System Identification on the Families of Auto-Regressive with Least-Square-Batch Algorithm

Moh Kamalul Wafi[1]

Abstract—The theories of system identification have been highly elaborated so as to achieve the true system. This paper much discuses regarding the stochastic processes along with the divergent of whether or not the system has zero-mean under scenario of either white and coloured noise. The mathematical foundations, including mean, variance, covariance, optimal parameters, along with some modified scenarios among them, are presented in detail from various system along with some basic idea behind them. The families of auto-regressive (1) and (2) are compared both mathematical and simulation in order to obtain the best design approaching the true system, Moreover, the least-square algorithm is used to examine the effectiveness of some number of iteration along with "batch" theorem.

Index Terms—System Identification, Auto-Regressive, Least-Square Algorithm, Batch Theorem

I. INTRODUCTION

System identification, an error-correcting training scheme, is becoming more important dealing with either a system or process requiring to estimate the states with certain unknown variables. Further research even has modified from classic statistical method to regularisation considering the unhealthy matrix of minimization scheme [1], [2]. The idea of auto-regressive involves current and preceding data to be estimated with mutually independent per-sequence [3]. This model is then modified so as to obtain the best prediction, reducing the error. This is well-conducted by [4] proposing unknown number of epoch formula each sequence inducing the Markov process. Thanks to [5], the very early idea in predicting one-step ahead using either the median and the log-normal process of distribution. Moreover, the procedural modification applying Lasso scenario is also a choice as presented in [6]. This research is to examine basic properties from the family of auto-regressive along with the benefit of some iteration N and batch κ. Beyond that, the assumptions of zero-mean and its counterpart along with either white and coloured noise are also considered.

The first and foremost is the stability properties of the system. The Lyapunov scenario as the most famed algorithm is proposed in this project [7]. While there has been a broad modification in determining the stability of the system, including strong passivity in nonlinear system as stated in [8], the mathematical foundation should be well-defined [9]. These basic information are then ready followed by some recursive or forecasting scheme with other bunch of methods. Least-square as prediction is to portray the best linear to show the data along

[1]Moh Kamalul Wafi is with Laboratory of Embedded Systems and Control, Department of Engineering Physics, Institut Teknologi Sepuluh Nopember, 60111, Indonesia, `kamalul.wafi at gmail.com`

with forecast the following data with respect to some instant time t, consisting the trend and the level of development. The more generalized with coloured noise system is presented in [10] referring to the original [11]. This paper comprises the problem formulation followed by the basic mathematical concepts. The next is the analytical result with three divergent scenarios ended by the numerical design and conclusion.

II. PROBLEM FORMULATION

This paper presents the objectives of system identification:

1) To elaborate the functions of expected value, variance, and covariance in a stationary process via static representation
2) To describe mathematically the identification abilities of AR models of increasing order
3) To compare the theoretical results with its simulation with empirical Least Square (LS) estimation
4) To highlight the impact of coloured noises on the LS estimation

III. BASIC MATHEMATICAL THEORIES

General design of the system model is presented in Fig. (1) defining that P is the system plant along with ξ as its parameter. ν comprises either fault (f) or disturbances (d) whilst u and z denote as the accessible input and the controlled variable in turn. The space of possible results is denoted by $\mathcal{S}$ and its single result of s is $\in \mathcal{S}$ while the events of certain interest are defined as s_ϵ which is also the subset of $\mathcal{S}$. Furthermore, the patterns of the output depend on the parameters on algebraic equation of the system P_ξ resulting its roots either real or complex. The common features of the dynamics constitute that the (t) continuous or (z) discrete-mathematical structure is built by differential equation and it has been a-priori knowledge being found from the physical properties of the system. The applied random variable is a
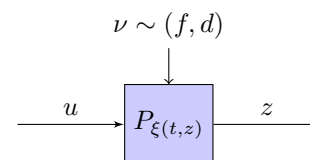

Figure 1: The general dynamic systems

variable of $\mathcal{F}(s)$ obtaining values according to the result $s \in \mathcal{S}$ of a random trial through a function $\Theta(\bullet)$. The probability distribution function supplies the data on the random variable so that the equation of $F(x) = \mathcal{P}(v \le x)$ can be denoted as,

$$\mathcal{P}(v \in [a,b]) = F(b) - F(a) \tag{1}$$

meaning that $\mathcal{P}$ is the specific probability distribution being made of function F with points between a and b while the probability density of $f^\dagger$ creating an area of $\mathcal{P}(v \in [a,b])$ is defined as follows,

$$f^\dagger(x) = \frac{\partial F}{\partial x} \tag{2}$$

Furthermore, the expected value $\mathbf{E}$ and its variance $\mathcal{Q}$ of a fault and disturbance are presented as,

$$\mathbf{E}(v) = \int_{-\infty}^{+\infty} x f^\dagger(x)\,dx \tag{3}$$

$$\mathcal{Q}_v = \int_{-\infty}^{+\infty} [x - \mathbf{E}(v)]^2 f^\dagger(x)\,dx \tag{4}$$

with standard deviation of $\sigma(v) = \sqrt{\mathcal{Q}_v}$. Keep in mind that, given two variables of $v_1(s)$ and $v_2(s)$, the results of $v(s)$, $\mathbf{E}(v)$, and $\mathcal{Q}_v$ are $v(s) = v_1(s) + v_2(s)$, $\mathbf{E}(v) = \mathbf{E}(v_1) + \mathbf{E}(v_2)$, and

$$\mathcal{Q}(v) \neq \mathcal{Q}(v_1) + \mathcal{Q}(v_2) \tag{5}$$

in turn, with $\mathcal{Q}_v = \mathbf{E}\left\{[v - \mathbf{E}(v)][v - \mathbf{E}(v)]^T\right\}$. A Gaussian random variable exists if the average $\mu = \mathbf{E}(v)$ and variance $\mathcal{Q}$ or σ^2 satisfies:

$$g(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-k} \longrightarrow k = \frac{(x-\mu)^2}{2\sigma^2} \tag{6}$$

with respect to the properties of the v in terms of correlation and independent, again, the two random variables v_1 and v_2 are **uncorrelated** if

$$\mathbf{E}\left\{[v_1 - \mathbf{E}(v_1)][v_2 - \mathbf{E}(v_2)]^\top\right\} = 0 \tag{7}$$

whereas they are called as independent if

$$\mathcal{P}(v_1, v_2) = \mathcal{P}(v_1)\mathcal{P}(v_2) \tag{8}$$

hence, the stochastic process originally requires the portrayal of probability distribution and is a collection of infinite random variable sequenced with respect to time. A stochastic process of $v(t)$ is called **white** being defined as $v \sim \mathcal{WN}(0, \lambda^2)$ if $\mathbf{E}[v(t)] = 0$ and

$$\mathcal{Q}(x) = \begin{cases} \lambda^2, & x = 0 \\ 0, & x \neq 0 \end{cases} \tag{9}$$

Beyond that, apart from examine the unbiased estimation, the minimization of the variance Φ could be tested. The estimator of $\bar{\theta}^1$ is better than that of $\bar{\theta}^2$ if $\Phi_{\bar{\theta}1} \leq \Phi_{\bar{\theta}2}$ or, if $\bar{\theta}^i$ is a vector, then $\Phi_{\bar{\theta}2} - \Phi_{\bar{\theta}1} \geq 0$. The asymptotic characteristic is affected by the number of simulated data N which also influence the better estimation and reduce the uncertainty. Suppose three different variances Φ_m with the same expected value $\mathbf{E}$ such that $\mathbf{E}[\bar{\theta}^i] = \theta$ with $i = 1 \rightarrow m$ and $m_3 > m_2 > m_1$, the estimate $\bar{\theta}^m$ is said good at converging to θ if

$$\lim_{m \to \infty} \Phi_{\bar{\theta}m} = 0 \longrightarrow \lim_{m \to \infty} \mathbf{E}\left[\|\bar{\theta}^m - \theta\|^2\right] = 0$$

where $\bar{\theta}^m, \theta$ and $\|\bar{\theta}^m - \theta\|$ are the random and constant vector along with a scalar random with good mean variable in turn.

