A particle marginal Metropolis-Hastings sampler for bilinear processes

P. de Zea Bermudez, K. F. Turkman and M. A. Amaral Turkman
DEIO, FCUL and CEAUL

Objective

The class of bilinear processes is quite versatile in modeling nonlinear time series. However, due to difficulties in inference, it is not much used in practice. This work aims to apply a particle Markov chain Monte Carlo algorithm to estimate the parameters of a simple bilinear model ([4]). The algorithm we use is the particle marginal Metropolis-Hastings sampler (PMMCMC) constructed by Andrieu et al. [2] as a inferential tool for state space models (SSM). We will show in the next section that bilinear models can be written in a state space form and inference on these on these models can be made using this algorithm.

Bilinear models

The bilinear models [4] are possibly the most natural way to extend the ARMA methodology in order to address nonlinear behavior, which is observed in many real time series. The class of bilinear models plays an important role in modeling non-linearity for various reasons:

1. The class is an obvious generalization of ARMA models resulting in non-linear conditional mean and conditional variance.
2. Under fairly general conditions, bilinear processes approximate finite order Volterra series expansions to any desired order of accuracy over finite time intervals (see Brockett [3]). Volterra series expansion are a dense class within the class of non-linear time series. Therefore, under fairly general conditions, bilinear processes are also a dense class within non linear processes, approximating any nonlinear process to a desired level of accuracy.
3. The class is fairly well-studied. Much is known regarding the existence of unique and stationary solutions. Although identification, estimation and diagnostic techniques are available, much of the work on the class remains to be completed.
4. Bilinear processes are often used in the control theory are somewhat different from the context within in which they are used in time series context. In the control theory, the output Xt, as well as the the input process Z_t are observable, making the probabilistic structure simple. In the context we use these models, the input random process Z_t is not observed. This somewhat restricts the use of these models within the context of time series. In estimation and prediction, it is important to know that the input process is measurable with respect to the Y_s, $s \le t$, i.e., invertible. Unfortunately, the lack of verifiable conditions for invertibility (except for very simple bilinear processes) limits the use of these processes as models.
5. Bilinear processes are capable of producing sudden bursts of large values and hence are very suitable for modeling time series showing burst-like phenomena.

These last particular features are evident in the simulated data sets that is plotted in Figure 1 (right). Very large values are observed when the innovation process is Pareto(2.5), whereas moderate observations are obtained when the errors are normally distributed.

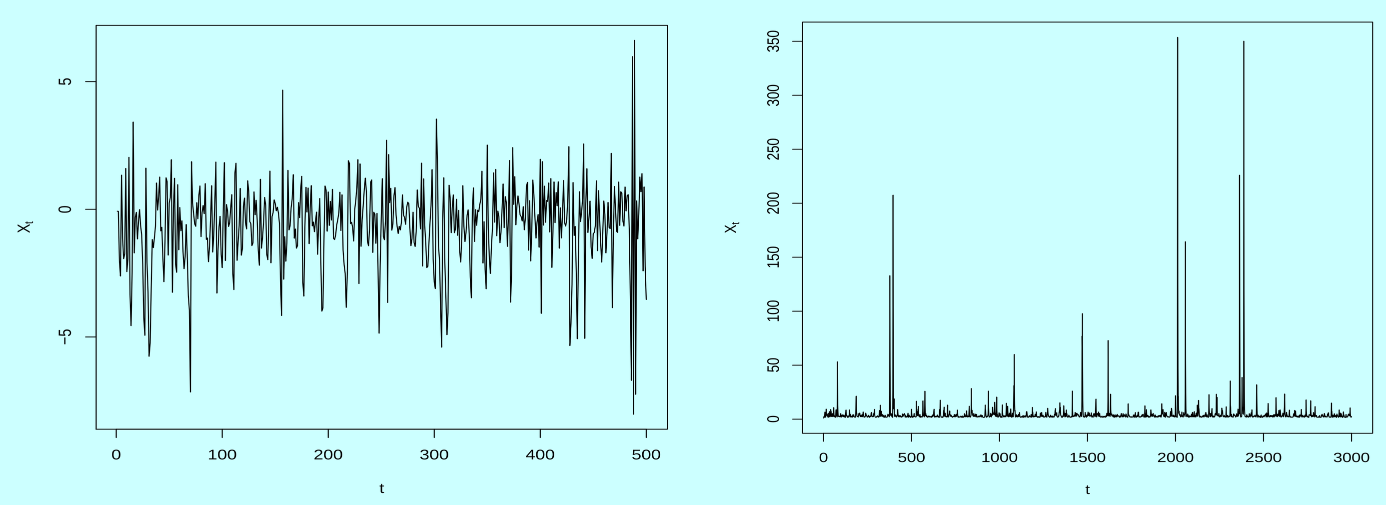

Figure 1: Simulated bilinear data sets - Standard gaussian (left) and Pareto(2.5) innovations (right)

