BAYESIAN ESTIMATION OF POLYNOMIAL MOVING AVERAGE MODELS WITH UNKNOWN DEGREE OF NONLINEARITY

Oktay Karakuş[*], *Ercan E. Kuruoğlu*[†], *Mustafa A. Altınkaya*[*]

[*] İzmir Institute of Technology
Electrical-Electronics Engineering
Urla, İzmir, Turkey

[†] ISTI-CNR,
via G. Moruzzi 1, 56124,
Pisa, Italy

ABSTRACT

Various real world phenomena such as optical communication channels, power amplifiers and movement of sea vessels exhibit nonlinear characteristics. The nonlinearity degree of such systems is assumed to be known as a general intention. In this paper, we contribute to the literature with a Bayesian estimation method based on *reversible jump Markov chain Monte Carlo* (RJMCMC) for *polynomial moving average* (PMA) models. Our use of RJMCMC is novel and unique in the way of estimating both model memory and the nonlinearity degree. This offers greater flexibility to characterize the models which reflect different nonlinear characters of the measured data. In this study, we aim to demonstrate the potentials of RJMCMC in the identification for PMA models due to its potential of exploring nonlinear spaces of different degrees by sampling.

Index Terms— Polynomial MA, Nonlinearity degree estimation, Reversible Jump MCMC.

1. INTRODUCTION

Modelling of real world problems employs nonlinear models, generally with fixed degree of polynomial nonlinearity or it necessitates an exhaustive search over all model order values. Hence, having an idea about this nonlinearity degree of the aforementioned models is of utmost importance. In this study, we propose a Bayesian approach which estimates the nonlinearity degree as well as the *moving average* (MA) order and the model coefficients of a Volterra series expansion based nonlinear model, namely *polynomial MA* (PMA).

In the literature, *nonlinear moving average* (NMA) models [1] have been preferred for modelling various real life signals and systems, such as radio frequency power amplifiers (RF-PAs) [2], bridge aerodynamics [3], finance [4] and adaptive control of the nonlinear systems [5]. This preference is motivated by the cases when the weighted sum of past values of errors (or shocks) is more important than the weighted sum of past values of data itself. Taking the errors into consideration rather than or in conjunction with autoregression, is cru-

cial in finance when modelling exchange rates and volatility in particular [4].

Polynomial moving average (PMA) models are Volterra based NMA models with *linear-in-the-parameters* property and defined as:

$$x(n) = \mu + \sum_{i=1}^{q} b_i^{(1)} e(n-i) + \sum_{i=1}^{q} \sum_{j=1}^{q} b_{i,j}^{(2)} e(n-i)e(n-j) + ...$$
$$+ \sum_{i,...}^{q,...} b_{i,...}^{(p)} e(n-i)... + e(n), \quad (1)$$

where $e(n)$ is the excitation sequence with distribution $\mathcal{N}(0, \sigma_e^2)$, $b_i^{(1)}$, $b_{i,j}^{(2)}$ and $b_{i,...}^{(p)}$ are PMA model coefficients for first order, second order and p^{th} order polynomials, respectively, p is the nonlinearity degree and q is the MA order of the PMA model. A PMA model can be represented with the notation: $P^{(p)}MA(q)$. In modelling problems, it is important to be able to estimate the degree of nonlinearity p as well as the MA order and the process coefficients. To the best of our knowledge, very limited work on the estimation of PMA nonlinearity order exists.

Reversible jump Markov chain Monte Carlo (RJMCMC) was first introduced in [6] as a Bayesian model identification tool which is an extended and generalized version of the MCMC algorithm. RJMCMC provides an algorithm for the construction of reversible Markov chain samplers which explores parameter subspaces of different dimensions. General intention is to employ RJMCMC on problems which include exploring spaces of varying dimensions of the same classes of models. In the literature, RJMCMC has been generally used in linear model identification problems, e.g. in [7, 8] for *autoregressive* (AR), in [9] for *autoregressive integrated moving average* (ARIMA) and in [10] for *fractional* ARIMA (ARFIMA) models.

