

# Nonlinear Model Selection for PARMA Processes Using RJMCMC

Oktay Karakuş  
İzmir Institute of Technology  
Electrical-Electronics Engineering  
Urla, İzmir, Turkey  
Email: oktaykarakuş@iyte.edu.tr

Ercan E. Kuruoğlu  
ISTI-CNR,  
via G. Moruzzi 1, 56124,  
Pisa, Italy  
Email: ercan.kuruoglu@isti.cnr.it

Mustafa A. Altinkaya  
İzmir Institute of Technology  
Electrical-Electronics Engineering  
Urla, İzmir, Turkey  
Email: mustafaaltinkaya@iyte.edu.tr

**Abstract**—Many prediction studies using real life measurements such as wind speed, power, electricity load and rain-fall utilize linear *autoregressive moving average* (ARMA) based models due to their simplicity and general character. However, most of the real life applications exhibit nonlinear character and modelling them with linear time series may become problematic. Among nonlinear ARMA models, *polynomial ARMA* (PARMA) models belong to the class of linear-in-the-parameters. In this paper, we propose a *reversible jump Markov chain Monte Carlo* (RJMCMC) based complete model estimation method which estimates PARMA models with all their parameters including the nonlinearity degree. The proposed method is unique in the manner of estimating the nonlinearity degree and all other model orders and model coefficients at the same time. Moreover, in this paper, RJMCMC has been examined in an anomalous way by performing transitions between linear and nonlinear model spaces.

## I. INTRODUCTION

*Autoregressive moving average* (ARMA) models have been utilized often especially in time series prediction studies due to the ease of estimation by methods such as that of Box-Jenkins [1] and their generality encompassing both *autoregressive* (AR) and *moving average* (MA) models. ARMA models have been used in diverse areas of applications such as speech [2], [3], seismology [4], video [5], image [6], etc. Particularly, they have been applied in energy and meteorological prediction studies of solar radiation [7], [8], electricity demand [9], [10] and wind speed [11], [12].

However, many real life problems exhibit nonlinear characters. *Nonlinear ARMA* (NARMA) or specifically, *polynomial ARMA* (PARMA) models appear as alternatives in modelling problems including nonlinearity and have been used in prediction studies in many different areas such as in energy prediction studies for electricity load [13], wind speed [14], modelling a nonlinear networked control system [15] and forecasting of financial returns [16].

Reversible jump Markov chain Monte Carlo (RJMCMC) is a Bayesian model identification method which was firstly introduced in [17]. RJMCMC has been defined as an extended and generalized version of the MCMC algorithm and offers transitions between parameter spaces of different dimensions. In the literature, RJMCMC has been generally used in linear model estimation studies such as AR models [18], [19],

*autoregressive integrated moving average* (ARIMA) models [20] and *fractional* ARIMA models [21].

Contrary to the general practice which employs RJMCMC on problems which include exploring spaces of varying dimensions of the same classes of models, RJMCMC offers much wider interpretation than just exploring spaces of different dimensions. In previous studies [22], [23], we have shown that RJMCMC can be used as a model determination method which performs transitions between linear and nonlinear model spaces for *polynomial AR* (PAR) and *polynomial MA* (PMA) models.

In this paper, the formulation in previous studies [22], [23] has been reconstructed into a general framework which estimates the nonlinearity degree of PARMA models. In addition to the nonlinearity degree, AR and MA orders and all the model coefficients have been estimated at the same time. Results have been depicted in figures and tables for synthetically generated PARMA models.

The rest of the paper is organized as follows: PARMA models and general RJMCMC definition are examined in Sections II and III, respectively. In Section IV, RJMCMC construction for PARMA model estimation is studied in detail. Numerical results about the study are presented with figures and a table in Section V. Section VI concludes the paper with a brief summary.

## II. PARMA MODELS

The Volterra based *linear-in-the-parameters* PARMA models can be defined as:

$$x_l = \mu + \sum_{d=1}^p \sum_{\tau_1=1}^k \dots \sum_{\tau_d=\tau_{d-1}}^k a_{\tau_1, \dots, \tau_d}^{(d)} \prod_{j=1}^d x_{l-\tau_j} + e_l + \sum_{d=1}^p \sum_{\tau_1=1}^q \dots \sum_{\tau_d=\tau_{d-1}}^q b_{\tau_1, \dots, \tau_d}^{(d)} \prod_{j=1}^d e_{l-\tau_j}, \quad (1)$$

where  $l = 1, \dots, n$  and  $p, k$  and  $q$  refer to the nonlinearity degree, the AR order and the MA memory of the model, respectively. In addition, vector  $\mathbf{e} = [e_1, e_2, \dots, e_n]$  represents the *iid* excitation sequence with distribution  $\mathcal{N}(\mathbf{0}, \sigma_e^2 \mathbf{I}_n)$ ,  $a_{\tau_1, \dots, \tau_d}^{(d)}$  and  $b_{\tau_1, \dots, \tau_d}^{(d)}$  are the  $d^{\text{th}}$  ordered coefficients of AR and

MA parts, respectively. A PARMA model can be represented with the notation:  $P^{(p)}\text{ARMA}(k, q)$ .