IV. ANALYTICAL RESULTS

This section acts as the numerical findings which are then compared to the simulations with Least Square algorithm being elaborated in the following section. There are three different scenarios. First is the basic with zero-mean and the second is to alter from zero-mean to non-zero mean yet the other conditions are kept the same and the following is to change the covariance instead of the mean.

A. Independent and Zero-mean

The dynamics of a stationary stochastic process $y(\bullet)$ is presented in the following Fig. (2) in which $q(\bullet)$ and $v(\bullet)$ are independent, such that:
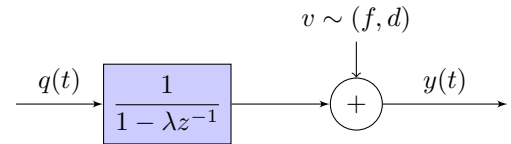
Figure 2: Dynamic systems of the proposed model

where $q \sim \mathcal{WN}\left(\bar{q}, \delta^2\right)$, $v \sim \mathcal{WN}\left(\bar{v}, \xi^2\right)$, with $[\bar{q}, \bar{v}] = 0$ and $\lambda \in \mathbb{R}, |\lambda| < 1$. It can clearly be shown from the Fig. (2) that there have been two variables which influence the point of output $y(t)$ so that, by applying some algebra shown in **Appendix A** from inversing the $z-$transform to (*time*) $t-$domain, it follows that:

$$y(t) = \frac{1}{1 - \lambda z^{-1}} q(t) + v(t) \tag{10}$$

can be then written as the true system of the following formula, therefore

$$\mathcal{S} : y(t) = \lambda y(t-1) + q(t) + v(t) - \lambda v(t-1) \tag{11}$$

As the inputs process of $q(\bullet)$ and $v(\bullet)$ are stationary and the point of λ lies strictly inside the unit circle as well as $|\lambda| < 0$ which makes the value of $|z| < 0$, the dynamic model generating the stochastic process is asymptotically stable which implies that the steady state stochastic process $y(t)$ is stationary. Furthermore, since the zero-mean, the expected value $\mathbf{E}[y(t)] = \bar{y}$ computed in **Appendix B** is zero,

$$\mathbf{E}[y(t)] = \bar{y} = 0 \longrightarrow [\bar{q}, \bar{v}] = 0 \tag{12}$$

or substituting the two mean of the independent $[q, v]$ by the end of the solution, the formula is then finished as $\bar{y} = \lambda \bar{y}$, concluding that $\bar{y} \to 0$ and $\lambda \neq 0$. As for the variance of $y(t)$ denoted as Φ_y, it is solved in the **Appendix C**, hence

$$\Phi_y = \frac{\delta^2 + (1 - \lambda^2)\xi^2}{1 - \lambda^2} \tag{13}$$

Since the expected value $(\bar{y})$ is zero, the covariance function Ψ_y with $\tau = 0, 1, \ldots$ is exactly the same as the correlation function being calculated in **Appendix D** and **D***, with

$$\Psi_y(\tau) = \mathbf{E}\left\{[y(t) - \bar{y}][y(t-\tau) - \bar{y}]\right\} \tag{14}$$

such that,

$$\Psi_y(\tau) = \begin{cases} \dfrac{\delta^2}{1-\lambda^2} + \xi^2 & , \tau = 0 \\[3mm] \dfrac{\delta^2}{1-\lambda^2}\lambda^\tau & , \tau > 0 \end{cases} \tag{15}$$

Furthermore, the analysis of the proposed dynamical process $y(\bullet)$ is required using a PEM identification algorithm of the family AR(1), such that

$$\mathscr{F}_1(\theta_1) : y(t) = \phi_1 y(t-1) + \psi(t) \tag{16}$$

and the corresponding stage of the design $\mathscr{F}_1$ in estimation form is,

$$\hat{\mathscr{F}}_1(\theta_1) : \hat{y}(t|t-1) = \phi_1 y(t-1) \tag{17}$$

As for the higher-order of the family AR(2), the $\mathscr{F}_2$ is denoted as follows,

$$\mathscr{F}_2(\theta_2) : y(t) = \phi_1 y(t-1) + \phi_2 y(t-2) + \psi(t) \tag{18}$$

along with its prediction form, such that

$$\hat{\mathscr{F}}_2(\theta_2) : \hat{y}(t|t-1) = \phi_1 y(t-1) + \phi_2 y(t-2) \tag{19}$$

where θ_1 and θ_2 equal to ϕ_1 and $[\phi_1, \phi_2]^\top$ in turn along with suitable white noise of $\psi(t)$. Keep in mind that to compute the optimal values of those two $(\theta_1^\star, \theta_2^\star)$ with respect to (θ_1, θ_2) respectively is to use the PEM algorithm and according to the asymptotic theory, the convergence to one of the minima of the two functions $\Gamma(\theta_1)$ and $\Gamma(\theta_2)$ can be guaranteed, such that

$$\Gamma(\theta_1) = \mathbf{E}\left\{[\varepsilon_{\theta_1}(t)]^2\right\} = \mathbf{E}\left\{[y(t) - \hat{y}(t|t-1)]^2\right\} \tag{20}$$

$$\Gamma(\theta_2) = \mathbf{E}\left\{[\varepsilon_{\theta_2}(t)]^2\right\} = \mathbf{E}\left\{[y(t) - \hat{y}(t|t-1)]^2\right\} \tag{21}$$

where $\varepsilon_{\theta_1}(t)$ is the prediction error with suitable $y(t)$ as shown in the $\mathscr{F}_1$ and $\mathscr{F}_2$ and regarding the cost (optimal) functions of the two, they are defined as follows,

$$\frac{\partial \Gamma(\theta_1)}{\partial \phi_1} = 0 \quad \text{and} \quad \frac{\partial \Gamma(\theta_2)}{\partial \phi_1 \partial \phi_2} = 0 \tag{22}$$

Furthermore, applying the concepts of [] above results in the optimality and recalling the $\Psi_y(\tau)$, the values of $(\theta_1^\star, \theta_2^\star)$ are written in the following according to the calculation being done in **Appendix E** and **F**,

$$\theta_1^\star := \phi_1 = \frac{\lambda \delta^2}{\delta^2 + (1-\lambda^2)\xi^2} \tag{23}$$

$$\theta_2^\star := [\phi_1, \phi_2]^\top = \begin{cases} \phi_1 = \dfrac{\lambda \delta^2 (\delta^2 + \xi^2)}{(\delta^2 + \xi^2)^2 - \lambda^2 \xi^4} \\[4mm] \phi_2 = \dfrac{\lambda^2 \delta^2 \xi^2}{(\delta^2 + \xi^2)^2 - \lambda^2 \xi^4} \end{cases} \tag{24}$$

