BAYESIAN ANALYSIS OF ORDER UNCERTAINTY IN ARIMA MODELS

BY RICARDO S. EHLERS AND STEPHEN P. BROOKS

Federal University of Paraná, Brazil and University of Cambridge, UK

Abstract. In this paper we extend the work of Brooks and Ehlers (2002) and Brooks et al. (2003) by constructing adaptive proposal schemes for reversible jump MCMC in the context of autoregressive moving average models. In particular, the full conditional distribution is not available for the added parameters and approximations to it are provided by suggesting an adaptive updating scheme which automatically selects proposal parameter values to improve the efficiency of between-model moves. The performance of the proposed algorithms is assessed by simulation studies and the methodology is illustrated by applying it to a real data set.

Keywords. Bayesian model selection, posterior model probability, Markov chain Monte Carlo, reversible jump MCMC, autoregressive moving average.

Introduction

In many applications, there is substantial prior uncertainty concerning the choice of most appropriate models for any given data.

Classical approach: use information criteria such as

Akaike (1974) $AIC(\boldsymbol{\theta}_i, M_i) = -2 \log p(y|\boldsymbol{\theta}_i, M_i) + 2n_i$

Bayesian approach: index all models under consideration, treating this index as another parameter and considering posterior model probabilities.

For k competing models $M_1, \ldots, M_k$, *a priori* we assign probabilities $p(M_i)$ to each model.

For each model there is a vector of parameters $\boldsymbol{\theta}_i \in \mathbb{R}^{n_i}$ with:

- a likelihood given all observations $p(\boldsymbol{y}|\boldsymbol{\theta}_i, M_i)$
- a prior distribution $p(\boldsymbol{\theta}_i|M_i)$.

We obtain the posterior distribution for model and associated parameters via Bayes theorem,

$$\pi(M_i, \boldsymbol{\theta}_i) \propto p(y|\boldsymbol{\theta}_i, M_i)\, p(\boldsymbol{\theta}_i|M_i)\, p(M_i)$$

Trans-dimensional Jumps

The reversible jump algorithm (Green 1995) is a general strategy for generating samples from $\pi(k, \boldsymbol{\theta}^{(k)})$ based upon the standard Metropolis-Hastings approach of proposing a move and defining a acceptance probability.

Suppose that the current state is $(k, \boldsymbol{\theta}^{(k)})$, where $\boldsymbol{\theta}^{(k)}$ has dimension n_k, and we have defined different move types allowing transitions between spaces of different dimensions. A move type r is performed with probability $p_k(r)$ by generating $\boldsymbol{u}$ from a proposal density $q(\cdot)$ and setting

$$(\boldsymbol{\theta}^{(k')}, \boldsymbol{u}') = g(\boldsymbol{\theta}^{(k)}, \boldsymbol{u})$$

where $n_k + dim(\boldsymbol{u}) = n_{k'} + dim(\boldsymbol{u}')$.

We accept $(k', \boldsymbol{\theta}^{(k')})$ as the new state with probability $\min(1, A)$ where

$$A = \frac{\pi(k', \boldsymbol{\theta}^{(k')})\; p_{k'}(r')\; q(\boldsymbol{u}')}{\pi(k, \boldsymbol{\theta}^{(k)})\; p_k(r)\; q(\boldsymbol{u})} \left| \frac{\partial g(\boldsymbol{\theta}^{(k)}, \boldsymbol{u})}{\partial(\boldsymbol{\theta}^{(k)}, \boldsymbol{u})} \right| \tag{1}$$

is called the acceptance ratio.

A class of moves for transitions between nested models consists of adding/deleting parameters. Assuming that $n_{k'} > n_k$ then $dim(\boldsymbol{u}) = n_{k'} - n_k$ and the transition from the larger model to the smaller one is deterministic. The acceptance ratio reduces to

$$A = \frac{\pi(k', \boldsymbol{\theta}^{(k')}|\boldsymbol{y})\ p_{k'}(r')}{\pi(k, \boldsymbol{\theta}^{(k)}|\boldsymbol{y})\ p_k(r)\ q(\boldsymbol{u})} \left| \frac{\partial g(\boldsymbol{\theta}^{(k)}, \boldsymbol{u})}{\partial(\boldsymbol{\theta}^{(k)}, \boldsymbol{u})} \right| \tag{2}$$