The system of equations in (1) can be easily represented in matrix vector form for a  $P^{(p)}\text{ARMA}(k, q)$  model by using the linear-in-the-parameters property:

$$\mathbf{x} = \mu + \mathbf{X}\mathbf{a}^{(p,k)} + \mathbf{E}\mathbf{b}^{(p,q)} + \mathbf{e}, \quad (2)$$

where  $n \times \eta$  matrix  $\mathbf{X}$ ,  $\eta$ -vector  $\mathbf{a}^{(p,k)}$ ,  $n \times \gamma$  matrix  $\mathbf{E}$  and  $\gamma$ -vector  $\mathbf{b}^{(p,q)}$  are given as:

$$\mathbf{X} = \begin{bmatrix} x_0 & \dots & x_{1-k} & x_0^2 & x_0x_{-1} & \dots & x_{1-k}^p \\ x_1 & \dots & x_{2-k} & x_1^2 & x_1x_0 & \dots & x_{2-k}^p \\ \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ x_{n-1} & \dots & x_{n-k} & x_{n-1}^2 & x_{n-1}x_{n-2} & \dots & x_{n-k}^p \end{bmatrix}, \quad (3)$$

$$\mathbf{a}^{(p,k)} = \begin{bmatrix} a_1^{(1)}, \dots, a_k^{(1)}, a_{1,1}^{(2)}, a_{1,2}^{(2)}, \dots, a_{k,k}^{(2)}, \dots, \underbrace{a_{k,k,\dots,k}^{(p)}}_{p^{\text{th}} \text{ order}} \end{bmatrix}^T, \quad (4)$$

$$\mathbf{E} = \begin{bmatrix} e_0 & \dots & e_{1-q} & e_0^2 & e_0e_{-1} & \dots & e_{1-q}^p \\ e_1 & \dots & e_{2-q} & e_1^2 & e_1e_0 & \dots & e_{2-q}^p \\ \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ e_{n-1} & \dots & e_{n-q} & e_{n-1}^2 & e_{n-1}e_{n-2} & \dots & e_{n-q}^p \end{bmatrix}, \quad (5)$$

$$\mathbf{b}^{(p,q)} = \begin{bmatrix} b_1^{(1)}, \dots, b_q^{(1)}, b_{1,1}^{(2)}, b_{1,2}^{(2)}, \dots, b_{q,q}^{(2)}, \dots, \underbrace{b_{q,q,\dots,q}^{(p)}}_{p^{\text{th}} \text{ order}} \end{bmatrix}^T, \quad (6)$$

where  $\eta$  and  $\gamma$  are the lengths of parameter vectors  $\mathbf{a}^{(p,k)}$  and  $\mathbf{b}^{(p,q)}$ , respectively.

### III. REVERSIBLE JUMP MCMC

Following [17], assume that we propose a move  $m$  with probability  $p_m$  from a Markov chain state  $\kappa$  to  $\kappa'$ . These states have parameter vectors  $\boldsymbol{\theta}$  and  $\boldsymbol{\theta}^*$ , respectively with different dimensions. Given the observed data  $\mathbf{x}$ , the general expression for the acceptance ratio,  $\alpha(\kappa \rightarrow \kappa')$ , is;

$$\min \left\{ 1, \frac{f(\boldsymbol{\theta}^*|\mathbf{x})p_m^R}{f(\boldsymbol{\theta}|\mathbf{x})p_m\chi(\mathbf{u})} \left| \frac{\partial \boldsymbol{\theta}^*}{\partial(\boldsymbol{\theta}, \mathbf{u})} \right| \right\}, \quad (7)$$

where  $p_m^R$  represents the probability for the reverse move of  $m$ ,  $\chi(\mathbf{u})$  is the proposal distribution for the auxiliary variable vector  $\mathbf{u}$  which is required to provide dimension matching for the move  $m$  and  $\left| \frac{\partial \boldsymbol{\theta}^*}{\partial(\boldsymbol{\theta}, \mathbf{u})} \right|$  is the magnitude of the Jacobian.