Those optimum values then are implemented in the models of AR(1) and AR(2) as the best design of either families, $\mathscr{F}_1(\theta_1^\star)$ with $\theta_1^\star := \phi_1$ and $\mathscr{F}_2(\theta_2^\star)$ with $\theta_2^\star := [\phi_1, \phi_2]^\top$, approximating

the true system. The variance of the prediction errors related to the two designs can be calculated from [] due to the fact that the expected values of the error from either families are zero, such that,

$$\bar{\varepsilon}_{\theta_1} = \mathbf{E}[y(t)] - \phi_1 \mathbf{E}[y(t-1)] = 0 \tag{25}$$

$$\bar{\varepsilon}_{\theta_2} = \mathbf{E}[y(t)] - \phi_1 \mathbf{E}[y(t-1)] - \phi_2 \mathbf{E}[y(t-2)] = 0 \tag{26}$$

and the variance of the families are enlightened in **Appendix G** by substituting the covariance, therefore

$$\Phi_{\varepsilon_{\theta_1^\star}} = \frac{(\delta^2 + \xi^2)^2 - \lambda^2 \xi^4}{\delta^2 + (1-\lambda^2)\xi^2} \tag{27}$$

$$\Phi_{\varepsilon_{\theta_2^\star}} = \frac{(\delta^2 + \xi^2)^3 - \lambda^2 (\delta^2 \xi + \xi^3)^2 + 2\lambda \delta^4 \xi^2}{1-\lambda^2} \tag{28}$$

B. Independent and Non-Zero Mean

Recalling the same dynamical process as Fig. (2), the q and v are no longer zero-mean, constituting $q \sim \mathcal{WN}(\bar{q}, \delta^2)$ and $v \sim \mathcal{WN}(\bar{v}, \xi^2)$ with $\bar{q}$ and $\bar{v}$ equal to 1 and 4 respectively. Keep in mind that when $q(\bullet)$ and $v(\bullet)$ are zero-mean, the process $y(\bullet)$ is also zero-mean and by contrast, if the two are not, this makes the expected value of the process is not zero anymore, recalling **Appendix B**, the $\bar{y}$ is,

$$\bar{y} := \mathbf{E}[y(t)] = \lambda \bar{y} + \bar{q} + \bar{v} - \lambda \bar{v}$$
$$= \bar{v} + \frac{\bar{q}}{1-\lambda} \longrightarrow \frac{5 - 4\lambda}{1-\lambda} \tag{29}$$

and this also leads to the introduction of the new process $\check{y}(t)$, where $\check{y}(t) = y(t) - \bar{y}$, therefore

$$y(t) = \check{y}(t) + \bar{y} \tag{30}$$

such that,

$$\Phi_y = \mathbf{E}\left\{[y(t) - \bar{y}]^2\right\} = \mathbf{E}\left\{[\check{y}(t)]^2\right\} \tag{31}$$

However, the not zero-mean process of $y(t)$ does not yield the changes of the variance because the mean is subtracted from it. Likewise, due to the coincidence between the correlation function in the zero mean process $\check{y}(t)$ and the covariance of the original process, this new process does not give divergent results in terms of covariance, meaning that both parameters are on par with the originals, therefore

$$\Psi_y(\tau) = \mathbf{E}\left\{[y(t) - \bar{y}][y(t-\tau) - \bar{y}]\right\}$$
$$= \mathbf{E}\left\{[\check{y}(t)][\check{y}(t-\tau)]\right\} \tag{32}$$

Furthermore, the AR families analysis of $\mathscr{F}_1$ and $\mathscr{F}_2$ regarding the optimality $(\theta_1^\star, \theta_2^\star)$ which are affected by the covariance as mentioned in Eqs. (21) and (22) result in the similarities of the original processes with slight differences. Recalling the new process of $\check{y}(t)$, the cost function $\Gamma(\theta_1)$, $\Gamma(\theta_2)$ along with their estimates,

$$\hat{y}(t|t-1)_{\hat{\mathscr{F}}_1} = \phi_1 y(t-1)$$
$$= \phi_1 [\check{y}(t-1) + \bar{y}] \tag{33}$$

$$\hat{y}(t|t-1)_{\hat{\mathscr{F}}_2} = \phi_1 y(t-1) + \phi_2 y(t-2)$$
$$= \phi_1 \check{y}(t-1) + \phi_1 \check{y}(t-2) + (\phi_1 + \phi_2)\bar{y} \tag{34}$$

respectively. The optimal value of $(\theta_1^\star, \theta_2^\star)$ is then obtained in **Appendix H** and **I**, such that

$$\theta_1^\star := \phi_1 = \frac{\lambda\delta^2 + (1-\lambda^2)\,\bar{y}^2}{\delta^2 + (1-\lambda^2)(\xi^2 + \bar{y}^2)} \qquad (35)$$

and $\theta_2^\star := [\phi_1, \phi_2]^\top$, with

$$\begin{bmatrix} \dfrac{\left[\lambda\delta^2 + (1-\lambda^2)\,\bar{y}^2\right](\delta^2+\xi^2)}{(\delta^2+\xi^2+2\bar{y}^2)(\delta^2+\xi^2) - \xi^2\lambda^2(\xi^2+2\bar{y}^2) - 2\lambda\delta^2\bar{y}^2} \\[2em] \dfrac{\lambda^2\delta^2\xi^2 + \left[\delta^2(1-\lambda)^2 + \xi^2(1-\lambda^2)\right]\bar{y}^2}{(\delta^2+\xi^2+2\bar{y}^2)(\delta^2+\xi^2) - \xi^2\lambda^2(\xi^2+2\bar{y}^2) - 2\lambda\delta^2\bar{y}^2} \end{bmatrix} \qquad (36)$$

Likewise, these result are applied as the most optimum designs in either families approaching the true systems, while the variance of them are modified using these following mean of prediction error, such that

$$\bar{\varepsilon}_{\theta_1} = (1-\phi_1)\bar{y} \qquad (37)$$
$$\bar{\varepsilon}_{\theta_2} = (1-\phi_1-\phi_2)\bar{y} \qquad (38)$$

and the results are showed in **Appendix J** with small divergent compared to the cost functions of $(\Gamma(\theta_1), \Gamma(\theta_2))$ by the square of the expected values. Finally, under the initial assumption of the characteristics $q(\bullet)$ and $v(\bullet)$, these influence almost the whole systems leading to the shifted distribution.