However, the formulation of Green in [6] provides a far more general usage potential for RJMCMC and its sampling strategy is not limited to linear models. RJMCMC can also be used for nonlinear model identification problems in a wide range of applications such as a model identification tool for

polynomial autoregressive (PAR) models in [11] and *threshold moving average* (TMA) models in [12].

In a previous study [11], we demonstrated the success of RJMCMC algorithm in the estimation of PAR processes with unknown degree of nonlinearity. In this study, we reformulate this PAR model estimation problem for the synthetically generated PMA models. The procedure differs from the previous studies [7–10] which apply RJMCMC for choosing a model from a set of linear models. In the present work, RJM-CMC explores linear and nonlinear models together and tests the nonlinearity of the model. Meanwhile, it estimates the nonlinearity degree and the MA order for the corresponding PMA model. In addition to the model selection procedure, PMA model coefficients are estimated as well.

The rest of the paper is organized as follows: PMA models and RJMCMC methodology for PMA model selection and the problem of estimating coefficients are examined in Section 2. The results of the simulations are provided in Section 2.5. Section 4 concludes the paper with a brief summary.

2. METHODOLOGY

2.1. PMA Models

A $\mathrm{P}^{(p)}\mathrm{MA}(q)$ model given by (1) can be represented in matrix-vector form by using the *linear-in-the-parameters* property:

$$\mathbf{x} = \mathbf{B}\epsilon^{(p,q)} + \mathbf{e}, \qquad (2)$$

where $\mathbf{x}$ is a $n-$vector of data samples, and $\mathbf{e}$ is a $n-$vector of excitation sequence with independent and identically distributed $\mathcal{N}(0, \sigma_e^2)$. $\mathbf{B}$ is a matrix whose rows are constructed in a way as to generate data in the correct form. Past samples and their polynomial products are included in a $(n \times w)-$vector, $\epsilon^{(p,q)}$. $\mathbf{B}$ and $\epsilon^{(p,q)}$ have the form:

$$\mathbf{B} = \begin{bmatrix} \mathbf{b^{(p,q)}} & \mathbf{0} & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \mathbf{b^{(p,q)}} & \mathbf{0} & \dots & \mathbf{0} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \dots & \mathbf{b^{(p,q)}} \end{bmatrix}, \qquad (3)$$

$$\epsilon^{(p,q)} = \begin{bmatrix} \epsilon_1^{(p,q)}, & \epsilon_2^{(p,q)}, & \epsilon_3^{(p,q)}, & \dots, & \epsilon_n^{(p,q)} \end{bmatrix}^{\mathrm{T}}. \quad (4)$$

where w refers to the number of coefficients of $\mathrm{P}^{(p)}\mathrm{MA(q)}$ model, $\mathbf{0}$ represents a zero row vector with w zeros, and $\mathbf{b^{(p,q)}}$ and $\epsilon_t^{(p,q)}$, for any $t = 1, 2, \dots, n$, have the form:

$$\mathbf{b^{(p,q)}} = \begin{bmatrix} b_1^{(1)} & b_2^{(1)} & \dots & b_q^{(1)} & b_{1,1}^{(2)} & b_{1,2}^{(2)} & \dots & b_{q,q}^{(2)} & \dots & b_{q,\dots}^{(q)} \end{bmatrix}, \quad (5)$$

$$\epsilon_t^{(p,q)} = [e(t-1), e(t-2), \dots, e(t-q), e^2(t-1),$$
$$e(t-1)e(t-2), \dots, e^2(t-q), \dots, e^p(t-q)]^{\mathrm{T}}. \quad (6)$$

2.2. Likelihood for PMA models

Studies [13,14], derive the likelihood for linear MA. In PMA, the Gaussianity of the likelihood is not guaranteed. However, in [15] it is shown that for white inputs and narrow-band Volterra systems, the output is Gaussian. We have experimentally verified this result and seen that the Gaussian likelihood is a good practical approximation for narrowband Volterra models.

In particular, for this case [13] provides an approximation on the likelihood function which attains estimations for the unobserved values in the model itself. This method has been employed in studies for Bayesian analysis of ARMA based time series models [8–10, 16].