In each RJMCMC run, the standard Metropolis-Hastings algorithm is applied for within-model moves, which we call as *life* move. For between-model moves, namely *birth* and *death* moves, reversible jump mechanism is applied.

### IV. RJMCMC CONSTRUCTION FOR PARMA MODEL ESTIMATION

PARMA model estimation via RJMCMC includes 3 stages at each iteration. At each stage, one of the model parameters ( $p, k$  or  $q$ ) has been proposed while the others remain at their recent values. Model order in question is increased and decreased with probabilities  $P_{\text{birth}}$  and  $P_{\text{death}}$ , respectively and remains at the same order with probability  $P_{\text{life}}$ . If the

proposed model order is accepted, RJMCMC performs a training procedure by applying 10 consecutive life moves to update the newly estimated model coefficients. All these steps have been repeated at each stage for each model order ( $p, k$  or  $q$ ).

In the case of a birth move corresponding to a model change from  $k = 2$  to  $k' = 3$  when  $q = 0$  and  $p = 2$ , in order to satisfy the dimension matching, it's required to propose  $\lambda = 9 - 5 = 4$  candidate model coefficients from a proposal distribution,  $\chi(\mathbf{u})$ , which is:

$$\chi(\mathbf{u}) = \prod_{i=1}^{\lambda} \mathcal{U}(-\delta, \delta), \quad \text{and} \quad \delta = \frac{0.02}{E[|\mathbf{x}|]}, \quad (8)$$

where  $E[|\mathbf{x}|]$  is the expectation of the absolute value of the observed data vector  $\mathbf{x}$ .

The upper and lower bounds of the joint distribution are chosen to depend on the given data. By doing so, distinct limits can be arisen due to the fact that the magnitudes of different data sets may vary in different ranges. Consequently, this *ad hoc* choice adds variety for the candidates. Moreover, after some trial-error process, we see that keeping the interval narrower (smaller  $\delta$  values) increases the number of acceptable moves. That's why we have chosen a value of 0.02 for the nominator part in equation (8).

Furthermore, each of the newly proposed model coefficients have been chosen to be independent from the recent coefficients which causes a unity Jacobian.

RJMCMC construction for PARMA model estimation is presented in detail in the sequel.

#### A. Bayesian Hierarchy

The target distribution of RJMCMC is a joint posterior density and can be expressed as  $f(\boldsymbol{\theta}|\mathbf{x})$  with parameter vector  $\boldsymbol{\theta} = \{p, k, q, \mathbf{a}^{(p,k)}, \mathbf{b}^{(p,q)}, \sigma_e^2, \sigma_a^2, \sigma_b^2\}$ . This density can be decomposed via Bayes Theorem:

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

#### B. Likelihood Definition

In a previous study [24], it has been stated that for white inputs and narrowband Volterra based models output follows Gaussian characteristics. We have experimentally verified this by testing the output process of a PARMA model.

Furthermore, in [25] a Gaussian approximate likelihood has been used in restoration of nonlinearly distorted AR signals when the whole system follows a nonlinear AR model characteristics. Consequently, choosing a Gaussian likelihood for PARMA models appears as a good practical approximation.

Moreover, in [1], an approximation of the likelihood function of ARMA models has been provided by performing a recursive estimation procedure in the model itself for the unobserved excitation values. This approach has been used in model estimation studies for Bayesian analysis of ARMA based time series models [19]–[21].

Thus, an approximate likelihood function for a  $P^{(p)}$ ARMA( $k, q$ ) has been examined in this study as,

$$f(\mathbf{x}|\boldsymbol{\theta}) = \frac{1}{\sqrt{(2\pi\sigma_e^2)^{(n-q_{\max})}}} \exp\left(\frac{-1}{2\sigma_e^2} \sum_{t=q_{\max}+1}^n e_t^2\right) \quad (10)$$

$$\approx \mathcal{N}(\mathbf{x} - \mathbf{X}\mathbf{a}^{(p,k)} - \mathbf{E}\mathbf{b}^{(p,q)} | \mathbf{0}, \sigma_e^2 \mathbf{I}_{n-q_{\max}}). \quad (11)$$

Please see Section IV-F for the details of the estimation process of the unobserved excitation sequence.