C. Assumption to Parameters of λ, δ^2, ξ^2

Turning to another design with similar dynamical process as Fig. (2), the q and v are set as zero-mean with certain true parameters, comprising $q \sim \mathcal{WN}(0,4)$ and $v \sim \mathcal{WN}(0,9)$. This design will not change the expected value with $\bar{y} = 0$ and the result makes the variance and covariance are counted the same as the originals (see **Appendix C** and **D**) with,

$$\Phi_y = 13.5 \qquad (39)$$

and,

$$\Psi_y(\tau) = \begin{cases} 13.5 & , \tau = 0 \\[1em] 4.5\left(\dfrac{1}{3}\right)^\tau & , \tau > 0 \end{cases} \qquad (40)$$

Furthermore, since the model of AR(1) is affected by $\Psi_y(\tau)$, the optimal value of $\theta_1^\star$ and the best design of $\mathcal{F}_1$ are

$$\theta_1^\star := \phi_1 = \frac{1}{9} \longrightarrow \mathcal{F}_1 : y(t) = \frac{1}{9}y(t-1) + \phi(t) \qquad (41)$$

which is approaching the true system with variance $\Phi_{\theta_1^\star}$ equals to 13.333. Likewise, as for the counterpart family of AR(2), the optimal value of $\theta_2^\star$ is

$$\theta_2^\star := [\phi_1, \phi_2]^\top = \left[\frac{39}{360}, \frac{1}{40}\right]^\top \qquad (42)$$

so that the best model of AR(2) can be written as follows

$$\mathcal{F}_2 : y(t) = \frac{39}{360}y(t-1) + \frac{1}{40}y(t-2) + \phi(t) \qquad (43)$$

with the variance $\Phi_{\theta_2^\star}$ equals to 13.325. From the model, the AR(2) is slightly better in terms of small variance with 0.008 difference compared to its counterpart.

V. NUMERICAL DESIGN

Linear regression as typical context related to least-square estimator is initialized with $n+1$ variables of $y(t)$, $\theta_1(t), \ldots, \theta_n(t)$ over an interval of $t = 1, 2, \ldots, N$. It is then required, if possible, n variables $\zeta_1, \zeta_2, \ldots, \zeta_n$, such that

$$y(t) = \zeta_1\theta_1 + \zeta_2\theta_2 + \cdots + \zeta_n\theta_n \qquad (44)$$

is denoted as linear regression according to $\theta(t)$ variable which is able to rewrite as a matrix function of $\zeta^\top\theta$ acting as row matrices. Recalling equation (11) with certain iteration of N along with some random sequence of either independent, the recursive equation is set as,

$$\Pi = \begin{bmatrix} y(1) & y(2) & \cdots & y(N) \end{bmatrix} \qquad (45)$$

While the appearance of error ϵ referring to $y(t) - \zeta^\top\theta$ is undeniable, to minimize it is a key by obtaining the optimal vector of $\theta^\star$. It can be achieved by the function of quadratic cost, therefore

$$\Gamma(\theta) = \sum_{t=1}^{N}\left[y(t) - \zeta^\top\theta\right]^2 \longrightarrow \theta^\star = \arg\min_\theta \Gamma(\theta) \qquad (46)$$

such that after differentiating the cost function in terms of θ-element generating the zero results, the θ could be determined by converting the equality of both row and column vectors. If sum of the product of $\zeta(t)\zeta(t)^\top$ from $1 \to N$ is non-singular,

$$\theta = \left[\sum_{t=1}^{N}\zeta(t)\zeta(t)^\top\right]^{-1}\sum_{t=1}^{N}\zeta(t)y(t)^\top \qquad (47)$$

it can be deduced that for the minimum of Γ, $\|y_N - \zeta_N\theta\|$ should be orthogonal to $\zeta_N\theta$. This is due to the symmetric and positive semi-definite so that θ_N yields in local minimum of $\Gamma(\theta)$. Moreover, the sum of $\zeta(t)\zeta(t)^\top$ results in two divergent scenarios of determinant whether it is zero or not. Unique global minimum or identifiability is situated if it is not zero while for the counterpart result, θ_N is one among the infinite global minima. The least-square algorithm in Eq. (47) is then used to computed the optimal $[\theta_1^\star, \theta_2^\star]$ of either families. The variation of N is also applied to observe the effectiveness of the recursive. Furthermore, the κ independent batches of the generated Π_i, such that

$$\Pi_i = \begin{bmatrix} y_i(1) & y_i(2) & \ldots & y_i(N) \end{bmatrix} \quad i = 1 \to \kappa \qquad (48)$$

with certain N is required to calculate the empirical mean for both families, such that

$$\bar{\theta}_1 = \frac{1}{\kappa}\sum_{i=1}^{\kappa}\hat{\theta}_1^{(i)} \quad \text{and} \quad \bar{\theta}_2 = \frac{1}{\kappa}\sum_{i=1}^{\kappa}\hat{\theta}_2^{(i)} \qquad (49)$$

along with some variances as follows. The broader batches of $\check{\Pi}_i$ with certain constant α, therefore αN is designed.

$$\sigma_{\theta_1}^2 := \Phi_{\theta_1} = \frac{1}{\kappa}\sum_{i=1}^{\kappa}\left(\hat{\theta}_1^{(i)} - \bar{\theta}_1\right)^2 \qquad (50)$$
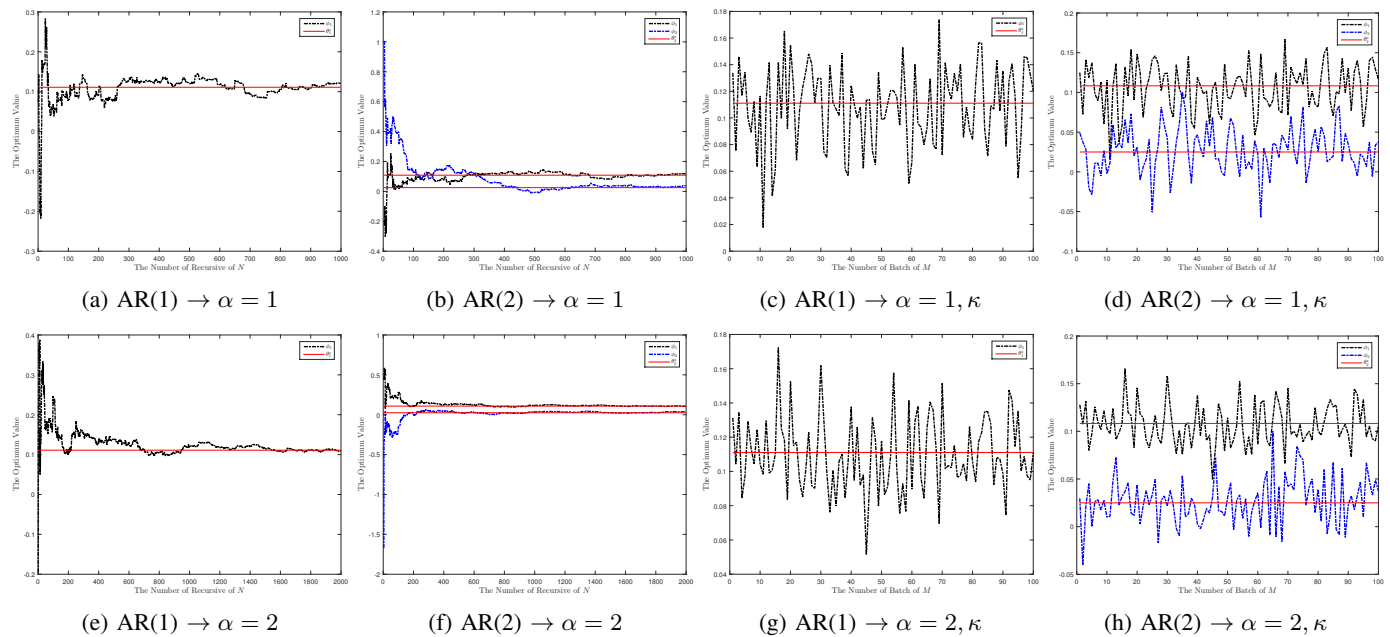
(a) AR(1) → α = 1 (b) AR(2) → α = 1 (c) AR(1) → α = 1, κ (d) AR(2) → α = 1, κ