An approximate likelihood function for a linear MA(q) process can be defined as [13],

$$f(\mathbf{x}|\theta) = \frac{1}{\sqrt{(2\pi\sigma_e^2)^{(n-q_{max})}}} \exp\left(\frac{-1}{2\sigma_e^2} \sum_{t=q_{max}+1}^{n} e_t^2\right) \quad (7)$$

$$\approx \mathcal{N}(\mathbf{e}|\mathbf{0}, \sigma_e^2 \mathbf{I_n}). \qquad (8)$$

where q_{max} represents the maximum MA order and n is the length of data vector $\mathbf{x}$ and excitation sequence vector $\mathbf{e}$. The authors of [13] have derived this form of likelihood under the assumption of both the MA process $\mathbf{x}$, and the excitation $\mathbf{e}$ are normally distributed.

The expression in (7) can be directly used for PMA models with $\mathbf{e} \sim \mathcal{N}(\mathbf{0}, \sigma_e^2 \mathbf{I_n})$ and parameter vector θ of $\{p, q, \mathbf{b^{(p,q)}}, \sigma_e^2, \sigma_b^2\}$, provided that the data coming from the PMA model is normally distributed.

As seen clearly, likelihood function in (7) requires elements of vector $\mathbf{e}$ to be known. However, excitation sequence is an unobserved quantity, approximations [17, 18] or sampling strategies [16] are available to solve this problem.

In this study, we apply a sampling strategy to employ the likelihood function in (7) by expressing the excitation in terms of $\mathbf{x}, \mathbf{B}$ and $\epsilon^{(p,q)}$ as $\mathbf{x} - \mathbf{B}\epsilon^{(p,q)}$ from (2). The details will be discussed in next sections.

2.3. Bayesian Hierarchy & Priors

The joint posterior density, i.e. target distribution of RJM-CMC, $f(\theta|\mathbf{x})$, can be easily written from Bayes Theorem:

$$f(p, q, \mathbf{b^{(p,q)}}, \sigma_e^2, \sigma_b^2|\mathbf{x}) \propto f(\mathbf{x}|p, q, \mathbf{b^{(p,q)}}, \sigma_e^2) \times$$
$$f(\mathbf{b^{(p,q)}}|p, q, \sigma_b^2) f(\sigma_b^2) f(\sigma_e^2) f(q) f(p). \quad (9)$$

In previous studies [7–11], making an assumption that all model subspaces are equally likely *a priori*, appears as a natural choice in the absence of real prior information about model orders of an observed data. Given these, we assume that the MA order q, and the nonlinearity degree p are uniformly distributed with upper bounds p_{max} and q_{max}:

$$f(q) = \mathcal{U}(1, q_{\max}) \qquad \text{and} \qquad f(p) = \mathcal{U}(1, p_{\max}). \quad (10)$$

In order to provide conditional conjugacy, prior for PMA coefficients are assumed to be normally distributed with zero mean and σ_b^2 variance. Also, the excitation variance σ_e^2 and variance of model coefficients, σ_b^2, are assumed as inverse-Gamma distributed. This choice of priors for scale parameters is due to the conditional conjugacy of the inverse-Gamma distribution. Then, the full posterior conditional distribution becomes also inverse-Gamma [9]:

$$f(\mathbf{b^{(p,q)}}|p, q, \sigma_b^2) = \mathcal{N}(\mathbf{b^{(p,q)}}|\mathbf{0}, \sigma_b^2 \mathbf{I_w}), \qquad (11)$$

$$f(\sigma_b^2) = \mathcal{IG}(\sigma_b^2|\alpha_b, \beta_b), \qquad (12)$$

$$f(\sigma_e^2) = \mathcal{IG}(\sigma_e^2|\alpha_e, \beta_e). \qquad (13)$$

2.4. RJMCMC Methodology

RJMCMC [6] is a general strategy of sampling from a target distribution, $f(\theta|\mathbf{x})$, whether the dimensions of parameter spaces are the same or not. It applies the standard MCMC strategy for within-model moves, which we call as *life* move, and reversible jump strategy for between-model moves, namely *birth* and *death* moves.