### C. Priors

In the previous studies, general practice is to use uniform priors for model order parameters [19]–[23]. Correspondingly, in this study, we assume that each model is equally likely and we choose uniform priors for model orders  $p, k$ , and  $q$  with upper bounds  $p_{\max}, k_{\max}$  and  $q_{\max}$ , respectively.

$$f(p) = \mathcal{U}(1, p_{\max}), \quad f(k) = \mathcal{U}(1, k_{\max}), \quad f(q) = \mathcal{U}(1, q_{\max}). \quad (12)$$

We choose priors for model coefficients and variances as shown below so as to provide conditional conjugacy:

$$f(\mathbf{a}^{(p,k)} | p, k, \sigma_a^2) = \mathcal{N}(\mathbf{a}^{(p,k)} | \mathbf{0}, \sigma_a^2 \mathbf{I}_w), \quad (13)$$

$$f(\mathbf{b}^{(p,q)} | p, q, \sigma_b^2) = \mathcal{N}(\mathbf{b}^{(p,q)} | \mathbf{0}, \sigma_b^2 \mathbf{I}_z), \quad (14)$$

$$f(\sigma_a^2) = \mathcal{IG}(\sigma_a^2 | \alpha_a, \beta_a), \quad (15)$$

$$f(\sigma_b^2) = \mathcal{IG}(\sigma_b^2 | \alpha_b, \beta_b), \quad (16)$$

$$f(\sigma_e^2) = \mathcal{IG}(\sigma_e^2 | \alpha_e, \beta_e). \quad (17)$$

where  $w$  and  $z$  are the length of coefficients vectors  $\mathbf{a}^{(p,k)}$  and  $\mathbf{b}^{(p,q)}$ , respectively,  $\mathcal{IG}(\cdot)$  refers to an inverse Gamma distribution and  $\alpha_a, \beta_a, \alpha_b, \beta_b, \alpha_e, \beta_e$ , are hyperparameters for variances.

### D. Between Model Moves

Between model moves require to change dimension of parameter space. If proposed move requires an increase in dimension, RJMCMC performs a **birth** move. In the contrary case, **death** move has been performed.

Assume a **birth** move has been proposed from AR order  $k$  to  $k'$  when  $p$  and  $q$  are fixed. Then, resulting acceptance ratio is  $\alpha_{\text{birth}} = \min\{1, r_{\text{birth}}\}$  where  $r_{\text{birth}}$  can be given as:

$$r_{\text{birth}} = \frac{f(\mathbf{x} | p, k', q, \mathbf{a}^{(p,k')}, \mathbf{b}^{(p,q)}, \sigma_e^2)}{f(\mathbf{x} | p, k, q, \mathbf{a}^{(p,k)}, \mathbf{b}^{(p,q)}, \sigma_e^2)} \times \frac{f(\mathbf{a}^{(p,k')} | p, k', \sigma_a^2)}{f(\mathbf{a}^{(p,k)} | p, k, \sigma_a^2)} \times \frac{P_{\text{death}}}{P_{\text{birth}} \chi(\mathbf{u})} \times \left| \frac{\partial \mathbf{a}^{(p,k')}}{\partial \mathbf{a}^{(p,k)}, \mathbf{u}} \right|. \quad (18)$$

When a **death** move has been proposed from  $k$  to  $k'$ , no new parameters are being proposed. Since birth and death moves are reversible move pairs,  $\alpha_{\text{death}}(k \rightarrow k')$  can be written in terms of  $\alpha_{\text{birth}}(k' \rightarrow k)$ . By definition, if

$$\alpha_{\text{birth}}(k' \rightarrow k) = \min\{1, r'_{\text{birth}}\}, \quad (19)$$

then we can directly write

$$\alpha_{\text{death}}(k \rightarrow k') = \min\{1, 1/r'_{\text{birth}}\}. \quad (20)$$

### E. Within Model Moves

When  $k' = k$ , a **life** move will be proposed with acceptance ratio,  $\alpha_{\text{life}} = \min\{1, r_{\text{life}}\}$  with

$$r_{\text{life}} = \frac{f(\mathbf{x} | p, k', q, \mathbf{a}^{(p,k')}, \mathbf{b}^{(p,q)}, \sigma_e^2)}{f(\mathbf{x} | p, k, q, \mathbf{a}^{(p,q)}, \mathbf{b}^{(p,q)}, \sigma_e^2)} \times \frac{f(\mathbf{a}^{(p,k')} | p, k', \sigma_a^2)}{f(\mathbf{a}^{(p,k)} | p, k, \sigma_a^2)} \times \frac{\psi(\mathbf{a}^{(p,k')} | p, k', \mathbf{a}^{(p,k')})}{\psi(\mathbf{a}^{(p,k')} | p, k, \mathbf{a}^{(p,k)})} \quad (21)$$

where the proposal distribution  $\psi(\cdot)$  is:

$$\psi(\mathbf{a}^{(p,k)} | p, k', \mathbf{a}^{(p,k')}) = \mathcal{N}(\mathbf{a}^{(p,k')} | \mu_n, \Sigma_n^{-1}), \quad (22)$$

with  $\mu_n = \sigma_e^{-2} \Sigma_n^{-1} \mathbf{X}^T (\mathbf{x} - \mathbf{B}\boldsymbol{\epsilon}^{(p,q)})$  and  $\Sigma_n = \sigma_e^{-2} \mathbf{X}^T \mathbf{X} + \sigma_a^{-2} \mathbf{I}_w$ .