(e) AR(1) → α = 2 (f) AR(2) → α = 2 (g) AR(1) → α = 2, κ (h) AR(2) → α = 2, κ

Figure 3: These show the different performance of either families with parameters of α and κ

$$\sigma^2_{\theta_2} := \Phi_{\theta_2} = \frac{1}{\kappa} \sum_{i=1}^{\kappa} \begin{bmatrix} (\varphi_1)^2 & \varphi_1\varphi_2 \\ \varphi_1\varphi_2 & (\varphi_2)^2 \end{bmatrix} \quad (51)$$

with,

$$\varphi_1 = \hat{\theta}^{(i)}_{2,1} - \bar{\theta}_{2,1} \qquad \varphi_2 = \hat{\theta}^{(i)}_{2,2} - \bar{\theta}_{2,2} \quad (52)$$

The process $q(\bullet)$ is supposed to be *coloured noise* based on the following recursive equation,

$$q(t) = -\frac{1}{2}q(t-1) + \eta(t) \quad (53)$$

where $\eta \sim \mathcal{WN}(0,1)$ and $\eta(\bullet)$ is independent from $v(\bullet)$. This scenario is then observed being compared to that of *white noise* in the preceding. It can be concluded that the variance Φ_{θ_2} is slightly smaller than that of Φ_{θ_1} meaning that the model approximation to the true system of AR(2) is better than that of AR(1). In terms of the simulation, the yields are closer to the hand-made calculation relative to certain $\alpha N, \alpha = 1 \to \infty$, the greater α results in better approximation of optimal value leading to the variance. With $\alpha = [1,2]$ and $N = 1000$ in AR(1) for instance, it generates

$$0.14 \to N, \quad 0.12 \to 2N \quad (54)$$

optimal value $\theta_1^\star$ respectively compared to the hand calculation 0.11 while as for AR(2) as $\theta_2^\star$ with the same design of α, it yields $[\phi_1, \phi_2]^\top$ as

$$\begin{bmatrix} 0.128 \\ 0.031 \end{bmatrix} \to N, \quad \begin{bmatrix} 0.123 \\ 0.028 \end{bmatrix} \to 2N \quad (55)$$

in turn against the origin of $[0.108, 0.025]^\top$. For certain $\kappa = 100, N = 1000$ and $\alpha = [1,2]$, the $\theta_1^\star$ AR(1) generates,

$$\begin{bmatrix} \theta_1^\star \\ \Phi_{\theta_1} \end{bmatrix} := \begin{bmatrix} 0.116 \\ 0.0011 \end{bmatrix} \to N, \quad \begin{bmatrix} 0.112 \\ 0.0005 \end{bmatrix} \to 2N \quad (56)$$

and these values of the estimated parameter is close to that of the optimal $\theta_1^\star = 0.11$. This is due to the more data provided leading to better approximation. With respect to AR(2), the optimal parameter $\theta_2^\star$ and the variance Φ_{θ_2} with the same scenarios of κ, N, α are obtained as follows,

$$\theta_2^\star = \begin{bmatrix} 0.1048 \\ 0.0216 \end{bmatrix} \quad \Phi_{\theta_2} = \begin{bmatrix} 0.00097 & -0.00006 \\ -0.00006 & 0.00111 \end{bmatrix} \quad (57)$$

for $N = 1000$ whereas with $\alpha = 2$, it gains

$$\theta_2^\star = \begin{bmatrix} 0.1086 \\ 0.0245 \end{bmatrix} \quad \Phi_{\theta_2} = \begin{bmatrix} 0.00046 & -0.00002 \\ -0.00002 & 0.00059 \end{bmatrix} \quad (58)$$

compared to the origin of $[\phi_1, \phi_2]^\top = [0.1083, 0.0250]^\top$. Moving to *coloured noise* which generates divergent approximation, for AR(1) with the same design as the *white noise* and hand-mad calculation, the optimal value $\theta_1^\star$ comprises,

$$-0.0137 \to N, \quad -0.0259 \to 2N \quad (59)$$

This not surprising though, as for each sequence iteration, it yields a divergent stimulated data influencing the system so that it is far compared to that of the *white*. As for AR(2), $\theta_2^\star$ is reached as follows,

$$\begin{bmatrix} -0.0686 \\ -0.0083 \end{bmatrix} \to N, \quad \begin{bmatrix} 0.0014 \\ 0.0538 \end{bmatrix} \to 2N \quad (60)$$

which show the same pattern as in AR(1). With respect to batches $\Pi_i \to \kappa$ scenario, the optimal value of AR(1) constitutes beyond the origin meaning that it is biased affecting the empirical and its variances, such that

$$\begin{bmatrix} \theta_1^\star \\ \Phi_{\theta_1} \end{bmatrix} := \begin{bmatrix} -0.025 \\ 0.0015 \end{bmatrix} \to N, \quad \begin{bmatrix} -0.022 \\ 0.0005 \end{bmatrix} \to 2N \quad (61)$$

and AR(2) is also performing the same another convergence number apart from the origin (hand-made calculation), it is proven as in the following,

$$\theta_2^\star = \begin{bmatrix} 0.1048 \\ 0.0216 \end{bmatrix} \quad \Phi_{\theta_2} = \begin{bmatrix} 0.00097 & -0.00006 \\ -0.00006 & 0.00111 \end{bmatrix} \quad (62)$$

for $N = 1000$ whereas with $\alpha = 2$, it gains

$$\theta_2^\star = \begin{bmatrix} 0.1086 \\ 0.0245 \end{bmatrix} \quad \Phi_{\theta_2} = \begin{bmatrix} 0.00046 & -0.00002 \\ -0.00002 & 0.00059 \end{bmatrix} \quad (63)$$

VI. Conclusion

The mathematical scenarios of switching from zero-mean to non zero-mean along with from white to coloured have been studied to give better picture of the mathematical characteristic from either families. The simulation designs have been proposed in order to observe the effectiveness of certain number of recursive N along with the batches concept κ compared to the hand calculation of the optimal values. The greater number of N and batches κ with the more complex design of estimate AR(2) results in better approximation of the true system in terms of empirical mean and variance. Further research is to compare another more complex families with modified stochastic scenarios.

Acknowledgment

References

[1] L. Ljung, T. Chen, and B. Mu, "A shift in paradigm for system identification," *International Journal of Control*, pp. 1–8, 2019.

[2] J. Nagumo and A. Noda, "A learning method for system identification," *IEEE Transactions on Automatic Control*, vol. 12, no. 3, pp. 282–287, June 1967.

[3] E. G. Hurst, "Bayesian autoregressive time series analysis," *IEEE Transactions on Systems Science and Cybernetics*, vol. 4, no. 3, pp. 317–324, Sep. 1968.

[4] J. Ding, S. Shahrampour, K. Heal, and V. Tarokh, "Analysis of multi-state autoregressive models," *IEEE Transactions on Signal Processing*, vol. 66, no. 9, pp. 2429–2440, May 2018.

[5] P. Stoica, "Prediction of autoregressive lognormal processes," *IEEE Transactions on Automatic Control*, vol. 25, no. 2, pp. 292–293, April 1980.

[6] Y. Nardi and A. Rinaldo, "Autoregressive process modeling via the lasso procedure," *Journal of Multivariate Analysis*, vol. 102, no. 3, pp. 528 – 549, 2011.