For $\alpha = 2.5$, this Pareto distribution has both finite mean and variance. The situation can be more extreme if $1 < \alpha \le 2$ or $\alpha \le 1$. In the first case, only the mean is finite, whereas in the second neither the mean nor the variance are finite.
Y_t is said to be a bilinear process, BL(p, q, m, k), if it satisfies the difference equation:

$$Y_t = \sum_{j=1}^{p} a_j Y_{t-j} + \sum_{j=1}^{q} c_j \epsilon_{t-j} + \sum_{l_1=1}^{m} \sum_{l_2=1}^{k} b_{l_1 l_2} Y_{t-l_1} \epsilon_{t-l_2} + \epsilon_t. \qquad (1)$$

where $\{\epsilon_t\}$ is a sequence of i.i.d. random variables with zero mean and variance σ^2.
In this work we consider that the observable process, Y_t, is a bilinear model of order (1,0,1,1) given by $Y_t = aY_{t-1} + bY_{t-1}\epsilon_{t-1} + \epsilon_t$, where $\epsilon_t \sim N(0, \sigma^2)$. The process is invertible and stationarity if the parameters a, b and σ^2 satisfy the condition $a^2 + b^2\sigma^2 < 1$. Solving iteratively for Y_t, we get, after n iterations,

$$Y_t = (\prod_{i=1}^{n} (a + b\epsilon_{t-i}) Y_{t-n} + \sum_{j=1}^{n-1} (\prod_{i=1}^{j} (a + b\epsilon_{t-i}))\epsilon_{t-j},$$

If $X_t = (a + b\epsilon_t)Y_t$ is written as $X_t = (a + b\epsilon_t)X_{t-1} + (a + b\epsilon_t)\epsilon_t$, and $Y_t = X_{t-1} + \epsilon_t$, X_t is a Markov process and Y_t has the standard latent Markov process state space representation.

The PMMCMC algorithm

Let $\{Y_t, t \in \mathbb{N}\}$ be the observed bilinear model and $\{X_t, t \in \mathbb{N}\}$ be the hidden (unobserved) Markov chain (MC). Let $y_{1:T} = (y_1, y_2, \ldots, y_T)$ be a the set of observations up to time T. The model is characterized by:

- the initial distribution, $\mu_\theta(.)$
- the transition probability of the MC, $f_\theta(x_t \mid x_{t-1})$,
- the distribution of the observations at time t, conditional on the state of the chain at time t, x_t, $g_\theta(y_t \mid x_t)$

for same value of θ. The focus is to approximate the density $p_\theta(x_{1:T} \mid y_{1:T})$ and the marginal density $p_\theta(y_{1:T})$, using some sequential MCMC scheme. In this general framework, the joint posterior distribution of all the unknowns of the model is given by

$$p(\theta, x_{1:T} \mid y_{1:T}) \propto \mu_\theta(x_1) \prod_{t=2}^{T} f_\theta(x_t \mid x_{t-1}) \prod_{t=1}^{T} g_\theta(y_t \mid x_t) p(\theta).$$

What is the procedure? (Andrieu et al. [2])
Basically, for each $t = 1, 2, \ldots, T$ and for some θ, a set of N particles, $x_t^{(1)}, x_t^{(2)}, \ldots, x_t^{(N)}$, which are values of the hidden MC, are simulated from some adequate distribution. Then, using Importance Sampling, a weight is calculated for every particle. A bootstrap procedure is applied to the N particles in order to eliminate the trajectories which are not sufficiently relevant. After that, Andrieu et al. [2] recommend carrying out a Metropolis-Hastings step which will accept (or not) a new candidate value for θ, θ^*. All this process is repeated some fixed number of iterations.

Details of the Sequential Importance Sampling (SIS) algorithm

(I) N particles, $x_1^{(1)}, x_1^{(2)}, \ldots, x_1^{(N)}$, are simulated from a proposal distribution $q_\theta(. \mid y_1)$.
(II) The weights associated to the particles are computed using the following expression:

$$w_1(x_1^k) = \frac{\mu_\theta(x_1^k) q_\theta(y_1 \mid x_1^k)}{q_\theta(x_1^k \mid y_1)}.$$

(III) Let $W_1^k = \frac{w_1(x_1^k)}{\sum_{m=1}^{N} w_1(x_1^m)}$, $k = 1, 2, \ldots, N$.
(IV) For $t = 2, 3, \ldots, T$

- A bootstrap procedure is performed to choose the particles with higher weights. Let A_{t-1}^k be the index of the k^{th} chosen particle at time $t-1$.
- Sample $x_t^k \sim q(. \mid y_t, x_{t-1}^{A_{t-1}^k})$ and let $x_{1:t}^k = (x_{1:t-1}^{A_{t-1}^k}, x_t^k)$.
- The associated weights are given by

$$w_t(x_{1:t}^k) = \frac{f_\theta(x_t^k \mid x_{t-1}^{A_{t-1}^k}) g_\theta(y_t \mid x_t^k)}{q_\theta(x_t^k \mid y_t, x_{t-1}^{A_{t-1}^k})}, k = 1, 2, \ldots, N.$$

- These weights are then normalized, as was done in step (III).