### F. Update Moves

After the 3 stage mechanism for model order estimation has been executed, at each iteration several update moves are performed to update variance parameters and initial unobserved excitation vector  $\mathbf{e}^{(0)} = [e_1, e_2, \dots, e_q]$  via Gibbs sampling.

The full conditional distribution for  $\sigma_e^2$  is [19]:

$$f(\sigma_e^2 | \mathbf{x}, p, q, \mathbf{b}^{(p,q)}) = \mathcal{IG}(\sigma_e^2 | \alpha_{en}, \beta_{en}), \quad (23)$$

where  $\alpha_{en} = \alpha_e + \frac{1}{2}n$ ,  $\beta_{en} = \beta_e + \frac{1}{2}\mathbf{e}^T \mathbf{e}$ . Similarly, the full conditionals for coefficient vector variances appear as:

$$f(\sigma_a^2 | \mathbf{x}, p, k, \mathbf{a}^{(p,k)}) = \mathcal{IG}(\sigma_a^2 | \alpha_{an}, \beta_{an}), \quad (24)$$

$$f(\sigma_b^2 | \mathbf{x}, p, q, \mathbf{b}^{(p,q)}) = \mathcal{IG}(\sigma_b^2 | \alpha_{bn}, \beta_{bn}), \quad (25)$$

where parameters for these inverse Gamma functions are  $\alpha_{an} = \alpha_a + \frac{1}{2}w$ ,  $\beta_{an} = \beta_a + \frac{1}{2}(\mathbf{a}^{(p,k)})^T \mathbf{a}^{(p,k)}$ ,  $\alpha_{bn} = \alpha_b + \frac{1}{2}z$  and  $\beta_{bn} = \beta_b + \frac{1}{2}(\mathbf{b}^{(p,q)})^T \mathbf{b}^{(p,q)}$ .

Initial excitation vector,  $\mathbf{e}^{(0)}$  can be sampled from the distribution:

$$\mathbf{e}^{(0)} \sim \mathcal{N}(\mathbf{0}, \sigma_e^2 \mathbf{I}_q) \quad (26)$$

where  $\sigma_e^2$  is the sampled value from (23). Complete excitation vector  $\mathbf{e}$  has been estimated at each iteration via (1) starting from  $e_{q+1}$  up to  $e_n$  by using  $\mathbf{e}^{(0)}$ .

## V. NUMERICAL RESULTS

In this study, 10 PARMA models have been generated for simulations. Each model generates 20 different data sets with zero mean ( $\mu = 0$ ) where each has a length of 750 samples. Each data set is driven with a Gaussian excitation sequence with variance of  $\sigma_e^2$ .

In all the MCMC methods, selection of the initial values of the parameters does not have a direct affect on the final result. It only changes the convergence time to be long or short according to the selection [26], [27]. We have chosen the initial values by taking these issues into consideration. Hyperparameters  $\alpha_a, \alpha_b, \alpha_e, \beta_a, \beta_b$  and  $\beta_e$  are selected as 1. Each of the initial values  $p_0, k_0$  and  $q_0$  is also set to 1. Upper bounds,  $p_{\max}, k_{\max}$  and  $q_{\max}$ , for model orders have been selected as 5, 6 and 6, respectively. Model orders change with probability of 0.15 to up (birth) and down (death) and remain