[7] R. K. Yedavalli, "Conditions for the existence of a common quadratic lyapunov function via stability analysis of matrix families," in *Proceedings of the 2002 American Control Conference (IEEE Cat. No.CH37301)*, vol. 2, May 2002, pp. 1296–1301 vol.2.

[8] C. Yang, J. Sun, Q. Zhang, and X. Ma, "Lyapunov stability and strong passivity analysis for nonlinear descriptor systems," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 60, no. 4, pp. 1003–1012, April 2013.

[9] W. R. Melvin, "Stability properties of functional difference equations," *Journal of Mathematical Analysis and Applications*, vol. 48, no. 3, pp. 749 – 763, 1974.

[10] M. V. Dragosevic and S. S. Stankovic, "A generalized least squares method for frequency estimation," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 37, no. 6, pp. 805–819, June 1989.

[11] M. Morf and T. Kailath, "Square-root algorithms for least-squares estimation," *IEEE Transactions on Automatic Control*, vol. 20, no. 4, pp. 487–497, August 1975.

VII. Appendices

Appendix A - $y(t)$

$$\mathcal{H} = \frac{1}{1 - \lambda z^{-1}}$$
$$Y(z) = \mathcal{Q}(z)\mathcal{H}(z) + \mathcal{V}(z)$$
$$Y(z) + \frac{Y(z)}{\mathcal{H}(z)} = Y(z) + \mathcal{Q}(z) + \frac{\mathcal{V}(z)}{\mathcal{H}(z)}$$
$$Y(z) = Y(z) - \frac{Y(z)}{\mathcal{H}(z)} + \mathcal{Q}(z) + \frac{\mathcal{V}(z)}{\mathcal{H}(z)}$$
$$= Y(z)\left[1 - \frac{1}{\mathcal{H}(z)}\right] + \mathcal{Q}(z) + \frac{\mathcal{V}(z)}{\mathcal{H}(z)}$$
$$= Y(z)\left[\lambda z^{-1}\right] + \mathcal{Q}(z) + \mathcal{V}(z)\left[1 - \lambda z^{-1}\right]$$
$$y(t) = \lambda y(t-1) + q(t) + v(t) - \lambda v(t-1)$$

Appendix B - Expected Value $\mathbf{E}_y$

$$\mathbf{E}[y(t)] = \mathbf{E}[\lambda y(t-1) + q(t) + v(t) - \lambda v(t-1)]$$
$$= \lambda\mathbf{E}[y(t-1)] + \mathbf{E}[q(t)] + \mathbf{E}[v(t)] - \lambda\mathbf{E}[v(t-1)]$$
$$= \lambda\bar{y} + \bar{q} + \bar{v} - \lambda\bar{v}$$
$$= \bar{v} + \frac{\bar{q}}{1 - \lambda}$$

Appendix C - Variance Φ_y

$$\Phi_y = \mathbf{E}\left\{[y(t) - \bar{y}]^2\right\}$$
$$= \mathbf{E}\left[y(t)^2 - 2y(t)\bar{y} + \bar{y}^2\right]$$
$$= \mathbf{E}\left[y(t)^2\right] - \mathbf{E}\left[\bar{y}^2\right] \longrightarrow \mathbf{E}\left[\bar{y}^2\right] = 0$$
$$= \mathbf{E}\left\{\left[\underbrace{\lambda y(t-1) + q(t)}_{A} + \overbrace{v(t) - \lambda v(t-1)}^{B}\right]^2\right\}$$
$$= \mathbf{E}\left[A^2 + 2AB + B^2\right]$$
$$\mathbf{E}\left[A^2\right] = \mathbf{E}\left[\lambda^2 y(t-1)^2 + 2\lambda y(t-1)q(t) + q(t)^2\right]$$
$$= \mathbf{E}\left[\lambda^2 y(t-1)^2\right] + 2\mathbf{E}\left[\lambda y(t-1)q(t)\right] + \mathbf{E}\left[q(t)^2\right]$$
$$= \lambda^2 \Phi_y + \delta^2$$
$$\mathbf{E}\left[2AB\right] = 2\mathbf{E}\left[\lambda y(t-1)v(t) - \lambda^2 y(t-1)v(t-1) + q(t)v(t)\right.$$
$$\left. - \lambda q(t)v(t-1)\right]$$
$$= 2\mathbf{E}\left[\lambda y(t-1)v(t)\right] - 2\lambda^2\mathbf{E}\left[y(t-1)v(t-1)\right]$$
$$+ 2\mathbf{E}\left[q(t)v(t)\right] - 2\mathbf{E}\left[\lambda q(t)v(t-1)\right]$$
$$= -2\lambda^2\xi^2$$
$$\mathbf{E}\left[B^2\right] = \mathbf{E}\left[v(t)^2 - 2\lambda v(t)v(t-1) + \lambda^2 v(t-1)^2\right]$$
$$= \mathbf{E}\left[v(t)^2\right] - 2\lambda\mathbf{E}\left[v(t)v(t-1)\right] + \lambda^2\mathbf{E}\left[v(t-1)^2\right]$$
$$= \xi^2 + \lambda^2\xi^2$$
$$\Phi_y = \lambda^2\Phi_y + \delta^2 - 2\lambda^2\xi^2 + \xi^2 + \lambda^2\xi^2$$
$$= \frac{\delta^2}{1 - \lambda^2} + \xi^2$$

Appendix D - Covariance Ψ_y

$$\Psi_y(\tau) = \mathbf{E}\left\{[y(t) - \bar{y}][y(t-\tau) - \bar{y}]\right\} \longrightarrow \bar{y} = 0$$
$$= \mathbf{E}\left\{[ay(t-1) + e(t) + v(t) - av(t-1)][y(t-\tau)]\right\}$$

$$= \lambda \mathbf{E}\left[y(t-1)y(t-\tau)\right] + \mathbf{E}\left[e(t)y(t-\tau)\right]$$
$$+ \mathbf{E}\left[v(t)y(t-\tau)\right] - \lambda \mathbf{E}\left[v(t-1)y(t-\tau)\right]$$
$$\Psi_y(0) = \lambda\Psi(1) + \delta^2 + \xi^2$$
$$\Psi_y(1) = \lambda\Psi(0) - \lambda\xi^2$$
$$\Psi_y(2) = \lambda\Psi(1)$$

Appendix D * - Covariance $\Psi_y(\tau) \longrightarrow \tau = 0, 1, 2$

$$\Psi_y(1) := \frac{\Psi(1)}{\lambda} + \xi^2 = \lambda\Psi(1) + \delta^2 + \xi^2$$
$$:= \frac{\lambda\delta^2}{1-\lambda^2}$$
$$\Psi_y(0) := \lambda\frac{\lambda\delta^2}{1-\lambda^2} + \delta^2 + \xi^2$$
$$:= \frac{\delta^2}{1-\lambda^2} + \xi^2 \longrightarrow \Phi_y$$
$$\Psi_y(2) := \lambda\Psi(1)$$
$$:= \frac{\lambda^2\delta^2}{1-\lambda^2} \longrightarrow \frac{\lambda^\tau\delta^2}{1-\lambda^2} \text{ for } \Psi_y(\tau > 1)$$