Following [6], when the current Markov chain state is κ with parameter vector θ, we propose a move type m with probability $Pr(\kappa \to \kappa')$, which changes dimension, and takes the state to κ' with parameter vector θ^*. The acceptance probability, which is denoted by $\alpha(\kappa \to \kappa')$, needs to be calculated to ensure convergence to the correct posterior.

Given the observed data $\mathbf{x}$, the general expression for $\alpha(\kappa \to \kappa')$ which is similar to eqn. (8) of [6], is;

$$\min\left\{1, \frac{f(\theta^*|\mathbf{x})Pr(\kappa' \to \kappa)}{f(\theta|\mathbf{x})Pr(\kappa \to \kappa')\chi(\mathbf{u})}\left|\frac{\partial \theta^*}{\partial(\theta, \mathbf{u})}\right|\right\}, \quad (14)$$

where $f(\cdot|\mathbf{x})$ is the target distribution of interest, $Pr(\kappa \to \kappa')$ and $Pr(\kappa' \to \kappa)$ represent the probabilities for the move m and its reverse move, $\chi(\mathbf{u})$ is the proposal distribution for the auxiliary variable vector $\mathbf{u}$ which is required to provide dimension matching for the move m and $\left|\frac{\partial \theta^*}{\partial(\theta, \mathbf{u})}\right|$ is the magnitude of the Jacobian.

2.4.1. Between-Model Moves (Birth & Death)

For a **birth** move from q to q' where p is fixed, the acceptance ratio is defined as $\alpha_{\text{birth}} = \min\{1, r_{\text{birth}}\}$. The corresponding value for r_{birth}:

$$r_{\text{birth}} = \frac{f(\mathbf{x}|p, q', \mathbf{b^{(p,q')}}, \sigma_e^2)}{f(\mathbf{x}|p, q, \mathbf{b^{(p,q)}}, \sigma_e^2)} \times \frac{f(\mathbf{b^{(p,q')}}|p, q', \sigma_b^2)}{f(\mathbf{b^{(p,q)}}|p, q, \sigma_b^2)}$$
$$\times \frac{P_{\text{death}}}{P_{\text{birth}}\chi(\mathbf{u})} \times \left|\frac{\partial \mathbf{b^{(p,q')}}}{\partial(\mathbf{b^{(p,q)}}, \mathbf{u})}\right|. \qquad (15)$$

If a move from q to q' where $q' < q$, is selected, no new parameters are proposed and a **death** move will be applied.

We remove the coefficients which belong to q of parameter vector $\mathbf{b^{(p,q)}}$. The acceptance ratio of the death move, $\alpha_{\text{death}}(q \to q') = \min\{1, 1/r'_{\text{birth}}\}$.

2.4.2. Within-Model Move (Life)

When the newly proposed model order is equal to the recent value, e.g. $q' = q$, a **life** move will be applied. Acceptance ratio is defined as $\alpha_{\text{life}} = \min\{1, r_{\text{life}}\}$. So r_{life} is:

$$r_{\text{life}} = \frac{f(\mathbf{x}|p, q', \mathbf{b^{(p,q')}}, \sigma_e^2)}{f(\mathbf{x}|p, q, \mathbf{b^{(p,q)}}, \sigma_e^2)} \times \frac{f(\mathbf{b^{(p,q')}}|p, q', \sigma_b^2)}{f(\mathbf{b^{(p,q)}}|p, q, \sigma_b^2)}$$
$$\times \frac{\psi(\mathbf{b^{(p,q)}}|p, q', \mathbf{b^{(p,q')}})}{\psi(\mathbf{b^{(p,q')}}|p, q, \mathbf{b^{(p,q)}})} \qquad (16)$$

where $f(\mathbf{x}|\cdot)$ is likelihood distribution and $f(\mathbf{b^{(p,q')}}|\cdot)$ is prior distribution for parameter vector. Updating model coefficients includes proposing from the distribution $\psi(\cdot)$;

$$\mathbf{b^{(p,q')}} \sim \psi(\mathbf{b^{(p,q')}}|p, q, \mathbf{b^{(p,q)}}), \qquad (17)$$

$$= \mathcal{N}(\mathbf{b^{(p,q')}}|\mu_n, \Sigma_n^{-1}), \qquad (18)$$

where $\mu_n = \sigma_e^{-2}\Sigma_n^{-1}\mathbf{X^T x}$ and $\Sigma_n = \sigma_e^{-2}\mathbf{X}^T\mathbf{X} + \sigma_b^{-2}\mathbf{I_w}$.