The applications in this paper will focus on a particular implementation in which $\boldsymbol{\theta}^{(k')} = (\boldsymbol{\theta}^{(k)}, \boldsymbol{u})$. In this case, the Jacobian term is equal to 1 and the acceptance ratio simplifies to

$$\begin{aligned} A &= \frac{p(\boldsymbol{y}|k', \boldsymbol{\theta}^{(k')})}{p(\boldsymbol{y}|k, \boldsymbol{\theta}^{(k)})} \frac{p(\boldsymbol{\theta}^{(k')}|k')p(k')}{p(\boldsymbol{\theta}^{(k)}|k)p(k)} \frac{p_{k'}(r')}{p_k(r)q(\boldsymbol{u})} \\ &= \text{likelihood ratio} \times \text{prior ratio} \times \text{proposal ratio} \end{aligned} \tag{3}$$

Choice of proposal distribution q is crucial to cover the parameter space.

When possible use the complete conditionals, or approximations for the complete conditionals **Brooks and Ehlers (2002)**.

Selection of ARIMA Models

Ehlers and Brooks (2004) ARIMA(p, d, q) models

$$X_t = \sum_{j=1}^{k} a_j \, X_{t-j} + \sum_{j=1}^{q} b_j \, \epsilon_{t-j} + \epsilon_t$$

are reparameterized in terms of inverse roots of characteristic polinomials

$$\prod_{i=1}^{k}(1 - \lambda_i L) \, X_t = \prod_{j=1}^{q}(1 - \delta_j L) \, \epsilon_t, \quad \epsilon_t \sim N(0, \sigma^2).$$

where $X_t = (1 - L)^d \, Y_t$.

Stationarity/inversibility are easily imposed: $|\lambda_i| < 1, \; i = 1, \ldots, k$ and $|\delta_i| < 1, \; i = 1, \ldots, q$.

k, q and d are unknown parameters and play the role of model indicator. Upper bounds on model order are fixed *a priori* and each model is equally likely.

Possible jumps: Addition/Deletion of 1 real root or a pair of complex conjugate roots.

Possible proposals: Truncated Normal, Beta-based, Logistic-based.

Parameter Priors

Conditional on model order we assume independent priors for the real reciprocal roots $r \in (-1, 1)$ and any pairs of complex conjugate reciprocal roots $(\lambda_j, \lambda_{j^*})$

$$\begin{aligned} \lambda_j &= r\cos\theta + ir\sin\theta \\ \lambda_{j^*} &= r\cos\theta - ir\sin\theta. \end{aligned}$$

The prior is in terms of θ and r and we assume prior independence,

$$p(\lambda_j, \lambda_{j^*}) = p(\theta)p(r)\left|\frac{\partial(\lambda_j, \lambda_{j^*})}{\partial(\theta, r)}\right|^{-1}.$$

We assume a $\theta \sim U(0, \pi)$ and a logistic-based prior for r

$$r = \frac{2e^x}{1+e^x} - 1, \quad x \sim N(0, \sigma_a^2). \tag{4}$$

Large values of $|x|$ correspond to values of $|r|$ close to 1. This prior becomes more concentrated around zero as σ_a^2 decreases and U-shaped and more concentrated near -1 and 1 as σ_a^2 increases.

Assigning $\sigma_a^2 \sim IG(\alpha, \beta)$ its full conditional distribution is given by

$$IG\left(\alpha + \frac{n_r + n_c}{2}, \beta + \frac{1}{2}\left[\sum_{i:\lambda_i \in \mathbb{R}} x_i^2 + \sum_{j:\lambda_j \in \mathbb{C}} x_j^2\right]\right)$$

where n_r and n_c are the number of real roots and the number of complex conjugate pairs respectively.

Within-Model Moves

Assigning $\sigma^2_\epsilon \sim IG(c, d)$ it is easy to see that its full conditional distribution has the inverse Gamma form. This parameter is then updated by a Gibbs move.

In order to update the ARMA coefficients we randomly choose one of the reciprocal roots and use Metropolis-Hastings updates with the proposal density centred on the current value as follows.

