Shahid Bahonar University of Kerman
Journal of Mahani Mathematical Research
Print ISSN: 2251-7952 Online ISSN: 2645-4505

INFERENCE IN UNIVARIATE AND BIVARIATE AUTOREGRESSIVE MODELS WITH NON-NORMAL INNOVATIONS

S. Zamani Mehreyan and A. Sayyareh ✉

Article type: Research Article
(Received: 22 January 2022, Received in revised form: 15 May 2022)
(Accepted: 07 June 2022, Published Online: 21 June 2022)

Abstract. In this paper we consider the estimation, order and model selection of autoregressive time series model which may be driven by non-normal innovations. The paper makes two contributions. First, we consider the method of moments for a univariate and also a bivariate time series model; the importance of using the method of moments is that it can provide us with consistent estimates easily for any model order and for any kind of distribution that we can assume for the non-normal innovations. Second, we provide methods for order and model selection, i.e., for selecting the order of the autoregression and the model for the innovation distribution. Our analysis provides analytic results on the asymptotic distribution of the method of moments estimators and also computational results via simulations. Our results show that although the performance of modified maximum likelihood estimators is better than method of moments estimators when the sample size is small but both methods have approximately same performance as the sample size increase and in misspecification case. Also, it is shown that focussed information criterion is an appropriate criterion for model selection for autoregressive models with non-normal innovations based on the method of moments estimators.

Keywords: Autoregressive order selection, Focussed information criterion, Method of moments estimation, Misspecified model, Nonnested models.
2020 MSC: 62M10, 62F10, 62F12, 62H12.

1. Introduction

Estimation, order and model selection are integral parts in time series analysis. When we collect a set of data and want to study its properties, it is important to select an appropriate functional form of the suggested model. How are we then to identify a model when it can not be completely specified from a priori knowledged? Although one might expect that a more complicated model will provide a better approximation to the data at hand, we are frequently faced with situations that a less complicated model is better, in terms of accuracy of parameter estimation or prediction of future values. Then, the performance of

✉ asayyareh@kntu.ac.ir, ORCID: 0000-0002-8585-9900
DOI: 10.22103/jmmr.2022.18874.1196
Publisher: Shahid Bahonar University of Kerman
How to cite: S.S. Zamani Mehreyan, A. Sayyareh, *Inference in Univariate and Bivariate Autoregressive Models with Non-Normal Innovations*, J. Mahani Math. Res. 2023; 12(1): 59-89.

a model selection criterion is optimal if the selected model is the most accurate model in the family of models that we might be contemplating.
Suppose therefore that we restrict attention to the family of autoregressive models of order p, i.e., AR(p):

$$x_t = \phi_1 x_{t-1} + ... + \phi_p x_{t-p} + \epsilon_t,$$

where the variables x_t's are correlated and the innovation terms ϵ_t's are independent and identically normally distributed with zero mean and finite variance. Determination of the order p of the model is an important step in autoregressive modelling. The Akaike information criterion, AIC, is an asymptotically unbiased estimator of the Kullback-Leibler criterion, KL, that can be used amongst other order selection criteria, see Akaike [2] and Kullback and Leibler [15]. This order selection criterion was first introduced by Akaike [1]. AIC is known to suffer from overfit, see Shibata [16]. The selected order of model can be greater than the optimal order. Claeskens et al. [11] proposed an adapted version of the Focussed Information Criterion, FIC, for order selection with focus on a high predictive accuracy as defined in Claeskens and Hjort [10]. The FIC estimates the Mean Squared Error, MSE, of the estimation of a focused parameter. This focused parameter is the h-step ahead prediction of the time series. It selects the model that yields the best estimation for the focused parameter from a proposed family of AR(p) models. Best here is defined in the sense of having the lowest mean squared forecast error where the parameter estimates are obtained using ordinary least squares, OLS, approach.

However, in modelling real-world time series data, the potential asymmetry of the marginal innovation distribution can frequently create some problems. The problem of a skewed distribution has been handled by a number of methods. A widely used technique is to use transformations to render a series close to normality, see Box and Jenkins [9]. However, there are cases where one might want to consider directly a class of non-normal models. For these cases, a number of non-normal models with autoregressive-type correlation structure have been proposed. The simplest of this kind of models corresponds to the so-called exponential autoregressive, EAR, model, see Gaver and Lewis [13]. They have shown that, there is an innovation process ϵ_t such that the sequence of random variables x_t generated by the first order autoregressive, AR, scheme, $x_t = \phi x_{t-1} + \epsilon_t$, is marginally distributed as a Gamma distribution if $\phi \in [0, 1]$. They claimed that this first order autoregressive Gamma sequence is useful for modelling a wide range of observed phenomena.
Steyn [17] proposed the generalized least squares estimation procedure using more than one sample statistic method for estimating the parameter of a standard Gamma distribution and derived new estimators for the parameter of a first order autoregressive process. Gourieroux and Jasiak [14] introduced the class of autoregressive Gamma processes with conditional distributions from the family of noncentred Gamma and provided the stationarity and ergodicity conditions for autoregressive Gamma processes of any autoregressive order p,

including long memory, and closed-form expressions of conditional moments. Bondon [8] introduced a non-Gaussian autoregressive model with epsilon-skew-normal innovations and derived the moments and maximum likelihood estimators of the parameters. Balakrishna [6] considered a variety of non-Gaussian autoregressive-type models to analyze time-series data and provided their probabilistic and inferential properties.

In this paper we consider univariate and bivariate autoregressive models where the innovation terms are assumed to be independent non-normal variables with mean μ_ϵ and finite variance σ^2_ϵ. We forecast the series x_t at horizon h by

$$x_{n+h} = \hat{\phi}_0 + \hat{\phi}_1 x_{n+h-1} + ... + \hat{\phi}_p x_{n+h-p},$$

where $\phi_0 = \mu_\epsilon$ and the parameters are estimated using the method of moments. Note that, while we can always use OLS to estimate the parameters of such a model it is clearly non-optimal, and maximum likelihood estimators, MLE, should be preferred. However, the MLEs are neither straightforward analytically nor easy (to specify) computationally and, most importantly, are difficult to specify and estimate for orders greater than p equal to one. Therefore, in this paper we consider the simpler, yet analytically and computationally more elegant, method of moments, MME. We show that the MME is applicable in any model order $p \geq 1$ and is easily extensible to bivariate time series model as well. Furthermore, with the use of the MME we can easily obtain estimators for both the autoregressive component of the model and the parameters of the marginal distribution of the innovations. We also extend the FIC where the focused parameter is the h-step ahead prediction of the non-normal autoregressive with the constant term being the mean of the innovations.