Each row of $(n \times w)$-matrix $\mathbf{X}$ consists of $\epsilon_t^{(p,q)}$ for row t (See expression in (6)).

2.4.3. Gibbs Moves

Excitation variance, σ_e^2 and variance of model coefficients, σ_b^2 are updated at each iteration via Gibbs Sampling. The full conditional distribution for σ_e^2 is [8];

$$f(\sigma_e^2|\mathbf{x}, p, q, \mathbf{b^{(p,q)}}) \propto f(\mathbf{x}|p, q, \mathbf{b^{(p,q)}}, \sigma_e^2)f(\sigma_e^2) \qquad (19)$$

$$\approx \mathcal{N}(\mathbf{e}|\mathbf{0}, \sigma_e^2\mathbf{I_n})\mathcal{IG}(\sigma_e^2|\alpha_e, \beta_e) \qquad (20)$$

$$= \mathcal{IG}(\sigma_e^2|\alpha_{en}, \beta_{en}), \qquad (21)$$

where $\alpha_{en} = \alpha_e + \frac{1}{2}n$, $\beta_{en} = \beta_e + \frac{1}{2}\mathbf{e}^T\mathbf{e}$ and excitation sequence vector $\mathbf{e} = \mathbf{x} - \mathbf{B}\epsilon^{(p,q)}$ from (2).

Similarly, the full conditional distribution for σ_b^2 is obtained as [8];

$$f(\sigma_b^2|\mathbf{x}, p, q, \mathbf{b^{(p,q)}}) \propto f(\mathbf{b^{(p,q)}}|\sigma_b^2)f(\sigma_b^2) \qquad (22)$$

$$\approx \mathcal{N}(\sigma_b^2|\mathbf{0}, \sigma_b^2\mathbf{I}_w)\mathcal{IG}(\sigma_b^2|\alpha_b, \beta_b) \qquad (23)$$

$$= \mathcal{IG}(\sigma_b^2|\alpha_{bn}, \beta_{bn}), \qquad (24)$$

where $\alpha_{bn} = \alpha_b + \frac{1}{2}w$ and $\beta_{bn} = \beta_b + \frac{1}{2}(\mathbf{b^{(p,q)}})^T\mathbf{b^{(p,q)}}$.

Model estimation procedure in this study requires past samples of unobserved excitation sequence. These samples and their polynomial extensions are elements of the vector, $\epsilon^{(p,q)}$ in (2) and the matrix, $\mathbf{X}$ in (18).

In [16], a Gibbs sampling methodology for ARMA models has been constructed for reconstruction of audio signals. Applying the method of [16] for MA models and employing

second order method used in [9], the full conditional distribution for vector $\epsilon^{(p,q)}$ is obtained as;

$$f(\epsilon^{(p,q)}|x,p,q,\mathbf{b}^{(\mathbf{p},\mathbf{q})},\sigma_e^2) = \mathcal{N}(\epsilon^{(p,q)}|\mu_\epsilon, \Sigma_\epsilon^{-1}), \quad (25)$$

where, $\mu_\epsilon = \sigma_e^{-2}\Sigma_\epsilon^{-1}\mathbf{B}^\mathsf{T}\mathbf{x}$ and $\Sigma_\epsilon = \sigma_e^{-2}\left(\mathbf{B}^\mathsf{T}\mathbf{B} + \mathbf{I}_n\right)$.