Appendix E - $\Gamma(\theta_1) \longrightarrow \phi_1$

$$\Gamma(\theta_1) = \mathbf{E}\left\{[y(t) - \hat{y}(t|t-1)]^2\right\}$$
$$= \mathbf{E}\left\{[y(t) - \phi_1 y(t-1)]^2\right\}$$
$$= \mathbf{E}\left[y(t)^2 - 2\phi_1 y(t)y(t-1) + \phi_1^2 y(t-1)^2\right]$$
$$= \Psi_y(0) - 2\phi_1\Psi_y(1) + \phi_1^2\Psi_y(0)$$
$$= \left(1 + \phi_1^2\right)\Psi_y(0) - 2\phi_1\Psi_y(1)$$
$$\frac{\partial\Gamma(\theta_1)}{\partial\phi_1} = 2\phi_1\Psi_y(0) - 2\Psi_y(1) = 0$$
$$\phi_1 = \frac{\Psi_y(1)}{\Psi_y(0)} \longrightarrow \frac{\lambda\Psi_y(0) - \lambda\xi^2}{\Psi_y(0)}$$
$$= \frac{\lambda\delta^2}{\delta^2 + (1-\lambda^2)\xi^2}$$

Appendix F - $\Gamma(\theta_2) \longrightarrow \theta_2^\star := [\phi_1, \phi_2]^\top$

$$\Gamma(\theta_2) = \mathbf{E}\left\{[y(t) - \hat{y}(t|t-1)]^2\right\}$$
$$= \mathbf{E}\left\{[y(t) - \phi_1 y(t-1) - \phi_2 y(t-2)]^2\right\}$$
$$= \mathbf{E}\left[y(t)^2 + \phi_1^2 y(t-1)^2 + \phi_2^2 y(t-2)^2\right.$$
$$- 2\phi_1 y(t)y(t-1) - 2\phi_2 y(t)y(t-2)$$
$$\left. + 2\phi_1\phi_2 y(t-1)y(t-2)\right]$$
$$= \Psi_y(0) + \phi_1^2\Psi_y(0) + \phi_2^2\Psi_y(0) - 2\phi_1\Psi_y(1)$$
$$- 2\phi_2\Psi_y(2) + 2\phi_1\phi_2\Psi_y(1)$$
$$= \left(1 + \phi_1^2 + \phi_2^2\right)\Psi_y(0) + 2\phi_1\left(\phi_2 - 1\right)\Psi_y(1)$$
$$- 2\phi_2\Psi_y(2)$$
$$\frac{\partial\Gamma(\theta_2)}{\partial\phi_1} = 2\phi_1\Psi_y(0) + 2\phi_2\Psi_y(1) - 2\Psi_y(1) = 0$$
$$\phi_1 = \frac{\Psi_y(1) - \phi_2\Psi_y(1)}{\Psi_y(0)}$$

$$= \frac{\Psi_y(1) - \left(\dfrac{\Psi_y(2)\Psi_y(0) - \Psi_y(1)^2}{\Psi_y(0)^2 - \Psi_y(1)^2}\right)\Psi_y(1)}{\Psi_y(0)}$$
$$= \frac{\Psi_y(1)\Psi_y(0) - \Psi_y(2)\Psi_y(1)}{\Psi_y(0)^2 - \Psi_y(1)^2}$$
$$= \frac{\lambda\delta^2\left(\delta^2 + \xi^2\right)}{\left(\delta^2 + \xi^2\right)^2 - \lambda^2\xi^4}$$
$$\frac{\partial\Gamma(\theta_2)}{\partial\phi_1} = 2\phi_2\Psi_y(0) + 2\phi_1\Psi_y(1) - 2\Psi_y(2) = 0$$
$$\phi_2 = \frac{\Psi_y(2) - \phi_1\Psi_y(1)}{\Psi_y(0)}$$
$$= \frac{\Psi_y(2) - \left(\dfrac{\Psi_y(1)\Psi_y(0) - \Psi_y(2)\Psi_y(1)}{\Psi_y(0)^2 - \Psi_y(1)^2}\right)\Psi_y(1)}{\Psi_y(0)}$$
$$= \frac{\Psi_y(2)\Psi_y(0) - \Psi_y(1)^2}{\Psi_y(0)^2 - \Psi_y(1)^2}$$
$$= \frac{\lambda^2\delta^2\xi^2}{\left(\delta^2 + \xi^2\right)^2 - \lambda^2\xi^4}$$

Appendix G - $\Phi_{\theta_1^\star}$ and $\Phi_{\theta_2^\star}$

$$\Phi_{\theta_1^\star} = \Gamma(\theta_1) = \mathbf{E}\left\{[\varepsilon_{\theta_1}(t) - \bar{\varepsilon}_{\theta_1}]^2\right\} \longrightarrow \bar{\varepsilon}_{\theta_1} = 0$$
$$= \left[1 + \phi_1^2\right]\Psi_y(0) - 2\phi_1\Psi_y(1) \longrightarrow \phi_1 \text{ from Eq. (21)}$$
$$= \left[1 + \left(\frac{\lambda\delta^2}{\delta^2 + (1-\lambda^2)\xi^2}\right)^2\right]\frac{\delta^2 + (1-\lambda^2)\xi^2}{1-\lambda^2}$$
$$- 2\left(\frac{\lambda\delta^2}{\delta^2 + (1-\lambda^2)\xi^2}\right)\frac{\lambda\delta^2}{1-\lambda^2}$$
$$= \frac{\left[\delta^2 + \left(1-\lambda^2\right)\xi^2\right]^2 - \lambda^2\delta^4}{\left[\delta^2 + \left(1-\lambda^2\right)\xi^2\right]\left[1-\lambda^2\right]}$$
$$= \frac{\delta^4\left(1-\lambda^2\right) + 2\delta^2\xi^2\left(1-\lambda^2\right) + \xi^4\left(1-\lambda^2\right)^2}{\left[\delta^2 + \left(1-\lambda^2\right)\xi^2\right]\left[1-\lambda^2\right]}$$
$$= \frac{\left(\delta^2 + \xi^2\right)^2 - \lambda^2\xi^4}{\delta^2 + \left(1-\lambda^2\right)\xi^2}$$
$$\Phi_{\theta_2^\star} = \Gamma(\theta_2) = \mathbf{E}\left\{[\varepsilon_{\theta_2}(t) - \bar{\varepsilon}_{\theta_2}]^2\right\} \longrightarrow \bar{\varepsilon}_{\theta_2} = 0$$
$$= \left[1 + \phi_1^2 + \phi_2^2\right]\Psi_y(0) - [2\phi_1 - 2\phi_1\phi_2]\Psi_y(1)$$
$$- 2\phi_2\Psi_y(2) \longrightarrow [\phi_1, \phi_2] \text{ from Eq. (22)}$$
$$= \Psi_y(0)^3 - \Psi_y(2)^2\Psi_y(0) - 2\Psi_y(1)^2\Psi_y(0)$$
$$+ 2\Psi_y(1)^2\Psi_y(2) \longrightarrow \text{after some complex-algebra}$$
$$= \left(\frac{\delta^2 + (1-\lambda^2)\xi^2}{1-\lambda^2}\right)^3 - \left(\frac{\lambda^2\delta^2}{1-\lambda^2}\right)^2\frac{\delta^2 + (1-\lambda^2)\xi^2}{1-\lambda^2}$$
$$- 2\left(\frac{\lambda\delta^2}{1-\lambda^2}\right)^2\frac{\delta^2 + (1-\lambda^2)\xi^2}{1-\lambda^2} + 2\lambda\left(\frac{\lambda\delta^2}{1-\lambda^2}\right)^3$$
$$= \frac{\left(\delta^2 + \xi^2\right)^3 - \lambda^2\left(\delta^2\xi + \xi^3\right)^2 + 2\lambda\delta^4\xi^2}{1-\lambda^2}$$