When this sequence is concluded, we possess T sets of N particles. These information is sufficient to estimate the marginal likelihood, $\hat{p}_\theta(y_{1:T}) = \hat{p}_\theta(y_1) \prod_{t=2}^{T} \hat{p}_\theta(y_t \mid y_{1:t-1})$, where

$$\hat{p}_\theta(y_t \mid y_{1:t-1}) = \frac{1}{N} \sum_{k=1}^{N} w_t(x_{1:t}^k).$$

Overall procedure

- Initial iteration - For a certain $\theta = \theta(0)$, apply the SIS in order to obtain a set of particles $x_0^{(1)}, x_0^{(2)}, \ldots, x_0^{(N)}$ and the corresponding estimates of the marginal likelihood distribution, $\hat{p}_{\theta(0)}(y_{1:T})$.
- The following steps are to be carried out a certain number of iterations ($i = 1, 2, \ldots tot$):
  1. generate a new value θ^* from the proposal $q\{. \mid \theta(i-1)\}$.
  2. run the SIS algorithm in order to produce a new set of MC trajectories $X_{1:T}^* \sim \hat{p}_{\theta^*}(. \mid y_{1:t})$ and compute the corresponding marginal likelihood estimate.
  (a) Compute

$$p = \min\left\{1, \frac{\hat{p}_{\theta^*}(y_{1:T}) p(\theta^*) q\left[\theta(i-1) \mid \theta^*\right]}{\hat{p}_{\theta(i-1)}(y_{1:T}) p\left[\theta(i-1)\right] q\left[(\theta^*) \mid \theta(i-1)\right]}\right\}.$$

  (b) The new value, θ^*, is accepted with probability given by the previous expression. In this case, the estimates of the marginal distribution and the optimal trajectories associated to the i^{th} iteration take the new values. Otherwise, neither of the quantities are updated, taking the previous values.

Which proposal distributions are being considered in the SIS?

For the BL(1,0,1,1) model, the proposal distributions considered are:

- $q_\theta(x_1^k \mid y_1) \sim N(-y_1, \sigma^2)$ and $q_\theta(y_1 \mid x_1^k) \sim N(x_1^k, \sigma^2)$. The p.f.d. of the hidden MC in x_1 is given by $\mu_\theta(x_1) = \Delta^{-1/2}\left[\phi(\frac{-a+\sqrt{\Delta}}{2b}) + \phi(\frac{-a-\sqrt{\Delta}}{2b})\right]$, where $\Delta = a^2/(4ax_1)$ and $\phi(.)$ is, as usual, the p.d.f. of the standard Normal distribution. This p.d.f. is only valid provided $b > 0$ and $x_1 < x$ or $b < 0$ and $x_1 < x$, where $x = -a^2/(4b)$.
- $q_\theta(x_t^k \mid y_t, x_{t-1}^{A_{t-1}^k})$ can be calculated as $x_t^k = ax_{t-1}^{A_{t-1}^k} + (a + bx_{t-1}^{A_{t-1}^k})\epsilon_t^k + b(\epsilon_t^2)^k$.
- In this framework, the weights $w_t(x_{1:t}^k)$ are $N(x_{t-1}^{A_{t-1}^k}, \sigma^2)$.

What about the prior distribution for θ?

The parameters a and b of the bilinear model are assumed to be independent with uniform distributions in (-1,1). Due to the invertibility and stationarity condition of the bilinear models, σ^2 has an upper bound equal to $(1-a)/b^2$. Then the prior distribution for σ^2 is uniform in $(0, (1-a)/b^2)$.

Ongoing work

1. R code has been developed to implement the SIS with a step of MH described above. It has been used to estimate the parameters of the BL(1,0,1,1) considering several values of a, b and σ^2.
2. The results obtained so far are not yet satisfactory. The inability of the procedure to accurately estimate the parameters of the model is related to inadequate choices of the proposal distributions.
3. Consequently, all the effort in the near future will be devoted to devising more appropriate proposal distributions.
4. When all these issues have been solved, we intend to extend the procedure to more complex bilinear models.

References

[1] Doucet, A., Godsill, S. and Andrieu, C. (2000). On sequential Monte Carlo sampling methods for Bayesian filtering. *Stat. and Computing*, **10**, 197-208.

[2] Andrieu, C., Doucet, A. and Holenstein, R. (2010). Particle Markov chain Monte Carlo methods. *J. R. Statist. Soc. B*, **72**, 269-342.

[3] Brockett, R. W. (1976). Non-linear systems and differential geometry. *Automatica*, 12, 167-176.

[4] Subba Rao, T. and Gabr, M. M. (1984). An Introduction to Bispectral Analysis and Bilinear Time Series Models. Springer-Verlag, New York.

2nd **Workshop on Bayesian Inference for Latent Gaussian Models with Applications**
Trondheim, 30 May to 01 June 2012
The research is (partially) supported by the projects PEst-OE/MAT/UI0006/2011 and PTDC/MAT/118335/2010.