_t$ **and** $\boldsymbol{G}_t$ **are known**, and, given η_t, y_t and $\boldsymbol{\theta}_t$ are independent.
  - West et al. (1985) proposed a system of recursions that uses the conjugate feature of the model to approximate sequentially, the posterior distributions of $\boldsymbol{\theta}_t$: **Conjugate Updating**.

- **Sequential Analysis**
  - Let $D_t = \{y_1, \ldots, y_t\}$ be the information at time t.
  - In **Dynamic Normal Linear Models**:

$$
\ldots \ (\boldsymbol{\theta}_{t-1}|D_{t-1}) \xrightarrow{\text{Evol.}} (\boldsymbol{\theta}_t|D_{t-1}) \xrightarrow{\text{Updat.}} (\boldsymbol{\theta}_t|D_t) \ \ldots
$$

  - In **Dynamic Generalized Linear Models**:

$$
\begin{array}{ccc}
\ldots \ (\boldsymbol{\theta}_{t-1}|D_{t-1}) \xrightarrow{\text{Evol.}} & (\boldsymbol{\theta}_t|D_{t-1}) & (\boldsymbol{\theta}_t|D_t) \ \ldots\\
& \downarrow & \uparrow\\
& (\eta_t|D_{t-1}) \xrightarrow{\text{Updat.}} & (\eta_t|D_t)
\end{array}
$$

Conjugate Updating Backward Sampling (CUBS)

- **CUBS approximation**
  - Let D_t be the information at t. Let $\Phi = (\boldsymbol{\psi}, \boldsymbol{\phi})$.
  - The **full conditional distribution** of $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_T)$ is:

$$
\begin{aligned}
p(\boldsymbol{\theta}|\boldsymbol{Y}, \Phi) &\propto p(\boldsymbol{\theta}_T|D_T, \Phi) \prod_{t=1}^{T-1} \underbrace{p(\boldsymbol{\theta}_t|\boldsymbol{\theta}_{t+1}, D_t, \Phi)}_{\text{Smoothing density}}\\
&\propto p(\boldsymbol{\theta}_T|D_T, \Phi) \prod_{t=1}^{T-1} p(\boldsymbol{\theta}_{t+1} \mid \boldsymbol{\theta}_t, D_t, \Phi)\, \underbrace{p(\boldsymbol{\theta}_t \mid D_t, \Phi)}_{\text{Filtering density}}
\end{aligned}
$$

  - The moments of the **filtering distributions** are approximated by:

$$
\begin{aligned}
p(\boldsymbol{\theta}_t \mid D_t, \Phi) &\propto p(\boldsymbol{\theta}_t \mid D_{t-1}, \Phi) p(Y_t \mid \boldsymbol{\theta}_t, \Phi)\\
&= \int p(\boldsymbol{\theta}_t \mid \eta_t, D_{t-1}, \Phi)\, \underbrace{p(\eta_t \mid D_{t-1}, \Phi) p(Y_t \mid \eta_t, \Phi)}_{\text{Conjugate analysis}} d\eta_t\\
&\propto \int \underbrace{p(\boldsymbol{\theta}_t \mid \eta_t, D_{t-1}, \Phi)}_{\text{Linear Bayes}} p(\eta_t \mid D_t, \Phi) d\eta_t = [\boldsymbol{m}_t, \boldsymbol{C}_t]\\
\boldsymbol{m}_t &= E[\boldsymbol{\theta}_t \mid D_t, \Phi]\\
&= E\big[\hat{E}\{\boldsymbol{\theta}_t \mid \eta_t, D_{t-1}\} \mid D_t, \Phi\big]\\
\boldsymbol{C}_t &= V[\boldsymbol{\theta}_t \mid D_t, \Phi]\\
&= V\big[\hat{E}\{\boldsymbol{\theta}_t \mid \eta_t, D_{t-1}\} \mid D_t, \Phi\big]\\
&\quad + E\big[\hat{V}\{\boldsymbol{\theta}_t \mid \eta_t, D_{t-1}\} \mid D_t, \Phi\big]
\end{aligned}
$$