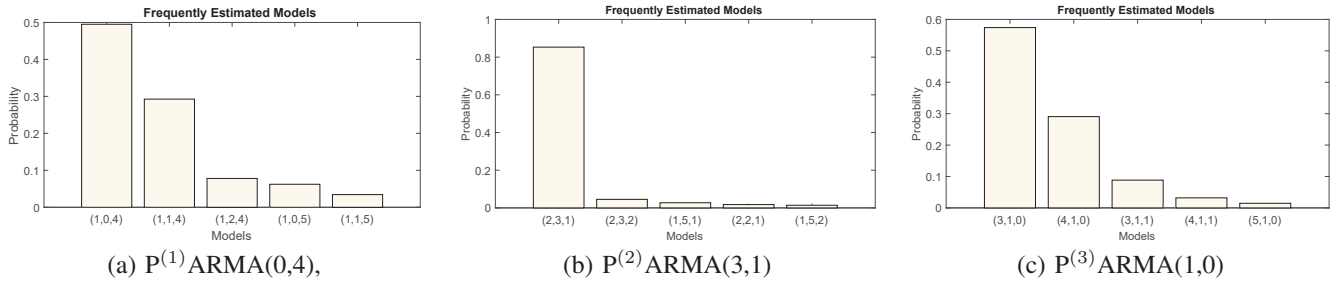
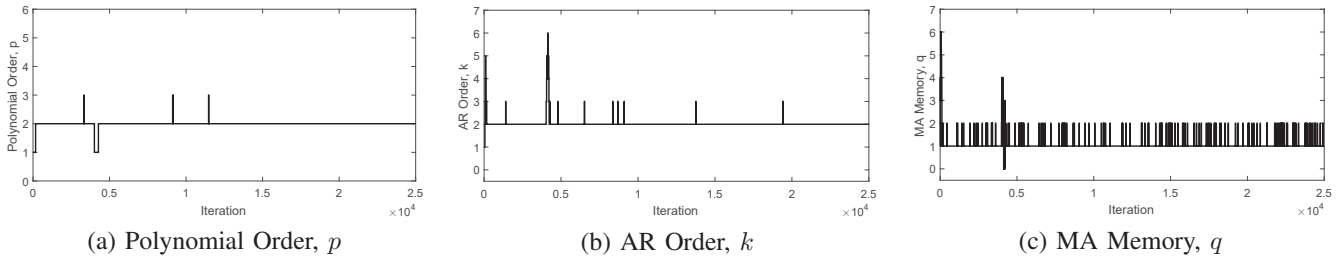
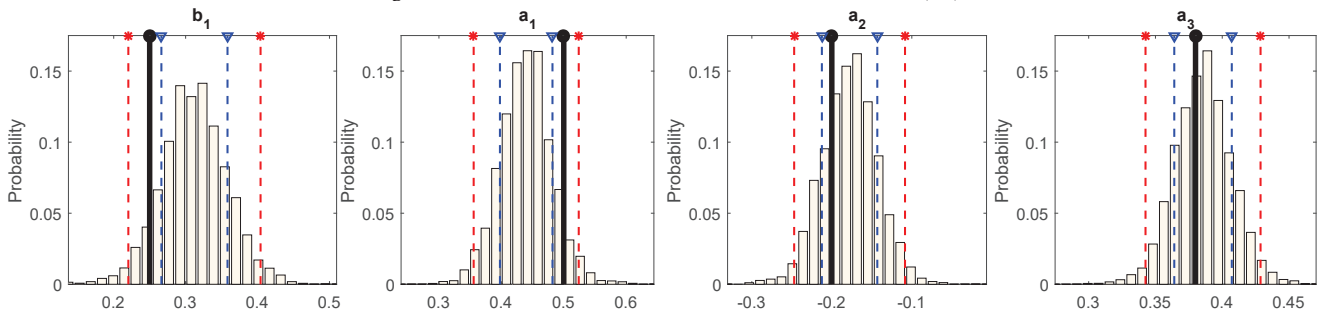


Fig. 1. Estimated joint posteriors for model orders in a single RJMCMC run.


 Fig. 2. The instantaneous model order estimates of  $P^{(2)}$ ARMA(2,1).

 Fig. 3. Estimated posteriors for  $P^{(1)}$ ARMA(3,1) Coefficients ( $\mathbf{b}^* = [0.25, 0.5, -0.2, 0.38]$ ). Black vertical line with "o" marker is the correct coefficient value, vertical dashed lines with "v" and "\*" markers refer to  $\pm\sigma$  and  $\pm 2\sigma$  CIs, respectively.

at the same (life) with a probability of 0.7. Each RJMCMC run performs 25000 iterations including a burn-in period of 10000 iterations.

In order to measure model coefficient estimation performance of the proposed method, the *normalized mean square error* (NMSE) which is defined as  $NMSE = \frac{1}{w+z} \sum_{i=1}^{w+z} \frac{(h_i - \hat{h}_i)^2}{\|\mathbf{h}\|_2^2}$  has been used. Here, vector  $\mathbf{h}$  includes both AR and MA model coefficients vectors  $\mathbf{a}^{(p,k)}$  and  $\mathbf{b}^{(p,q)}$  and  $\hat{\mathbf{h}}$  is its estimate.