The rest of the paper is structured as follows: in Section 2, the method of moment estimation of univariate and bivariate autoregressive parameters and there asymptotic distribution are derived. In Section 3, the focussed information criterion is improved for the class of models that we consider. In Section 4, we study the obtained theoretical results by simulation. Also, we illustrated our theoretical results with the analysis of a real dataset in Section 5 while Section 6 offers some concluding remarks and directions for future research.

2. Model and parameter estimation

We start off by listing the most commonly used methods of parameter estimation in autoregressive models, they are:

1. Method of Moments, MM,
2. Least Squares Method, OLS,
3. Maximum Likelihood Method, ML,
4. Modified Maximum Likelihood Method, MML.

The least squares method is inefficient when the innovations have a non-normal distribution. Also the maximum likelihood method might be computationally problematic, since explicit solutions from the likelihood equations cannot be obtained and iterative methods have to be used for which consistent initial estimates must be available. Hence, the modified maximum likelihood method applied to first order autoregressive model can be a viable alternative. The modified maximum likelihood method has been developed by Tiku [18] and applied to some non-normal time series models. This method is based on linearization of intractable terms of the log-likelihood function using first-order Taylor series expansion. See, for example, Akkaya and Tiku [3, 4], Akkaya and Tiku [5] and Bayrak and Akkaya [7]. The MMLE would complicate the computations too and is not easily extendable to orders greater than $p = 1$. Hence, the univariate and bivariate p-order autoregressive models are considered and estimated using the method of moments.

2.1. **Univariate autoregressive model.** A p-order univariate autoregressive is specified as

$$x_t = \sum_{j=1}^{p} \phi_j x_{t-j} + \epsilon_t, \tag{1}$$

where x_t is the n-dimensional vector of dependent variables for a sample of size n, the ϵ_t's are independent and identically distributed, i.i.d, with mean μ_ϵ and variance σ_ϵ^2 and are uncorrelated with $x_1, ..., x_{t-1}$. The assumption that we make in this paper is $\sum_{j=1}^{p} \phi_j < 1$, so that x_t is a stationary process. Since the mean of process is not zero, we can use the standard form where the time series model consist of a constant,

$$x_t = \phi_0 + \sum_{j=1}^{p} \phi_j x_{t-j} + \epsilon_t^*, \tag{2}$$

where $\phi_0 = \mu_\epsilon = \left(1 - \sum_{j=1}^{p} \phi_j\right) \mu_x$, $\mu_x = E(X_t)$ and $\epsilon_t^*,\ t = 1, ..., n$ are i.i.d random variable with location parameter $\mu_{\epsilon^*} = E(\epsilon_t^*) = 0$.

The method of moments is the oldest method that we can use for the estimation of unknown parameters. Although MMEs may not be the best estimators they almost always produce some asymptotically unbiased and consistent estimators. Let $\mu_{\epsilon,j} = E(\epsilon^j) = h(\theta)$ be the jth moment and let $\hat{\mu}_{\epsilon,j} = \frac{1}{n} \sum_{t=1}^{n} \epsilon_t^j$ be the jth sample moment, which is an unbiased estimator of $\mu_{\epsilon,j},\ j = 1, ..., k$, where θ is a vector of unknown model parameters. We obtain a moments-based estimator $h(\hat{\theta})$ as

$$\hat{\mu}_{\epsilon,j} = h(\hat{\theta}),\ \ j = 1, ..., k.$$

If the inverse function $g = h^{-1}$ exists, then the unique moments estimation of θ is $\hat{\theta} = h^{-1}(\hat{\mu}_\epsilon)$. If g is continuous at μ_ϵ, then $\hat{\theta}$ is strongly consistent for θ, since

$$\hat{\mu}_{\epsilon,j} \xrightarrow{a.s} \mu_{\epsilon,j}.$$

Let us therefore start-off by pointing to the relationship between the mean of x_t and the mean of the innovations ϵ_t, i.e.,

$$\mu_x = \mu_x \sum_{j=1}^{p} \phi_j + \mu_\epsilon.$$

Or equivalently

$$\mu_x \left(1 - \mathbf{e}^t \phi\right) = \mu_\epsilon, \tag{3}$$

where $\mathbf{e} = (1, ..., 1)^t$ and $\phi = (\phi_{\mathbf{1}}, ..., \phi_{\mathbf{p}})^{\mathbf{t}}$ are $(p \times 1)$ vectors. Next, take the usual moment conditions as in the case of the Yule-Walker estimators, Yule [20] and Walker [19], and write:

$$E\left(X_t X_{t-k}\right) = \sum_{j=1}^{p} \phi_j E\left(X_{t-j} X_{t-k}\right) + E\left(\epsilon_t X_{t-k}\right).$$

For $k = 1, ..., p$, we have the following system of equations

$$\begin{aligned}
\sigma(1) &= \sigma(0)\phi_1 + \sigma(1)\phi_2 + ... + \sigma(p-1)\phi_p + \mu_\epsilon \mu_x \\
\sigma(2) &= \sigma(1)\phi_1 + \sigma(0)\phi_2 + ... + \sigma(p-2)\phi_p + \mu_\epsilon \mu_x \\
&\vdots \\
\sigma(p) &= \sigma(p-1)\phi_1 + \sigma(p-2)\phi_2 + ... + \sigma(0)\phi_p + \mu_\epsilon \mu_x
\end{aligned}$$

where $\sigma(k) = E(X_t X_{t-k})$ for $k = 1, ..., p$. The matrix notation of system of equations is given by:

$$\sigma = \Sigma\phi + \mu_\epsilon \mu_x \mathbf{e},$$

where $\mu_x = E(X_t)$, $\sigma = (\sigma(1), ..., \sigma(p))^t$, and

$$\Sigma = \begin{bmatrix} \sigma(0) & \sigma(1) & ... & \sigma(p-1) \\ \sigma(1) & \sigma(0) & ... & \sigma(p-2) \\ . & & & \\ . & & & \\ . & & & \\ \sigma(p-1) & \sigma(p-2) & ... & \sigma(0) \end{bmatrix}.$$

So

$$\hat{\phi} = \hat{\Sigma}^{-1}\hat{\sigma} - \hat{\mu}_\epsilon \hat{\mu}_x \hat{\Sigma}^{-1}\mathbf{e} \tag{4}$$