2.5. PMA Model Selection Procedure

PMA model selection procedure requires 2-stage RJMCMC procedure for q and p at each iteration. Equations between (15)-(18) correspond to the 1st RJMCMC stage used to estimate q when p is fixed to its value in previous iteration. In the 2nd stage, q is set to the recently estimated value in the first stage and p is estimated which can be described modifying Equations (15)-(18) suitably.

Each RJMCMC stage updates one of the model parameters that is either p or q. In case of a birth move corresponding to a model change from $p = 2$ to $p = 3$ when $q = 2$, that move requires $\lambda = 9 - 5 = 4$ candidate coefficients to be proposed from a proposal distribution to satisfy the dimension matching criteria. Each element, u_i, of the proposed vector $\mathbf{u}$, where $i = 1, \ldots, \lambda$ is proposed from a uniform distribution $\mathcal{U}(-\delta, \delta)$ and the joint distribution $\chi(\mathbf{u})$ is defined as;

$$\chi(\mathbf{u}) = \prod_{i=1}^{\lambda}\mathcal{U}(-\delta, \delta), \quad \text{and} \quad \delta = \frac{0.05}{E[|\mathbf{x}|]}, \quad (26)$$

where $E[|\mathbf{x}|]$ is the expected value of the absolute value of the data vector $\mathbf{x}$. In particular, proposal distribution borders, (δ and $-\delta$) are selected as depending on observed data due to sample meaningful candidates under varying conditions of data. Moreover, proposal distribution is chosen to make the candidates independent from recent coefficients, thus the increase in dimensionality of the parameter space is accomplished through an identity function. This selection makes the Jacobian equal to unity.

3. SIMULATION & RESULTS

6 different PMA models (2 linear and 4 nonlinear) are generated for simulations. Each data set has a length of 500 samples. Mean value, μ, for each data sets is chosen as 0 for simplicity. Each data set is driven with a Gaussian excitation sequence with variance of σ_e^2. Normality of the each data set is tested via Kolmogorov-Smirnov and Kullback-Leibler tests in order to provide that the PMA process, $\mathbf{x}$ is normally distributed.

Hyperparameters are set to values $\alpha_e = \alpha_b = 1$ and $\beta_e = \beta_b = 2$. The initial MA order q_0 and nonlinearity degree p_0 are set to 1 and both of upper bounds p_{max} and q_{max} are set to 6. $\mathbf{b}^{(\mathbf{p_0},\mathbf{q_0})}$ is sampled from the prior distribution in (11). Move probabilities for P_{birth}, P_{death} and P_{life} are selected as 0.15, 0.15 and 0.7, respectively. $10,000$ iterations are simulated to let sampled parameters converge.

Table 1. Model Estimation Results

	$P^{(1)}$MA(4)	$P^{(1)}$MA(5)	$P^{(2)}$MA(2)
Percentage of Detection	70%	70%	100%
Avg. NMSE of Coeff. Vector Estimate	0.0287	0.0448	0.0174
# of Coeff. (w)	4	5	5
	$P^{(2)}$MA(3)	$P^{(3)}$MA(2)	$P^{(4)}$MA(1)
Percentage of Detection	100%	65%	70%
Avg. NMSE of Coeff. Vector Estimate	0.0198	0.0308	0.0531
# of Coeff. (w)	9	9	4

Because of the space limitation, 2 results out of 6 PMA models will be shown as examples. In Figure 1 histograms of the four model coefficients of $P^{(2)}$MA(3) and the excitation sequence are plotted. The *normalized mean square error* (NMSE) of the coefficient vector in Figure 1 is 1.39×10^{-2} which is defined as NMSE $= \frac{1}{w}\sum_{i=1}^{w}\frac{(b_i - \hat{b}_i)^2}{\|\mathbf{b}\|_2}$ where b_i and $\hat{b}_i$ are the ith element of the w-dimensional coefficient vector $\mathbf{b}$ and its estimate $\hat{\mathbf{b}}$ and $\|\mathbf{b}\|_2$ is the l_2-norm of $\mathbf{b}$.