- **MCMC+CUBS**

1. **Initialization**: set initial values $\boldsymbol{\theta}^{(0)}$, $\boldsymbol{\psi}^{(0)}$ and $i = 1$;
2. **Sample** $\boldsymbol{\theta}^{(i)}$ using CUBS:
   (a) Compute the moments of $p(\boldsymbol{\theta}_t|D_t, \boldsymbol{\psi}^{(i-1)})$, $\boldsymbol{m}^{(i)}$ e $\boldsymbol{C}^{(i)}$, with the *Conjugate Updating*;
   (b) Sample $\boldsymbol{\theta}^*$ with the *Backward Sampling* (Frühwirth-Schnater, 1994).
      i. Sample $\boldsymbol{\theta}_T^*$ from Normal$(\boldsymbol{m}_T^{(i)}, \boldsymbol{C}_T^{(i)})$
      ii. Sample $\boldsymbol{\theta}_t^*, t = T-1, \ldots, 1$, from $p(\boldsymbol{\theta}_t|\boldsymbol{\theta}_{t+1}^*, \boldsymbol{\psi}^{(i-1)})$
   (c) Set $\boldsymbol{\theta}^{(i)} = \boldsymbol{\theta}^*$ with probability p_t and $\boldsymbol{\theta}^{(i)} = \boldsymbol{\theta}^{(i-1)}$ with probability $1 - p_t$, where $p_t = \min(1, A)$ and A is the acceptance rate of the Metropolis-Hastings:

$$
A = \min\left\{1, \frac{\omega(\boldsymbol{\theta}^*)}{\omega(\boldsymbol{\theta})}\right\}, \quad \omega(\boldsymbol{\theta}^*) = \frac{\pi(\boldsymbol{\theta}^*)}{q(\boldsymbol{\theta}^*)},
$$

3. **Sample** $\boldsymbol{\psi}^{(i)}$ using, in general, a Metropolis-Hastings step ;
4. **Sample** $\boldsymbol{\phi}^{(i)}$ using, in general, a Metropolis-Hastings step;
5. **Update**: Set $i = i + 1$ and return to 2 until convergence.

Example: Dynamic Gamma Model

1. Set $t = 1$
2. Compute $\boldsymbol{m}_t$ and $\boldsymbol{C}_t$:
   (a) Compute the prior moments of $\boldsymbol{\theta}_t$ and $g(\eta_t)$, using the model:

$$
\begin{aligned}
\boldsymbol{\theta}_t|D_{t-1} &\sim [\boldsymbol{a}_t, \boldsymbol{R}_t] : \boldsymbol{a}_t = \boldsymbol{G}_t\boldsymbol{m}_{t-1}, \ \boldsymbol{R}_t = \boldsymbol{G}_t\boldsymbol{C}_{t-1}\boldsymbol{G}_t' + \boldsymbol{W}_t\\
g(\eta_t)|D_{t-1} &\sim [f_t, q_t] : f_t = \boldsymbol{F}_t'\boldsymbol{a}_t, \qquad q_t = \boldsymbol{F}_t'\boldsymbol{R}_t\boldsymbol{F}_t
\end{aligned}
$$

   (b) Compute the moments of the conjugate prior for $g(\eta_t)$:

$$
\begin{aligned}
E[g(\eta_t)|D_{t-1}] &= \log r_t - \gamma(s_t + 1) = f_t\\
Var[g(\eta_t)|D_{t-1}] &= \gamma'(s_t + 1) \qquad\quad = q_t
\end{aligned}
$$

   (c) Compute the posterior moments of $g(\eta_t)$:

$$
\begin{aligned}
E[g(\eta_t)|D_t] &= \log(r_t + \phi y_t) - \gamma(s_t + \phi + 1) = f_t^*\\
Var[g(\eta_t)|D_t] &= \gamma'(s_t + \phi + 1) \qquad\qquad\quad = q_t^*
\end{aligned}
$$

   (d) Compute the posterior moments of $\boldsymbol{\theta}_t$, $\boldsymbol{\theta}_t|D_t \sim [\boldsymbol{m}_t, \boldsymbol{C}_t]$:

$$
\boldsymbol{m}_t = \boldsymbol{a}_t + \boldsymbol{R}_t\boldsymbol{F}_t(f_t^* - f_t)\frac{1}{q_t} \quad \boldsymbol{C}_t = \boldsymbol{R}_t - \boldsymbol{R}_t\boldsymbol{F}_t\boldsymbol{F}_t'\boldsymbol{R}_t(1 - \frac{q_t^*}{q_t})\frac{1}{q_t}
$$