The first term of (4) is indeed the usual Yule-Walker estimator for the autoregressive terms and second term is related to $\hat{\mu}_\epsilon$, so we need to estimate the mean of the innovations from knowledge only of the autoregressive parameters and the moments of the observations. Substituting (4) in (3) we obtain:

$$\mu_\epsilon = \mu_x \left(1 - \mathbf{e}^t \Sigma^{-1} \sigma\right) + \mu_\epsilon \mu_x^2 \mathbf{e}^t \Sigma^{-1} \mathbf{e},$$

which implies

$$\hat{\mu}_\epsilon = \frac{\hat{\mu}_x \left(1 - \mathbf{e}^t \hat{\Sigma}^{-1} \hat{\sigma}\right)}{1 - \hat{\mu}_x^2 \mathbf{e}^t \hat{\Sigma}^{-1} \mathbf{e}}, \tag{5}$$

which is also an MME for the mean of the innovations. Similarly, the relationship between the variance of x_t and the variance of the innovations is:

$$\begin{aligned} var(X_t) &= \sum_{j=1}^{p} \phi_j^2 var(X_{t-j}) + \sum_{i \neq j} \phi_i \phi_j cov(X_{t-i}, X_{t-j}) + var(\epsilon_t) \\ &= \phi \Sigma_x \phi^t + \sigma_\epsilon^2, \end{aligned}$$

where

$$\Sigma_X = \begin{bmatrix} \gamma(0) & \gamma(1) & ... & \gamma(p-1) \\ \gamma(1) & \gamma(0) & ... & \gamma(p-2) \\ . & & & \\ . & & & \\ . & & & \\ \gamma(p-1) & \gamma(p-2) & ... & \gamma(0) \end{bmatrix}$$

and $\gamma(k) = E((X_t - \mu_x)(X_{t-k} - \mu_x))$. Therefore

$$\sigma_\epsilon^2 = \sigma_x^2 - \phi^t \Sigma_x \phi,$$

where $\sigma_\epsilon^2 = var(\epsilon_t)$ and $\sigma_x^2 = var(X_t)$, and of course we obtain consistent estimators by estimating σ_x^2 by the data and σ_ϵ^2 by the data and the MMEs of the autoregressive parameters.

2.2. **Bivariate autoregressive model.** The above discussion is conceptually very easy to be expanded in a bivariate context. To this end, let $\mathbf{X}_t = (X_{1t}, X_{2t})$ denote a (2×1) vector of time series variables. To simplify our explicitly equations we present the basic 2-lag vector autoregressive, VAR(2), model as

$$\begin{bmatrix} x_{1,t} \\ x_{2,t} \end{bmatrix} = \begin{bmatrix} \phi_{11}^{(1)} & \phi_{12}^{(1)} \\ \phi_{12}^{(1)} & \phi_{22}^{(1)} \end{bmatrix} \begin{bmatrix} x_{1,t-1} \\ x_{2,t-1} \end{bmatrix} + \begin{bmatrix} \phi_{11}^{(2)} & \phi_{12}^{(2)} \\ \phi_{12}^{(2)} & \phi_{22}^{(2)} \end{bmatrix} \begin{bmatrix} x_{1,t-2} \\ x_{2,t-2} \end{bmatrix} + \begin{bmatrix} \epsilon_{1,t} \\ \epsilon_{2,t} \end{bmatrix}$$

or

$$\mathbf{X}_t = \phi_1 \mathbf{X}_{t1} + \phi_2 \mathbf{X}_{t2} + \epsilon_t, \;\; t = 1, ..., n, \tag{6}$$

where ϕ_i's are (2×2) coefficient matrices and $\epsilon_{\mathbf{t}}$ is an (2×1) unobservable non-zero mean vector of variables with time invariant covariance matrix. The material that follows is immediately adaptable to the case where $p > 2$ and the same equation structure applies (in fact in our computations we provide for the full bivariate VAR(p) model). Under the assumption of stationarity (based on the eigenvalues of the companion form of the model)

$$\mu_{\mathbf{x}} = \phi_{\mathbf{1}} \mu_{\mathbf{x}} + \phi_{\mathbf{2}} \mu_{\mathbf{x}} + \mu_\epsilon$$

so

$$\mu_\epsilon = (I - \phi_1 - \phi_2)\mu_x, \tag{7}$$

where $\mu_{\mathbf{x}} = \mathbf{E}(\mathbf{X_t})$, $\mu_\epsilon = \mathbf{E}(\epsilon_\mathbf{t})$ and I is an identity matrix. The autocovariance is obtained as:

$$E(X_t X_{t-k}^t) = \phi_1 E(X_{t-1} X_{t-k}^t) + \phi_2 E(X_{t-2} X_{t-k}^t) + E(\epsilon_t X_{t-k}^t).$$

For $k = 1, 2$, we have:

$$\Gamma(1) = \phi_1 \Gamma(0) + \phi_2 \Gamma(1) + \mu_\epsilon \mu_x$$

and

$$\Gamma(2) = \phi_1 \Gamma(1)^t + \phi_2 \Gamma(0) + \mu_\epsilon \mu_x.$$

In matrix notation of system of equations is thus given by:

$$\Gamma = \Sigma_B \phi + e \otimes \mu_x \mu_\epsilon,$$

where:

$$\Gamma = \begin{bmatrix} \Gamma(1)^t \\ \Gamma(2)^t \end{bmatrix}_{(4\times 2)}, \Sigma_B = \begin{bmatrix} \Gamma(0) & \Gamma(1) \\ \Gamma(1)^t & \Gamma(0) \end{bmatrix}_{(4\times 4)}, \phi = \begin{bmatrix} \phi_1^t \\ \phi_2^t \end{bmatrix}_{(4\times 2)}, e = \begin{bmatrix} 1 \\ 1 \end{bmatrix}_{(2\times 1)}.$$

So, collecting terms we can now have:

$$\phi = \Sigma_B^{-1} \Gamma - \Sigma_B^{-1} \left(e \otimes \mu_x \right) \mu_\epsilon^t. \tag{8}$$

Define the matrix operation $S = e^t \otimes I$ and see that $S\phi = \phi_1^t + \phi_2^t$. We can now re-write (7) as:

$$\begin{aligned} \mu_\epsilon &= (I - \phi^t S^t)\mu_x = \mu_x - \phi^t S^t \mu_x \\ &= \mu_x - \left(\Gamma \Sigma_B^{-1} S^t - \mu_\epsilon \left(e^t \otimes \mu_x^t \right) \Sigma_B^{-1} S^t \right) \mu_x. \end{aligned}$$

Hence, we arrive at an explicit expression for the innovation mean vector as:

$$\mu_\epsilon = \left(I - \Gamma \Sigma_B^{-1} S^t \right) \mu_x \left(I - (e^t \otimes \mu_x^t) \Sigma_B^{-1} S^t \mu_x \right)^{-1}.$$