In Figure 2, the instantaneous estimates and the joint posterior density of the model orders p and q for model $P^{(2)}$MA(2) are shown for one of the iterations. The proposed method estimates the true order pair over 50% of the iterations. Consequently, by using RJMCMC, we can decide the nonlinearity degree of PMA model and the resulting parameters can be used in an application to find the best model.

Table 1 depicts the RJMCMC percentage of detection of true model order pairs and estimated model coefficient error values in terms of NMSE values. 20 different data sets are generated for each model and RJMCMC model detection performance is measured. RJMCMC decides true model order pairs with highest percentage within 20 simulations for all 6 candidate PMA models. Moreover, RJMCMC also estimates model coefficients of unknown models and achieves remarkable performance in terms of average NMSE of around 10^{-2} for all models.

4. CONCLUSIONS

RJMCMC algorithm has been known for its success in the solution of model uncertainty of linear processes. This study have used RJMCMC in an anomalous case and demonstrates the potential of it in estimating the degree of nonlinearity of a linear in the parameters nonlinear model.

Furthermore, by sampling in linear and nonlinear spaces of varying dimensions for the estimation of orders p and q, we also show that RJMCMC algorithm can jump (hence explore) between not only spaces with different dimensions, but also different classes of models (MA and PMA).

RJMCMC is shown to be a complete model estimation tool which not only determines model order by applying a search on parameter space but also estimates model coefficients, concurrently.

As a future work, the proposed method will be adapted to a Volterra Systems Identification problem to apply in problems such as dynamical system identification, equalizing the nonlinearities in fiber optical systems and communications systems etc.
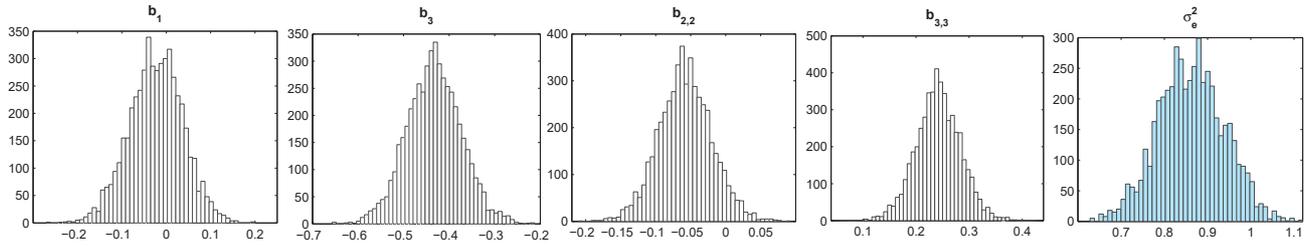
Fig. 1. Histograms for $P^{(2)}MA(3)$ Parameters ($\mathbf{b}^* = [0, -0.5, -0.18, 0.45], \sigma_e^2 = 0.7$)
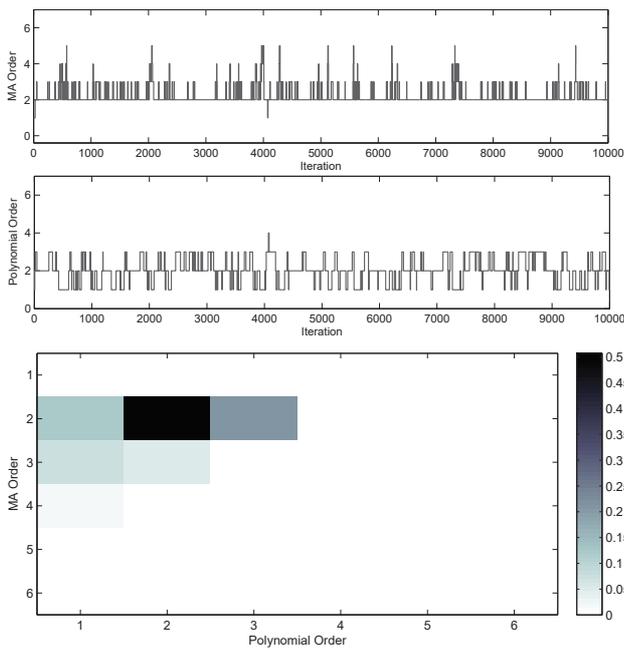
Fig. 2. The instantaneous model order estimates and the joint posterior density of the model orders of $P^{(2)}MA(2)$

5. REFERENCES

[1] P. M. Robinson, "The estimation of a nonlinear moving average model," *Stochastic processes and their applications*, vol. 5, no. 1, pp. 81–90, 1977.