3. Set $t = t + 1$ and return to 2 if $t < T$;
4. Sample $\boldsymbol{\theta}_T$ from $N(m_T, C_T)$;
5. Set $t = T - 1$, sample $\boldsymbol{\theta}_t$ from $p(\boldsymbol{\theta}_t \mid \boldsymbol{\theta}_{t+1}, D_t, \boldsymbol{\theta}) = N(\boldsymbol{m}_t^s, \boldsymbol{C}_t^s)$;
6. Set $t = t - 1$ and return to 5 if $t > 1$;

Monte Carlo Study

- **A Simulation Study**

We generated data from a **first order dynamic Poisson model**:

$$
\begin{aligned}
Y_t &\sim \text{Poisson}(\lambda_t), & t = 1, \ldots, T && \text{(3a)}\\
\log(\lambda_t) &= \theta_t & && \text{(3b)}\\
\theta_t &= \theta_{t-1} + w_t & w_t \sim N(0, W) && \text{(3c)}\\
\theta_0 &\sim N(m_0, C_0). & && \text{(3d)}
\end{aligned}
$$

  - Prior: $IG(1e-03, 1e-03)$, for W and $N(0, 1e+03)$, for θ_0.
  - We used (3), with $\theta_0 = 0.50$ and $W = 0.01$, to generate 300 artificial time series of different sizes, $T = (50, 100, 300)$.

- **Comparing 7 different sampling schemes**

I. Conjugate Updating and Backward Sampling (CUBS): Multimove sampling.

II. Conjugate Updating - Single move: The proposal is a Normal density with mean and variance based on the smoothed moments of the Conjugate Updating at time t.

III. From Gamerman (1998) - Single move, sampling from the system disturbances: The proposal is obtained from an adjusted normal dynamic linear model but re-parametrizated in terms of the system disturbances.

IV. From Gamerman (1998) - Single move, sampling from the state parameters: Proposal obtained from an adjusted normal dynamic linear model.

V. From Geweke and Tanizaki (2001) - Proposal Density I: The proposal is the density function obtained from the system equation.

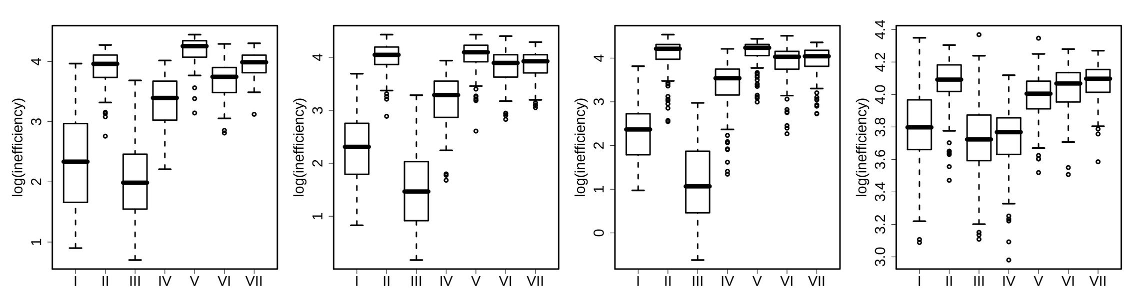
VI. From Geweke and Tanizaki (2001) - Proposal Density II: The proposal is a normal density with mean and variance based on the extended Kalman smoothed estimates at time t.

VII. From Geweke and Tanizaki (2001) - Proposal Density III: The proposal is a normal density with mean based in a random walk and variance based in the extended Kalman smoothed estimates at time t.