We can also go a step further here. The covariance of the model in (6) is given as:

$$
\begin{aligned}
cov(X_{1,t},X_{2,t}) &= E\left(X_{1,t}-\mu_x\right)\left(X_{2,t}-\mu_x\right)\\
&= E\Big\{\Big[\phi_{11}^{(1)}(X_{1,t-1}-\mu_{1,x})+\phi_{12}^{(1)}(X_{2,t-1}-\mu_{2,x})+\phi_{11}^{(2)}(X_{1,t-2}-\mu_{1,x})\\
&+\phi_{12}^{(2)}(X_{2,t-2}-\mu_{2,x})+(\epsilon_{1,t}-\mu_{1,\epsilon})\Big]\times\Big[\phi_{21}^{(1)}(X_{1,t-1}-\mu_{1,x})+\phi_{22}^{(1)}(X_{2,t-1}-\mu_{2,x})\\
&+\phi_{21}^{(2)}(X_{1,t-2}-\mu_{1,x})+\phi_{22}^{(2)}(X_{2,t-2}-\mu_{2,x})+(\epsilon_{2,t}-\mu_{2,\epsilon})\Big]\Big\}\\
&= \phi_{11}^{(1)}\phi_{21}^{(1)}E\left(X_{1,t-1}-\mu_{1,x}\right)\left(X_{1,t-1}-\mu_{1,x}\right)+\phi_{12}^{(1)}\phi_{21}^{(1)}E\left(X_{2,t-1}-\mu_{2,x}\right)\left(X_{1,t-1}-\mu_{1,x}\right)\\
&+\phi_{11}^{(2)}\phi_{21}^{(1)}E\left(X_{1,t-1}-\mu_{1,x}\right)\left(X_{1,t-1}-\mu_{1,x}\right)+\phi_{12}^{(2)}\phi_{21}^{(1)}E\left(X_{2,t-1}-\mu_{2,x}\right)\left(X_{1,t-1}-\mu_{1,x}\right)\\
&+\phi_{11}^{(1)}\phi_{22}^{(1)}E\left(X_{1,t-1}-\mu_{1,x}\right)\left(X_{2,t-1}-\mu_{2,x}\right)+\phi_{12}^{(1)}\phi_{22}^{(1)}E\left(X_{2,t-1}-\mu_{2,x}\right)\left(X_{2,t-1}-\mu_{2,x}\right)\\
&+\phi_{11}^{(2)}\phi_{22}^{(1)}E\left(X_{1,t-1}-\mu_{1,x}\right)\left(X_{2,t-1}-\mu_{2,x}\right)+\phi_{12}^{(2)}\phi_{22}^{(1)}E\left(X_{2,t-2}-\mu_{2,x}\right)\left(X_{2,t-1}-\mu_{2,x}\right)\\
&+\phi_{11}^{(1)}\phi_{21}^{(2)}E\left(X_{1,t-1}-\mu_{1,x}\right)\left(X_{1,t-2}-\mu_{1,x}\right)+\phi_{12}^{(1)}\phi_{21}^{(2)}E\left(X_{2,t-1}-\mu_{2,x}\right)\left(X_{1,t-2}-\mu_{1,x}\right)\\
&+\phi_{11}^{(2)}\phi_{21}^{(2)}E\left(X_{1,t-2}-\mu_{1,x}\right)\left(X_{1,t-2}-\mu_{1,x}\right)+\phi_{12}^{(2)}\phi_{21}^{(2)}E\left(X_{2,t-2}-\mu_{2,x}\right)\left(X_{1,t-2}-\mu_{1,x}\right)\\
&+\phi_{11}^{(1)}\phi_{22}^{(2)}E\left(X_{1,t-1}-\mu_{1,x}\right)\left(X_{2,t-2}-\mu_{2,x}\right)+\phi_{12}^{(1)}\phi_{22}^{(2)}E\left(X_{2,t-1}-\mu_{2,x}\right)\left(X_{2,t-2}-\mu_{2,x}\right)\\
&+\phi_{11}^{(2)}\phi_{22}^{(2)}E\left(X_{1,t-2}-\mu_{1,x}\right)\left(X_{2,t-2}-\mu_{2,x}\right)+\phi_{12}^{(2)}\phi_{22}^{(2)}E\left(X_{2,t-2}-\mu_{2,x}\right)\left(X_{2,t-2}-\mu_{2,x}\right)\\
&+\phi_{11}^{(1)}\phi_{21}^{(1)}E\left(X_{1,t-1}-\mu_{1,x}\right)\left(X_{1,t-1}-\mu_{1,x}\right)+\phi_{12}^{(1)}\phi_{21}^{(1)}E\left(X_{2,t-1}-\mu_{2,x}\right)\left(X_{1,t-1}-\mu_{1,x}\right)\\
&+E\left(\epsilon_{1,t}-\mu_{1,\epsilon}\right)\left(\epsilon_{2,t}-\mu_{2,\epsilon}\right)
\end{aligned}
$$

where ϵ_t is uncorrelated with X_{t-1}, X_{t-2}. Hence note that:

$$
\begin{aligned}
cov(X_{1,t},X_{2,t}) &= \phi_{11}^{(1)}\phi_{21}^{(1)}\gamma_1(0)+\phi_{12}^{(1)}\phi_{21}^{(1)}\gamma_{21}(0)+\phi_{11}^{(2)}\phi_{21}^{(1)}\gamma_1(1)+\phi_{12}^{(2)}\phi_{21}^{(1)}\gamma_{21}(1)\\
&+\phi_{11}^{(1)}\phi_{22}^{(1)}\gamma_{12}(0)+\phi_{12}^{(1)}\phi_{22}^{(1)}\gamma_2(0)+\phi_{11}^{(2)}\phi_{22}^{(1)}\gamma_{12}(1)+\phi_{12}^{(2)}\phi_{22}^{(1)}\gamma_2(1)\\
&+\phi_{11}^{(1)}\phi_{21}^{(2)}\gamma_1(1)+\phi_{12}^{(1)}\phi_{21}^{(2)}\gamma_{21}(1)+\phi_{11}^{(2)}\phi_{21}^{(2)}\gamma_1(0)+\phi_{12}^{(2)}\phi_{21}^{(2)}\gamma_{21}(0)\\
&+\phi_{11}^{(1)}\phi_{22}^{(2)}\gamma_{12}(1)+\phi_{12}^{(1)}\phi_{22}^{(2)}\gamma_2(1)+\phi_{11}^{(2)}\phi_{22}^{(2)}\gamma_{12}(0)+\phi_{12}^{(2)}\phi_{22}^{(2)}\gamma_2(0)\\
&+cov(\epsilon_{1,t},\epsilon_{2,t})\\
&= \begin{bmatrix}\phi_1^{(1)} & \phi_1^{(2)}\end{bmatrix}\begin{bmatrix}\Gamma_x(0) & \Gamma_x(1)\\ \Gamma_x^t(1) & \Gamma_x(0)\end{bmatrix}\begin{bmatrix}\phi_2^{(1)}\\ \\ \phi_2^{(2)}\end{bmatrix}+\sigma_{\epsilon_1\epsilon_2}
\end{aligned}
$$