Appendix H - $\Gamma(\theta_1) \longrightarrow \phi_1$; (non) zero-mean

$$\Gamma(\theta_1) = \mathbf{E}\left[y(t)^2 - 2y(t)y(t|t-1) + y(t|t-1)^2\right]$$
$$= \mathbf{E}\left\{[\breve{y}(t) + \bar{y}]^2\right\} - 2\phi_1\mathbf{E}\left\{[\breve{y}(t) + \bar{y}][\breve{y}(t-1) + \bar{y}]\right\}$$

$$+ \phi_1^2 \mathbf{E}\left\{ [\breve{y}(t-1) + \bar{y}]^2 \right\}$$
$$= \left[1 + \phi_1^2\right] \Psi_y(0) - 2\phi_1 \Psi_y(1) + (1 - \phi_1)^2 \bar{y}^2$$
$$\frac{\partial \Gamma(\theta_1)}{\partial \phi_1} = -2\Psi_y(1) + 2\phi_1 \Psi_y(0) - 2(1 - \phi_1)\bar{y}^2 = 0$$
$$\theta_1^\star := \frac{\Psi(1) + \bar{y}^2}{\Psi(0) + \bar{y}^2}$$

Appendix I - $\Gamma(\theta_2) \longrightarrow \theta_2^\star := [\phi_1, \phi_2]^\top$; (non) zero-mean

$$\Gamma(\theta_2) = \mathbf{E}\left[y(t)^2 - 2y(t)y(t|t-1) + y(t|t-1)^2 \right]$$
$$= \mathbf{E}\left\{ [\breve{y}(t) + \bar{y}]^2 \right\} - 2\mathbf{E}\Big\{ [\breve{y}(t) + \bar{y}] [\phi_1 \breve{y}(t-1) $$
$$+ \phi_2 \breve{y}(t-2) + (\phi_1 + \phi_2)\bar{y}] \Big\} + \phi_1^2 \mathbf{E}\left\{ [\breve{y}(t-1)]^2 \right\}$$
$$+ \phi_2^2 \mathbf{E}\left\{ [\breve{y}(t-2)]^2 \right\} + (\phi_1 + \phi_2) \mathbf{E}\left\{ [\bar{y}]^2 \right\}$$
$$+ 2\phi_1 (\phi_1 + \phi_2) \mathbf{E}\left\{ [\breve{y}(t-1)\bar{y}] \right\}$$
$$+ 2\phi_2 (\phi_1 + \phi_2) \mathbf{E}\left\{ [\breve{y}(t-2)\bar{y}] \right\}$$
$$+ 2\phi_1 \phi_2 \mathbf{E}\left\{ [\breve{y}(t-1)\breve{y}(t-2)] \right\}$$
$$= \Psi_y(0) + \phi_1^2 \Psi_y(0) + \phi_2^2 \Psi_y(0) - 2\phi_1 \Psi_y(1)$$
$$- 2\phi_2 \Psi_y(2) + 2\phi_1 \phi_2 \Psi_y(1) + \left[1 - (\phi_1 + \phi_2)\right]^2 \bar{y}^2$$
$$\frac{\partial \Gamma(\theta_2)}{\partial \phi_1} = 0 \text{ and } \frac{\partial \Gamma(\theta_2)}{\partial \phi_2} = 0$$

after some complex-algebra

$$\phi_1 = \frac{\Psi_y(1)\Psi_y(0) - \Psi_y(2)\Psi_y(1) + \bar{y}^2 [\Psi(0) - \Psi(2)]}{\Psi_y(0)^2 - \Psi_y(1)^2 + 2\bar{y}^2 [\Psi(0) - \Psi(1)]}$$
$$\phi_2 = \frac{\Psi_y(2)\Psi_y(0) - \Psi_y(1)^2 + \bar{y}^2 [\Psi(0) - 2\Psi(1) + \Psi(2)]}{\Psi_y(0)^2 - \Psi_y(1)^2 + 2\bar{y}^2 [\Psi(0) - \Psi(1)]}$$

Appendix J - $\Phi_{\theta_1^\star}$ and $\Phi_{\theta_2^\star}$; (non) zero-mean

$$\bar{\varepsilon}_{\theta_1} = \mathbf{E}\left[\breve{y}(t)\right] + \mathbf{E}\left[\bar{y}\right] - \phi_1 \mathbf{E}\left[\breve{y}(t-1)\right] - \phi_1 \mathbf{E}\left[\bar{y}\right]$$
$$= (1 - \phi_1)\bar{y}$$
$$\Phi_{\theta_1^\star} = \Gamma(\theta_1) = \mathbf{E}\left\{ \left[\varepsilon_{\theta_1}(t) - \bar{\varepsilon}_{\theta_1}\right]^2 \right\}$$
$$= \mathbf{E}\left\{ [\breve{y}(t) + \bar{y} - \phi_1 \breve{y}(t-1) - \phi_1 \bar{y} - (1 - \phi_1)\bar{y}]^2 \right\}$$
$$= \mathbf{E}\left\{ [\breve{y}(t) - \phi_1 \breve{y}(t-1)]^2 \right\} \longrightarrow \textbf{Appendix E}$$
$$= \frac{(\delta^2 + \xi^2)^2 - \lambda^2 \xi^4}{\delta^2 + (1 - \lambda^2)\xi^2} \longrightarrow \textbf{Appendix G}$$
$$\bar{\varepsilon}_{\theta_2} = \mathbf{E}\left[\breve{y}(t)\right] + \mathbf{E}\left[\bar{y}\right] - \phi_1 \mathbf{E}\left[\breve{y}(t-1)\right] - \phi_2 \mathbf{E}\left[\breve{y}(t-2)\right]$$
$$- (\phi_1 + \phi_2)\mathbf{E}\left[\bar{y}\right]$$
$$= (1 - \phi_1 - \phi_2)\bar{y}$$
$$\Phi_{\theta_2^\star} = \Gamma(\theta_2) = \mathbf{E}\left\{ \left[\varepsilon_{\theta_2}(t) - \bar{\varepsilon}_{\theta_2}\right]^2 \right\}$$
$$= \mathbf{E}\Big\{ [\breve{y}(t) + \bar{y} - \phi_1 \breve{y}(t-1) - \phi_2 \breve{y}(t-2)$$
$$- (\phi_1 + \phi_2)\bar{y} - (1 - \phi_1 - \phi_2)\bar{y}]^2 \Big\}$$
$$= \mathbf{E}\left\{ [\breve{y}(t) - \phi_1 \breve{y}(t-1) - \phi_2 \breve{y}(t-2)]^2 \right\}$$
$$= \frac{(\delta^2 + \xi^2)^3 - \lambda^2 (\delta^2 \xi + \xi^3)^2 + 2\lambda \delta^4 \xi^2}{1 - \lambda^2}$$

AUTHORS

First and Correspondence Author - Moh Kamalul Wafi is with Laboratory of Embedded Systems and Control, Department of Engineering Physics, Institut Teknologi Sepuluh Nopember, 60111, Indonesia