- If we choose $\lambda_j \in \mathbb{R}$ we propose a new value $\lambda'_j \sim U[\max(\lambda_j - \delta, -1), \min(\lambda_j + \delta, 1)]$.
- If we choose $\lambda_j \in \mathbb{C}$ we propose a new value for $(\lambda_j, \lambda_{j^*})$ by sampling $\theta^* \sim U[\max(0, \theta - \delta), \min(\pi, \theta + \delta)$ and $r^* \sim U[\max(r - \delta, -1), \min(r + \delta), 1]$, and setting the proposed new values as $r^* \cos\theta^* \pm ir^* \sin\theta^*$.

This scheme ensures that new values are proposed in a neighbourhood of the current ones and are restricted to stationarity/inversibility.

Model Priors

The root structure is not unique, except for $k = 1$ or $q = 1$. In order to assign a uniform prior on the AR order, the prior probability for a certain value of k should be split uniformly over the possible configurations of real and complex roots corresponding to that order. Likewise for the MA component.

This can be accomplished by assigning,

$$\mathbb{P}(r \text{ real and } c \text{ complex roots}) \propto \frac{1}{[k/2]+1}.$$

This prior specification differs from Huerta and West (1999) where a uniform distribution is assigned to the possible configurations of real and complex roots thus leading to a non-uniform prior distribution on model order.

Between-Model Moves

Four model move types are allowed each one proposed with probability 1/4.

We propose a real birth by adding one real reciprocal root $r \in (-1, 1)$.

A complex birth is propose by adding a pair of complex reciprocal roots

$$\begin{aligned} u &= r\cos\theta + ir\sin\theta \\ \bar{u} &= r\cos\theta - ir\sin\theta \end{aligned}$$

with new values sampled for (θ, r).

Under this parameterisation and updating scheme, the models can be treated as nested so that the Jacobian of the transformation from $(\lambda_1, \ldots, \lambda_k)$ to either $(\lambda_1, \ldots, \lambda_k, r)$ or $(\lambda_1, \ldots, \lambda_k, u, \bar{u})$ equals 1.

Conversely, a real (or complex) death is proposed by randomly selecting one real (or complex) roots and deleting it (or the pair of complex conjugates).

We consider three families of proposal densities,

- sample $r \sim N(\mu, \sigma^2)$ truncated to $(-1, 1)$,
- sample $u \sim Beta(\alpha_1, \alpha_2)$ and set $r = 2u - 1$,
- sample $z \sim N(\mu, \sigma^2)$ and set $r = \dfrac{2e^z}{1 + e^z} - 1$.

Updating the Number of Unit Roots

A change in d implies a change in k only. For example, we can propose a move from ARIMA$(k, 0, q)$ to

$$\text{ARIMA}(k-1, 1, q) \text{ or } \text{ARIMA}(k-2, 2, q)$$

so we allow unit roots (if they exist) to be either complex or real.

After we accept or reject $d+1$ new values of k and q are proposed and accepted or rejected. So it is possible to go from (k, d, q) to $(k, d+1, q)$ in two steps. In practice, values other than $d = 0$, 1 or 2 would not make much sense, so this is the range of possible values that we adopt here.

The criteria for proposing these moves are as follows:

- we randomly choose one root which is greater (in absolute value) than a prespecified lower bound L and propose 1 or 2 differences depending on the root being real or complex (this implies deleting 1 or 2 roots).
- Otherwise, the number of differences is decreased by 1 or 2, which implies adding 1 or 2 roots by sampling from U(-1,-L) or U(L,1) with probability 1/2.

Adaptive Proposals for Reciprocal Roots

When updating the AR component we can use the same expressions that appear in Brooks and Ehlers (2002) with the error terms redefined for ARMA models.

Proposing a move from q to $q+1$ in the MA component by adding one real reciprocal root, r the representation for the higher dimensional model is

$$\prod_{i=1}^{k}(1-\lambda_i L)y_t = (1-rL)\prod_{j=1}^{q}(1-\delta_j L)\epsilon'_t$$

so that $\epsilon_t = (1-rL)\epsilon'_t = \epsilon'_t - r\epsilon'_{t-1}$ where the ϵ_t denote the error terms in the original model and ϵ'_t depends on r in a complicated non-linear way.

Treating ϵ'_{t-1} as if it were fixed in the larger model the likelihood function under the larger model is given by,

$$L(\boldsymbol{y} \mid k, q, \boldsymbol{\lambda}, \boldsymbol{\delta}, \sigma^2_\epsilon) \propto \exp\left[-\frac{1}{2\sigma^2_\epsilon}\sum(\epsilon_t + r\epsilon_{t-1})^2\right].$$

and again expressions given in Brooks and Ehlers (2002) can be used here with slight modifications. Likewise for complex birth and death in the MA component.