where for $i,j=1,2$ and $k=0,1$:

$$
\sigma_{\epsilon_1\epsilon_2}=cov(\epsilon_{1,t},\epsilon_{2,t}),\ \ \phi_i^{(j)}=\begin{bmatrix}\phi_{i1}^{(j)} & \phi_{i2}^{(j)}\end{bmatrix},\ \ \Gamma_x(k)=\begin{bmatrix}\gamma_1(k) & \gamma_{12}(k)\\ \gamma_{21}(k) & \gamma_2(k)\end{bmatrix}.
$$

Similarly, we can obtain the variances of the two series as below:

$$
\begin{aligned}
var(X_{1,t}) &= \begin{bmatrix}\phi_1^{(1)} & \phi_1^{(2)}\end{bmatrix}\begin{bmatrix}\Gamma_x(0) & \Gamma_x^t(1)\\ \Gamma_x(1) & \Gamma_x(0)\end{bmatrix}\begin{bmatrix}\phi_1^{(1)}\\ \\ \phi_1^{(2)}\end{bmatrix}+var(\epsilon_{1,t})\\
&= \begin{bmatrix}\phi_1^{(1)} & \phi_1^{(2)}\end{bmatrix}\Sigma_{B,x}\begin{bmatrix}\phi_1^{(1)}\\ \\ \phi_1^{(2)}\end{bmatrix}+\sigma_{\epsilon_1}^2
\end{aligned}
$$

and

$$var(X_{2,t}) = \begin{bmatrix} \phi_2^{(1)} & \phi_2^{(2)} \end{bmatrix} \begin{bmatrix} \Gamma_x(0) & \Gamma_x(1) \\ \Gamma_x^t(1) & \Gamma_x(0) \end{bmatrix} \begin{bmatrix} \phi_2^{(1)} \\ \\ \phi_2^{(2)} \end{bmatrix} + var(\epsilon_{2,t})$$

$$= \begin{bmatrix} \phi_2^{(1)} & \phi_2^{(2)} \end{bmatrix} \Sigma_{B,x} \begin{bmatrix} \phi_2^{(1)} \\ \\ \phi_2^{(2)} \end{bmatrix} + \sigma_{\epsilon_2}^2$$

where $\sigma_{\epsilon_2}^2 = var(\epsilon_{2,t})$, $\sigma_{\epsilon_1}^2 = var(\epsilon_{1,t})$, $\Sigma_{B,x} = \begin{bmatrix} \Gamma_x(0) & \Gamma_x(1) \\ \Gamma_x^t(1) & \Gamma_x(0) \end{bmatrix}$. As in the case of the univariate model in all the above expressions we can obtain consistent estimators of the various parameters (autoregressive and those of the innovations) by the use of the sample moments of the data for the vector of the means and the various autocovariance matrices.

2.3. **Parameter estimation.** Once we have estimates of the autoregressive parameters and the means and variances of the innovations, we can postulate a model for the marginal distribution of the innovations and use again the method of moments to obtain estimates for the structural parameters of this distribution. We will give various estimators from practical, for non-normal distributions. Thus, consider again the p-order autoregressive model in (1) where the innovation terms are independent and distributed as $f(\epsilon_t; \theta)$. Let ϕ denote the $(p \times 1)$ vector of autoregressive coefficients $(\phi_1, ..., \phi_p)^t$, θ denote the $(q \times 1)$ vector of distributional parameters. Let $\hat{\mu}_{\epsilon,j}$ denote the appropriate sample moments vector $(q \times 1)$ as well, then in all proposed models, ϕ is estimated as

$$\hat{\phi} = \hat{\Sigma}^{-1}\hat{\sigma} - \hat{\mu}_\epsilon \hat{\mu}_x \hat{\Sigma}^{-1}\mathbf{e}.$$

An estimator for θ is given by $\hat{\theta} = h^{-1}(\hat{\mu}_{\epsilon,j})$ and by solving the system of equations $h(\theta) = \hat{\mu}_{\epsilon,j}$.

Let us start with the example of the the Gamma distribution, $G(\alpha, \beta)$, with mean $\alpha\beta$ and variance $\alpha\beta^2$. Here we can have two equations with two unknown parameters and the equations of moment estimators are to be found as follow. First, express the mean and variance as functions of the estimated AR parameters, for the mean:

$$\hat{\mu}_\epsilon = \frac{\hat{\mu}_x \left(1 - e^t \hat{\Sigma}^{-1}\hat{\sigma}\right)}{1 - \hat{\mu}_x^2 e^t \hat{\Sigma}^{-1} e} = \frac{\hat{\mu}_x (1 - C_m)}{1 - \hat{\mu}_x^2 S_m},$$

where $C_m = e^t \hat{\Sigma}^{-1}\hat{\sigma}$, $S_m = e^t \hat{\Sigma}^{-1} e$, and for the variance $\hat{\sigma}_\epsilon^2 = \hat{\sigma}_x^2 - \hat{\phi}^t \hat{\Sigma} \hat{\phi}$. The set these estimators equal to the corresponding theoretical expressions as follows:

$$\alpha\beta = \hat{\mu}_\epsilon$$
$$\alpha\beta^2 = \hat{\sigma}_\epsilon^2$$

and then solve for the distributional parameters. The calculations are straightforward and yield:

$$\begin{aligned}
\hat{\beta} &= \frac{\hat{\alpha}\hat{\beta}^2}{\hat{\alpha}\hat{\beta}} = \frac{\hat{\sigma}_\epsilon^2}{\hat{\mu}_\epsilon} \\
&= \frac{\hat{\sigma}_x^2 - D_m}{\hat{\mu}_\epsilon} + 2\hat{\mu}_x C_m - \hat{\mu}_\epsilon \hat{\mu}_x^2 S_m \\
&= \frac{\hat{\sigma}_x^2 - D_m}{\hat{\mu}_\epsilon} + \hat{\mu}_x C_m - \hat{\mu}_\epsilon + \hat{\mu}_x \\
&= \frac{\hat{\sigma}_x^2 - D_m}{\hat{\mu}_x} \left(\frac{1 - \hat{\mu}_x^2 D_m}{\hat{\mu}_x (1 - C_m)} \right) + \hat{\mu}_x (1 + C_m) - \frac{\hat{\mu}_x (1 - C_m)}{1 - \hat{\mu}_x^2 D_m}
\end{aligned}$$