- **Results**

Table 1: Root mean square error (RMS), acceptance rate and time (in seconds) for $T = 50$

	RMS_Y		RMS_θ		Acceptance rate		Time(sec)
Scheme	mean	sd	mean	sd	mean	sd	
I	1.2525	0.1095	0.2366	0.0361	42.6297	–	280.3062
II	1.2766	0.1169	0.2342	0.0263	33.5253	7.1575	214.9814
III	1.2599	0.1107	0.2492	0.0513	97.2529	1.2287	629.0840
IV	1.2443	0.1148	0.2583	0.0290	98.1449	0.6814	120.7988
V	1.3155	0.1322	0.2376	0.0345	51.4230	4.5316	89.0456
VI	1.2403	0.1137	0.2611	0.0280	37.7968	6.3019	248.2544
VII	1.2365	0.1109	0.2596	0.0303	44.6136	5.8161	219.6554

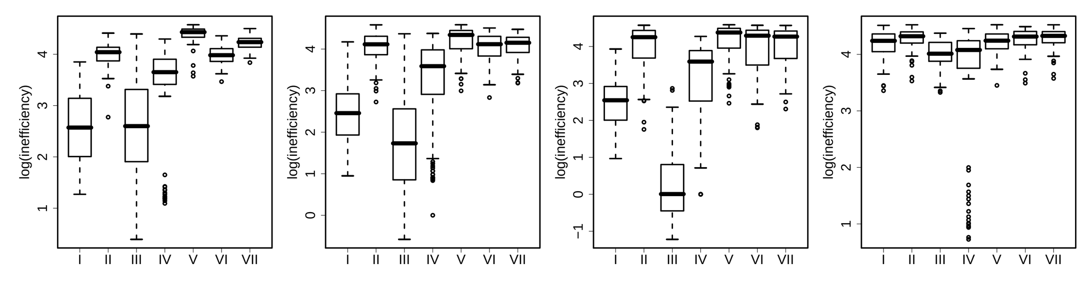


(a) $t = 5$ (b) $t = 25$ (c) $t = 45$ (d) W

Figure 1: Box plots of the (log)inefficiencies of $\theta_t, t = 5, 25, 45$ and W, for $T = 50$.

Table 2: Root mean square error (RMS), acceptance rate and time (in seconds) for $T = 300$

	RMS_Y		RMS_θ		Acceptance rate		Time(sec)
Scheme	mean	sd	mean	sd	mean	sd	
I	1.6085	0.3039	0.2281	0.0352	31.3526	–	247.76
II	1.6158	0.3065	0.3798	0.2213	32.1778	9.4706	314.09
III	1.6107	0.2980	0.2273	0.0376	98.4285	0.6872	2158.93
IV	4.5229	3.7538	0.5459	0.2526	95.5152	14.1218	163.87
V	1.7110	0.3295	0.4718	0.1330	51.3656	4.1081	74.44
VI	1.6076	0.3012	0.2365	0.0344	32.8817	8.9858	371.31
VII	1.6066	0.3025	0.2739	0.0876	40.3615	7.8025	338.89



(a) $t = 5$ (b) $t = 145$ (c) $t = 295$ (d) W

Figure 2: Box plots of the (log)inefficiencies of $\theta_t, t = 5, 145, 195$ and W, for $T = 300$.

- **Final Remarks**
  - We reviewed the seminal work of Dynamic Generalized Linear Models of West et al. (1985) and showed that their proposed algorithm can be used with satisfactory results, to construct a proposal density in a Metropolis-Hastings step to sample in block, all the state parameters of a dynamic model.
  - We performed an extensive comparison between our proposal and other previously established and noted that CUBS is much **simpler** to implement and the results are quite satisfactory.
  - One of our current topics of research is the application of the scheme proposed to make inference in k-parameters distributions.

References

Frühwirth-Schnater, S. (1994). Data augmentation and dynamic linear models. *Journal of Time Series Analysis*, *15*(2), 183–202.

Gamerman, D. (1998). Markov chain Monte Carlo for dynamic generalised linear models. *Biometrika*, *85*(1), 215–227.

Geweke, J., & Tanizaki, H. (2001). Bayesian estimation of state space models using Metropolis-Hastings algorithm within Gibbs sampling. *Computacional Statistics & Data Analysis*, *37*, 151-170.

West, M., Harrison, J., & Migon, H. (1985). Dynamic generalized linear models and Bayesian forecasting. *Journal of the American Statistical Association*, *80*(389), 73–83.