Table I demonstrates results after 20 RJMCMC runs. These results reveal that for all 10 PARMA models, RJMCMC estimates correct model orders with the highest percentages. In addition, RJMCMC exhibits great performance on model coefficients estimation with NMSE values between  $6 \times 10^{-4}$  and  $1 \times 10^{-2}$  for all the models.

Figure 1 depicts the estimated joint posteriors of model order for 3 example PARMA models in a single RJMCMC run. Each sub-figure shows 5 frequently estimated models and their probabilities after burn-in period. Examining the figure, we can

 TABLE I  
 MODEL ESTIMATION RESULTS

|                      | (1,3,0) | (1,0,4) | (1,3,1) | (1,2,2) | (2,1,1) |
|----------------------|---------|---------|---------|---------|---------|
| % Perf. Of Detection | 85      | 65      | 80      | 80      | 95      |
| Avg. NMSE            | 0.0015  | 0.0022  | 0.0081  | 0.0145  | 0.0079  |
| # of Coeff.          | 3       | 4       | 4       | 4       | 4       |
|                      | (2,2,1) | (2,3,1) | (3,1,0) | (3,0,1) | (3,1,1) |
| % Perf. Of Detection | 90      | 100     | 70      | 80      | 80      |
| Avg. NMSE            | 0.0050  | 0.0056  | 0.0006  | 0.0011  | 0.0030  |
| # of Coeff.          | 7       | 11      | 3       | 3       | 6       |

clearly state that models which have highest probabilities are the correct models and these posteriors depict that RJMCMC estimates the correct models.

In Figure 2, instantaneous estimates for each model order  $p$ ,  $k$  and  $q$  for a  $P^{(2)}$ ARMA(2,1) model, have been shown for a single RJMCMC run with 25000 iterations. Burn-in period is also shown in Figure 2 in order to see the transient character of the algorithm.

Figure 3 shows the posterior probabilities for models coefficients of  $P^{(1)}$ ARMA(3,1) in a single RJMCMC run. Vertical black line corresponds to the correct model coefficient value for each sub-figure. Coefficient estimates for only correctly estimated model orders after burn-in period have been used to obtain the distributions in this figure. Resulting posteriors show that RJMCMC estimates model coefficients with a remarkable performance and the correct model coefficient values stand in 95% confidence interval (CI) ( $\pm 2\sigma$ ) for  $b_1$  and  $a_1$  and 68% CI ( $\pm\sigma$ ) for  $a_2$  and  $a_3$  of the estimated posteriors.

## VI. CONCLUSIONS

In this study we present a general framework for Volterra based linear-in-the-parameters nonlinear model estimation problems by extending our previous works on PAR and PMA models to PARMA models. We have utilized RJMCMC algorithm which has been generally used for linear model estimation problems, in the manner of performing transitions between linear and nonlinear model spaces.

We propose a method which can estimate all parameters of PARMA or its special cases PAR, PMA, linear AR, MA, ARMA processes. In addition, RJMCMC has been used to estimate the nonlinearity degree of the models which is generally assumed to be known in the literature.

Numerical results demonstrate promising performance of the proposed method in estimating model orders and corresponding model coefficients concurrently.

## ACKNOWLEDGEMENT

Oktay Karakuş is funded as a visiting scholar at ISTI-CNR, Pisa, Italy by The Scientific and Technological Research Council of Turkey (TUBITAK) under grant program 2214/A.