where $D_m = \hat{\sigma}^t \hat{\Sigma}^{-1} \hat{\sigma}$ and then:

$$\begin{aligned}
\hat{\alpha} &= \frac{\hat{\mu}_\epsilon^2}{\hat{\sigma}_\epsilon^2} = \frac{\hat{\mu}_\epsilon^2}{\hat{\sigma}_x^2 - D_m + \hat{\mu}_\epsilon \hat{\mu}_x (1 + C_m) - \hat{\mu}_\epsilon^2} \\
&= \frac{\hat{\mu}_x^2 (1 - C_m)^2}{(\hat{\sigma}_x^2 - D_m)(1 - \hat{\mu}_x^2 S_m)^2 + \hat{\mu}_x^2 (1 - C_m^2)(1 - \hat{\mu}_x^2 S_m) - \hat{\mu}_x^2 (1 - C_m)^2}. \qquad (9)
\end{aligned}$$

In a similar fashion we can obtain distributional parameter estimators for other non-negative distributions. We continue our illustration using ϵ_t's that are distributed as log-normal, $LN(\mu, \sigma)$, or Weibull, $W(\gamma, \tau)$. For the log-normal case the estimators are quite straightforward and are given as:

$$\hat{\mu} = 2\log(\hat{\mu}_\epsilon) - \frac{1}{2}\log(\hat{\sigma}_\epsilon^2 + \hat{\mu}_\epsilon^2)$$

and

$$\begin{aligned}
\hat{\sigma} &= \log(\hat{\sigma}_\epsilon^2 + \hat{\mu}_\epsilon^2) - 2\log\hat{\mu}_\epsilon) \\
&= log\left(\left(\hat{\sigma}_x^2 - \hat{\phi}^t \hat{\Sigma} \hat{\phi} \right) \left(1 - \hat{\mu}_x^2 S_m\right)^2 + \hat{\mu}_x^2 (1 - C_m)^2 \right) - 2log\left(\hat{\mu}_x (1 - C_m) \right), \qquad (10)
\end{aligned}$$

while for the Weibull case we obtain a non-linear system, which can be solved numerically very easily. Specifically, we have:

$$\hat{\tau}\Gamma(1 + \frac{1}{\hat{\gamma}}) - \hat{\mu}_\epsilon = 0$$

and

$$\hat{\tau}^2 \left(\Gamma(1 + \frac{2}{\hat{\gamma}}) - \left(\Gamma(1 + \frac{2}{\hat{\gamma}}) \right)^2 \right) - \hat{\sigma}_\epsilon^2 = 0 \qquad (11)$$

for the mean and variance equations respectively, and since we have the presence of a distributional parameter inside the Gamma function we must solve the system numerically.

It is interesting to further consider the Generalized Lambda Distribution, GLD, which is extremely versatile in fitting a probability distribution to observed data. Freimer et al. [12] devise a parametrization for the GLD, denoted FMKL, which is given by

$$Q(u) = \lambda_1 + \frac{1}{\lambda_2}\left(\frac{u^{\lambda_1}-1}{\lambda_3} - \frac{(1-u)^{\lambda_4}-1}{\lambda_4}\right),$$

where λ_1 is the location parameter, λ_2 determines the scale parameter, λ_3 and λ_4 capture the shape characteristics of the empirical distribution generated by the data. Note, however, in order to have a finite k^{th} order moment, it is necessary that $min(\lambda_3, \lambda_4) > -\frac{1}{k}$. Given the GLD with quantile function Q(u), find parameters λ_1, λ_2, λ_3 and λ_4 so that the mean, μ_ϵ, variance, σ_ϵ^2, skewness

$$\begin{aligned}\alpha_{3,\epsilon} &= E\left(\epsilon_t - \mu_\epsilon\right)^2 = E\left((X_t-\mu_x) - \phi_1(X_{t-1}-\mu_x) - ... - \phi_p(X_{t-p}-\mu_x)\right)^3 \\ &= E\left(X_t-\mu_x\right)^3 - 3\sum_{j=1}^{p}\phi_j E\left((X_t-\mu_x)^2(X_{t-j}-\mu_x)\right) \\ &+ 3\sum_{j=1}^{p}\phi_j^2 E\left((X_t-\mu_x)(X_{t-j}-\mu_x)^2\right) - 3\sum\sum_{i\neq j}\phi_i^2\phi_j E\left((X_{t-i}-\mu_x)^2(X_{t-j}-\mu_x)\right) \\ &= \mu_{3,x} - 3\sum_{j=1}^{p}\phi_j\mu_{21,x}^{j} + 3\sum_{j=1}^{p}\phi_j^2\mu_{12,x}^{j} - 3\sum\sum_{i\neq j}\phi_i^2\phi_j\mu_{21,x}^{ij}\end{aligned}$$

and kurtosis,

$$\begin{aligned}\alpha_{4,\epsilon} &= E\left(\epsilon_t - \mu_\epsilon\right)^4 = E\left((X_t-\mu_x) - \phi_1(X_{t-1}-\mu_x) - ... - \phi_p(X_{t-p}-\mu_x)\right)^4 \\ &= E\left(X_t-\mu_x\right)^4 - 4\sum_{j=1}^{p}\phi_j E\left((X_t-\mu_x)^3(X_{t-j}-\mu_x)\right) + 6\sum_{j=1}^{p}\phi_j^2 E\left((X_t-\mu_x)^2(X_{t-j}-\mu_x)^2\right) \\ &+ 4\sum\sum_{i\neq j}\phi_i^3\phi_j E\left((X_{t-i}-\mu_x)^3(X_{t-j}-\mu_x)\right) + 6\sum\sum_{i\neq j}\phi_i^2\phi_j^2 E\left((X_{t-i}-\mu_x)^2(X_{t-j}-\mu_x)^2\right) \\ &= \mu_{4,x} - 4\sum_{j=1}^{p}\phi_j\mu_{31,x}^{j} + 6\sum_{j=1}^{p}\phi_j^2\mu_{22,x}^{j} + 4\sum\sum_{i\neq j}\phi_i^3\phi_j\mu_{31,x}^{ij} + 6\sum\sum_{i\neq j}\phi_i^2\phi_j^2\mu_{22,x}^{ij}\end{aligned}$$