[2] M. Isaksson, D. Wisell, and D. Rönnow, "A comparative analysis of behavioral models for RF power amplifiers," *Microwave Theory and Techniques, IEEE Transactions on*, vol. 54, no. 1, pp. 348–359, 2006.

[3] A. Kareem and T. Wu, "Changing dynamic of bridge aerodynamics," *Proceedings of the ICE-Structures and Buildings*, vol. 168, no. 2, pp. 94–106, 2014.

[4] P. Zaffaroni and B. d'Italia, "Gaussian inference on certain long-range dependent volatility models," *Journal of econometrics*, vol. 115, no. 2, pp. 199–258, 2003.

[5] Y. Chikkula and J. H. Lee, "Robust adaptive predictive control of nonlinear processes using nonlinear moving average system models," *Industrial & Engineering Chemistry Research*, vol. 39, no. 6, pp. 2010–2023, 2000.

[6] P. J. Green, "Reversible jump Markov chain Monte Carlo computation and Bayesian model determination," *Biometrika*, vol. 82, no. 4, pp. 711–732, 1995.

[7] J. Vermaak, C. Andrieu, A. Doucet, and S. J. Godsill, "Reversible jump Markov chain Monte Carlo strategies for Bayesian model selection in autoregressive processes," *Journal of Time Series Analysis*, vol. 25, no. 6, pp. 785–809, 2004.

[8] P. T. Troughton and S. J. Godsill, "A reversible jump sampler for autoregressive time series," in *ICASSP 1998 IEEE*. IEEE, 1998, vol. 4, pp. 2257–2260.

[9] R. S. Ehlers and S. P. Brooks, "Bayesian analysis of order uncertainty in ARIMA models," *Federal University of Parana, Tech. Rep*, 2004.

[10] E. Eğri, S. Günay, et al., "Bayesian model selection in ARFIMA models," *Expert Systems with Applications*, vol. 37, no. 12, pp. 8359–8364, 2010.

[11] O. Karakuş, E. E. Kuruoğlu, and M. A. Altınkaya, "Estimation of the nonlinearity degree for polynomial autoregressive processes with RJMCMC," in *Signal Processing Conference (EUSIPCO), 2015 23rd European*. IEEE, 2015, pp. 953–957.

[12] Q. Xia, J. Pan, Z. Zhang, and J. Liu, "A Bayesian nonlinearity test for threshold moving average models," *Journal of Time Series Analysis*, vol. 31, no. 5, pp. 329–336, 2010.

[13] G. E. P. Box, G. M. Jenkins, and G. C. Reinsel, *Time series analysis: forecasting and control*, vol. 734, John Wiley & Sons, 2011.

[14] S. C. Hillmer and G. C. Tiao, "Likelihood function of stationary multiple autoregressive moving average models," *Journal of the American Statistical Association*, vol. 74, no. 367, pp. 652–660, 1979.

[15] M. A. Ivanov, "Distribution of the output reponse of Volterra-type narrowband nonlinear-systems," *Telecommunications and Radio Engineering*, vol. 41, no. 12, pp. 99–102, 1987.

[16] S.J. Godsill, "Bayesian enhancement of speech and audio signals which can be modelled as ARMA processes," *International Statistical Review*, vol. 65, no. 1, pp. 1–21, 1997.

[17] S.M. Shaarawy and S.S. Ali, "Bayesian model order selection of vector moving average processes," *Communications in Statistics-Theory and Methods*, vol. 41, no. 4, pp. 684–698, 2012.

[18] J. Marriott, N. Ravishanker, and A.E. Gelfand, "Bayesian analysis of ARMA processes: complete sampling based inference under full likelihoods," Tech. Rep., DTIC Document, 1993.