## REFERENCES

- [1] G. E. P. Box, G. M. Jenkins, and G. C. Reinsel, *Time series analysis: forecasting and control*. John Wiley & Sons, 2011, vol. 734.
- [2] S.-H. Peng, C.-W. Li, and Y.-W. Liu, "Joint estimation of vocal tract and nasal tract area functions from speech waveforms via auto-regression moving-average modeling and a pole assignment method," in *Acoustics, Speech and Signal Processing (ICASSP), 2015 IEEE International Conference on*. IEEE, 2015, pp. 4644–4648.
- [3] S. Ganapathy, "Robust speech processing using arma spectrogram models," in *Acoustics, Speech and Signal Processing (ICASSP), 2015 IEEE International Conference on*. IEEE, 2015, pp. 5029–5033.
- [4] F. Kozin, "Autoregressive moving average models of earthquake records," *Probabilistic Engineering Mechanics*, vol. 3, no. 2, pp. 58–63, 1988.
- [5] J. Zhong *et al.*, "Segmenting foreground objects from a dynamic textured background via a robust kalman filter," in *Computer Vision, 2003. Proceedings. Ninth IEEE International Conference on*. IEEE, 2003, pp. 44–50.
- [6] J. A. Cadzow, D. M. Wilkes, R. A. Peters, and X. Li, "Image texture synthesis-by-analysis using moving-average models," *IEEE transactions on aerospace and electronic systems*, vol. 29, no. 4, pp. 1110–1122, 1993.
- [7] W. Ji and K. Chee, "Prediction of hourly solar radiation using a novel hybrid model of ARMA and TDNN," *Solar Energy*, vol. 85, no. 5, pp. 808–817, 2011.
- [8] K. Benmouiza and A. Chekane, "Small-scale solar radiation forecasting using ARMA and nonlinear autoregressive neural network models," *Theoretical and Applied Climatology*, vol. 124, no. 3-4, pp. 945–958, 2016.
- [9] Y. Ohtsuka, T. Oga, and K. Kakamu, "Forecasting electricity demand in japan: A Bayesian spatial autoregressive ARMA approach," *Computational Statistics & Data Analysis*, vol. 54, no. 11, pp. 2721–2735, 2010.
- [10] S. Pappas, L. Ekonomou, P. Karampelas, D. Karamousantas, S. Katsikas, G. Chatzarakis, and P. Skafidas, "Electricity demand load forecasting of the Hellenic power system using an ARMA model," *Electric Power Systems Research*, vol. 80, no. 3, pp. 256–264, 2010.
- [11] J. Flores, M. Graff, and H. Rodriguez, "Evolutive design of ARMA and ANN models for time series forecasting," *Renewable Energy*, vol. 44, pp. 225–230, 2012.
- [12] E. Erdem and J. Shi, "ARMA based approaches for forecasting the tuple of wind speed and direction," *Applied Energy*, vol. 88, no. 4, pp. 1405–1414, 2011.
- [13] J. Connor, L. Atlas, and D. Martin, "Recurrent networks and NARMA modeling," in *NIPS*, 1991, pp. 301–308.
- [14] M. Lydia, S. Kumar, A. Selvakumar, and G. Kumar, "Linear and nonlinear autoregressive models for short-term wind speed forecasting," *Energy Conversion and Management*, vol. 112, pp. 115–124, 2016.
- [15] J. Zhang and H. Wang, "Minimum entropy control of nonlinear ARMA systems over a communication network," *Neural Computing and Applications*, vol. 17, no. 4, pp. 385–390, 2008.
- [16] S. Chen, K. Jeong, and W. Härdle, "Recurrent support vector regression for a non-linear ARMA model with applications to forecasting financial returns," *Computational Statistics*, vol. 30, no. 3, pp. 821–843, 2015.
- [17] P. J. Green, "Reversible jump Markov chain Monte Carlo computation and Bayesian model determination," *Biometrika*, vol. 82, no. 4, pp. 711–732, 1995.
- [18] 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.
- [19] P. T. Troughton and S. J. Godsill, "A reversible jump sampler for autoregressive time series," in *ICASSP 1998 IEEE*, vol. 4. IEEE, 1998, pp. 2257–2260.
- [20] R. S. Ehlers and S. P. Brooks, "Bayesian analysis of order uncertainty in ARIMA models," *Federal University of Parana, Tech. Rep.*, 2004.
- [21] E. Eğri, S. Günay *et al.*, "Bayesian model selection in ARFIMA models," *Expert Systems with Applications*, vol. 37, no. 12, pp. 8359–8364, 2010.
- [22] O. Karakuş, E. Kuruoğlu, and M. Altinkaya, "Estimation of the nonlinearity degree for polynomial autoregressive processes with rjmc," in *Signal Processing Conference (EUSIPCO), 2015 23rd European*. IEEE, 2015, pp. 953–957.
- [23] —, "Bayesian estimation of polynomial moving average models with unknown degree of nonlinearity," in *Signal Processing Conference (EUSIPCO), 2016 24th European*. IEEE, 2016.
- [24] M. A. Ivanov, "Distribution of the output response of Volterra-type narrowband nonlinear-systems," *Telecommunications and Radio Engineering*, vol. 41, no. 12, pp. 99–102, 1987.
- [25] P. T. Troughton and S. J. Godsill, "MCMC methods for restoration of nonlinearly distorted autoregressive signals," *Signal Processing*, vol. 81, no. 1, pp. 83–97, 2001.
- [26] A. Gelman and D. B. Rubin, "Inference from iterative simulation using multiple sequences," *Statistical science*, pp. 457–472, 1992.
- [27] A. E. Raftery and S. M. Lewis, "[Practical Markov Chain Monte Carlo]: comment: one long run with diagnostics: implementation strategies for Markov Chain Monte Carlo," *Statistical science*, vol. 7, no. 4, pp. 493–497, 1992.