of the GLD match the corresponding mean, $\hat{\mu}_\epsilon$, variance, $\hat{\sigma}_\epsilon^2$, skewness,

$$\hat{\alpha}_{3,\epsilon} = \hat{\mu}_{3,x} - 3\sum_{j=1}^{p}\hat{\phi}_j\hat{\mu}_{21,x}^{j} + 3\sum_{j=1}^{p}\hat{\phi}_j^2\hat{\mu}_{12,x}^{j} - 3\sum\sum_{i\neq j}\hat{\phi}_i^2\hat{\phi}_j\hat{\mu}_{21,x}^{ij}$$

and kurtosis,

$$\hat{\alpha}_{4,\epsilon} = \hat{\mu}_{4,x} - 4\sum_{j=1}^{p}\hat{\phi}_j\hat{\mu}_{31,x}^{j} + 6\sum_{j=1}^{p}\hat{\phi}_j^2\hat{\mu}_{22,x}^{j} + 4\sum\sum_{i\neq j}\hat{\phi}_i^3\hat{\phi}_j\hat{\mu}_{31,x}^{ij} + 6\sum\sum_{i\neq j}\hat{\phi}_i^2\hat{\phi}_j^2\hat{\mu}_{22,x}^{ij}$$

of the sample, where

$$\mu_{3,x} = E\left(X_t - \mu_x\right)^3,$$
$$\mu^j_{21,x} = E\left((X_t - \mu_x)^2(X_{t-j} - \mu_x)\right),$$
$$\mu^j_{12,x} = E\left((X_t - \mu_x)(X_{t-j} - \mu_x)^2\right),$$
$$\mu^{ij}_{21,x} = E\left((X_{t-i} - \mu_x)^2(X_{t-j} - \mu_x)\right),$$
$$\mu_{4,x} = E\left(X_t - \mu_x\right)^4,$$
$$\mu^j_{31,x} = E\left((X_t - \mu_x)^3(X_{t-j} - \mu_x)\right),$$
$$\mu^j_{22,x} = E\left((X_t - \mu_x)^2(X_{t-j} - \mu_x)^2\right),$$

$$\mu^{ij}_{31,x} = E\left((X_{t-i} - \mu_x)^3(X_{t-j} - \mu_x)\right),$$
$$\mu^{ij}_{22,x} = E\left((X_{t-i} - \mu_x)^2(X_{t-j} - \mu_x)\right)^2,$$
$$\hat{\mu}_{3,x} = \frac{1}{n}\sum\left(x_t - \mu_x\right)^3,$$
$$\hat{\mu}^j_{21,x} = \frac{1}{n}\left((x_t - \mu_x)^2(x_{t-j} - \mu_x)\right),$$
$$\hat{\mu}^j_{12,x} = \frac{1}{n}\left((x_t - \mu_x)(x_{t-j} - \mu_x)^2\right),$$
$$\hat{\mu}^{ij}_{21,x} = \frac{1}{n}\left((x_{t-i} - \mu_x)^2(x_{t-j} - \mu_x)\right),$$
$$\hat{\mu}_{4,x} = \frac{1}{n}\left(x_t - \mu_x\right)^4,$$
$$\hat{\mu}^j_{31,x} = \frac{1}{n}\left((x_t - \mu_x)^3(x_{t-j} - \mu_x)\right),$$
$$\hat{\mu}^j_{22,x} = \frac{1}{n}\left((x_t - \mu_x)^2(x_{t-j} - \mu_x)^2\right),$$
$$\hat{\mu}^{ij}_{31,x} = \frac{1}{n}\left((x_{t-i} - \mu_x)^3(x_{t-j} - \mu_x)\right),$$

and

$$\hat{\mu}^{ij}_{22,x} = \frac{1}{n}\left((x_{t-i} - \mu_x)^2(x_{t-j} - \mu_x)\right)^2.$$

The distributional parameters λ_3 and λ_4 can now be computed by solving the following system of nonlinear equations:

$$G_3(\lambda_3, \lambda_4) = \hat{\alpha}_{3,\epsilon},$$

$$G_4(\lambda_3, \lambda_4) = \hat{\alpha}_{4,\epsilon}, \tag{12}$$

where

$$G_3(\lambda_3, \lambda_4) = \frac{\upsilon_3 - 3\upsilon_1\upsilon_2 + 2\upsilon_1^3}{(\upsilon_2 - \upsilon_1^2)^{\frac{3}{2}}},$$

$$G_4(\lambda_3,\lambda_4)=\frac{\upsilon_4-4\upsilon_1\upsilon_3+6\upsilon_1^2\upsilon_2-3\upsilon_1^4}{\left(\upsilon_2-\upsilon_1^2\right)^2},$$
$$\upsilon_1=\frac{1}{\lambda_3(\lambda_3+1)}-\frac{1}{\lambda_4(\lambda_4+1)},$$
$$\upsilon_2=\frac{1}{\lambda_3^2(2\lambda_3+1)}+\frac{1}{\lambda_4^2(2\lambda_4+1)}-\frac{2}{\lambda_3\lambda_4}\beta(\lambda_3+1,\lambda_4+1),$$
$$\upsilon_3=\frac{1}{\lambda_3^3(3\lambda_3+1)}-\frac{1}{\lambda_4^3(3\lambda_4+1)}-\frac{3}{\lambda_3^2\lambda_4}\beta(2\lambda_3+1,\lambda_4+1)+\frac{3}{\lambda_3\lambda_4^2}\beta(\lambda_3+1,2\lambda_4+1),$$
$$\upsilon_4=\frac{1}{\lambda_3^4(4\lambda_3+1)}+\frac{1}{\lambda_4^4(4\lambda_4+1)}+\frac{6}{\lambda_3^2\lambda_4^2}\beta(2\lambda_3+1,2\lambda_4+1)$$
$$-\frac{4}{\lambda_3^3\lambda_4}\beta(3\lambda_3+1,\lambda_4+1)-\frac{4}{\lambda_3\lambda_4^3}\beta(\lambda_3+1,3\lambda_4+1),$$

and $\beta(.,.)$ is beta function. Finally, once the values for λ_3 and λ_4 are obtained, the remaining parameters are computed using the formulae:

$$\lambda_2=\frac{\sqrt{\upsilon_2-\upsilon_1^2}}{\hat{\sigma}_\epsilon},$$

$$\lambda_1=\hat{\mu}_\epsilon+\frac{1}{\lambda_2}\left(\frac{1}{\lambda_3+1}-\frac{1}{\lambda_4+1}\right). \tag{13}$$