Simulation Study

Data sets were simulated from AR(3), MA(3) and ARMA(3,3) processes and for the three proposals we ran our algorithm for 1 million iterations (discarding the first 500,000) using the first 20, 50, 100, 500 and 1000 observations and recorded the estimated posterior probability of the true model. Maximum model orders are $k_{\max} = q_{\max} = 5$ and $d = 0, 1, 2$.

In Table 1 we show the results for the simulated AR(3). For each proposal distribution the first row refers to the average posterior probability of the true model while the second row shows the proportion of correct choices of the true model.

Clearly, the performance of the algorithm improves as the sample size increases and a similar pattern is observed for the three proposals considered. Acceptable performances seem to be achieved for at least 200 observations.

Table 1: Model probabilities and proportion of correct model for a simulated AR(3).

	Sample size					
proposal	20	50	100	200	500	1000
Truncated normal	0.0092	0.0398	0.1074	0.2677	0.4565	0.5480
	0.0000	0.1500	0.3500	0.6000	0.9000	0.9500
Beta-based	0.0096	0.0441	0.1059	0.2702	0.4812	0.5404
	0.0000	0.1500	0.3500	0.6500	0.9500	0.9000
Logistic-based	0.0092	0.0414	0.1058	0.2628	0.4822	0.5436
	0.0000	0.1500	0.3000	0.6000	0.9500	0.9000

Figure 1: Southern oscillation index (SOI), 540 measurements taken between 1950-1995.

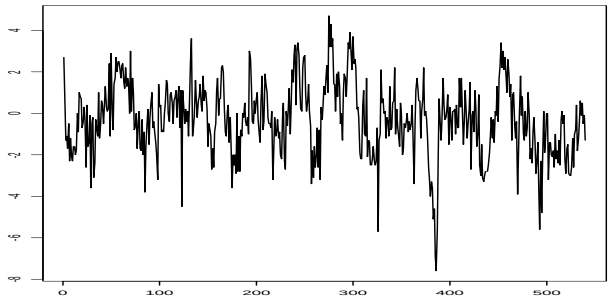

Table 2: Posterior model order probabilities, 500,000 iterations after a 500,000 burn-in,

Proposal	d	(p,q)	0	1	2	3	4	5
Truncated normal	0	1	0.0000	**0.2245**	0.0518	0.0503	0.0398	0.0336
		2	0.0024	0.0164	0.0556	0.0292	0.0249	0.0184
		3	0.0111	0.0173	0.0374	0.0286	0.0247	0.0186
		4	0.0130	0.0083	0.0178	0.0134	0.0140	0.0104
		5	0.0200	0.0071	0.0148	0.0113	0.0103	0.0085
	1	0	0.0000	0.0135	0.0089	0.0156	0.0147	0.0193
		1	0.0000	0.0023	0.0042	0.0067	0.0078	0.0105
		2	0.0001	0.0018	0.0032	0.0053	0.0067	0.0086
		3	0.0004	0.0020	0.0024	0.0041	0.0042	0.0058
		4	0.0004	0.0012	0.0024	0.0034	0.0050	0.0060

References

Akaike, H. (1974). A new look at the statistical identification model. *IEEE Transactions on Automatic Control 19*, 716–723.

Brooks, S. P. and R. S. Ehlers (2002). Efficient construction of reversible jump MCMC proposals for autoregressive time series models. Technical report, University of Cambridge.

Brooks, S. P., P. Giudici, and G. O. Roberts (2003). Efficient construction of reversible jump MCMC proposal distributions (with discussion). *Journal of the Royal Statistical Society, Series B 65*, 1–37.

Ehlers, R. S. and S. P. Brooks (2004). Bayesian analysis of order uncertainty in arima models. Technical Report 2004/05-B, Federal University of Paran.

Green, P. J. (1995). Reversible jump MCMC computation and Bayesian model determination. *Biometrika 82*, 711–732.

Huerta, G. and M. West (1999). Priors and component structures in autoregressive time series models. *Journal of the Royal Statistical Society, Series B 51*, 881–899.