Similarly, we can consider the estimation of distributional parameters on the case of our bivariate autoregression. The novelty here is that the distributional parameters will contain the correlation among the two series. Let us again take as an illustrative example the case of a bivariate Gamma distribution. We have that a random vector $(\epsilon_{1,t},\epsilon_{2,t})$ is distributed according to the bivariate Gamma distribution on R_+^2 with shape parameter α and scale parameter vector β if it's moment generating function is defined as:

$$\psi_{\alpha,\beta}(z)=E\left(exp\left(-\sum_{i=1}^{2}\epsilon_{i,t}z_i\right)\right)=\left(\beta(z)\right)^{-\alpha},$$

where $z=(z_1,z_2)$, $\alpha\geq 0$ and $\beta(z)=1+\beta_1z_1+\beta_2z_2+\beta_{12}z_1z_2$ with condition $\beta_1>0$, $\beta_2>0$, $\beta_{12}>0$ and $\beta_1\beta_2-\beta_{12}>0$. This condition ensures that $\psi_{\alpha,\beta}(z)$ is a moment generating function of a probability density defined as:

$$f(\epsilon_{1t},\epsilon_{2,t})=exp\left(-\frac{\beta_2\epsilon_{1,t}+\beta_1\epsilon_{2,t}}{\beta_{12}}\right)\frac{\epsilon_{1t}^{\alpha-1}\epsilon_{2t}^{\alpha-1}}{\beta_{12}^{\alpha}\Gamma(\alpha)}f(C\epsilon_{1t}\epsilon_{2t})I_{R_+^2}(\epsilon_{1t},\epsilon_{2t}),$$

where $C=\frac{\beta_1\beta_2-\beta_{12}}{\beta_{12}^2}$, $f(z)=\sum_{k=0}^{\infty}\frac{z^k}{k!\Gamma(\alpha+k)}$ and

$$I_{R_+^2}(\epsilon_{1t},\epsilon_{2t})=\begin{cases}1 & \epsilon_{1,t}>0,\epsilon_{2,t}>0\\ 0 & otherwise\end{cases}.$$

To proceed we have to match the mean, $E(\epsilon_i)=\alpha\beta_i$, variance, $var(\epsilon_i)=\alpha\beta_i^2$, $i=1,2$, covariance, $cov(\epsilon_1,\epsilon_2)=\alpha(\beta_1\beta_2-\beta_{1,2})$, and correlation, $\rho(\epsilon_1,\epsilon_2)=$

$\frac{\beta_1\beta_2-\beta_{12}}{\beta_1\beta_2}$, by their corresponding sample moments – which, again, are related to the AR parameters. Doing so we obtain:

$$\hat{\beta}_1 = \frac{\hat{\sigma}^2_{\epsilon_1}}{\hat{\mu}_{\epsilon_1}}$$

$$\hat{\beta}_2 = \frac{\hat{\sigma}^2_{\epsilon_2}}{\hat{\mu}_{\epsilon_2}}$$

$$\hat{\beta}_{12} = \left(1 - \frac{\hat{\sigma}_{\epsilon_1\epsilon_2}}{\hat{\sigma}_{\epsilon_1}\hat{\sigma}_{\epsilon_2}}\right)\frac{\hat{\sigma}^2_{\epsilon_1}\hat{\sigma}^2_{\epsilon_2}}{\hat{\mu}_{\epsilon_1}\hat{\mu}_{\epsilon_2}}$$

and

$$\hat{\alpha} = \frac{\hat{\mu}_{\epsilon_1}\hat{\mu}_{\epsilon_2}}{\hat{\sigma}_{\epsilon_1}\hat{\sigma}_{\epsilon_2}}, \tag{14}$$

where $\hat{\sigma}_{\epsilon_1\epsilon_2}$ is the corresponding sample covariance of $\sigma_{\epsilon_1\epsilon_2}$. Here, we obtain $\sigma_{\epsilon_1\epsilon_2}$ for model (6).

$$\begin{aligned}
\sigma_{\epsilon_1\epsilon_2} &= E\left(\epsilon_{1,t} - \mu_{\epsilon_1}\right)\left(\epsilon_{2,t} - \mu_{\epsilon_2}\right)\\
&= C_{12}(0) - \phi_{12}^{(1)}C_{11}(1) - \phi_{22}^{(1)}C_{12}(1) - \phi_{21}^{(2)}C_{11}(2) - \phi_{22}^{(2)}C_{12}(2)\\
&- \phi_{11}^{(1)}C_{12}(-1) + \phi_{11}^{(1)}\phi_{12}^{(1)}C_{11}(0) + \phi_{11}^{(1)}\phi_{22}^{(1)}C_{12}(0) + \phi_{11}^{(1)}\phi_{21}^{(2)}C_{11}(1) + \phi_{11}^{(1)}\phi_{22}^{(2)}C_{12}(1)\\
&- \phi_{12}^{(1)}C_{22}(-1) + \phi_{12}^{(1)}\phi_{12}^{(1)}C_{21}(0) + \phi_{12}^{(1)}\phi_{22}^{(1)}C_{22}(0) + \phi_{12}^{(1)}\phi_{21}^{(2)}C_{21}(1) + \phi_{12}^{(1)}\phi_{22}^{(2)}C_{22}(1)\\
&- \phi_{11}^{(2)}C_{12}(-2) + \phi_{11}^{(2)}\phi_{12}^{(1)}C_{11}(-1) + \phi_{11}^{(2)}\phi_{22}^{(1)}C_{12}(-1) + \phi_{11}^{(2)}\phi_{21}^{(2)}C_{11}(0) + \phi_{11}^{(2)}\phi_{22}^{(2)}C_{12}(0)\\
&- \phi_{12}^{(2)}C_{22}(-2) + \phi_{12}^{(2)}\phi_{12}^{(1)}C_{21}(-1) + \phi_{12}^{(2)}\phi_{22}^{(1)}C_{22}(-1) + \phi_{12}^{(2)}\phi_{21}^{(2)}C_{21}(0) + \phi_{12}^{(2)}\phi_{22}^{(2)}C_{22}(0)\\
&= \Phi_1\mathcal{C}\Phi_2^t
\end{aligned} \tag{15}$$

where:

$$\Phi_1 = \begin{pmatrix} 1 & \phi_{11}^{(1)} & \phi_{12}^{(2)} & \phi_{11}^{(2)} & \phi_{12}^{(1)} \end{pmatrix}, \quad \Phi_2 = \begin{pmatrix} 1 & \phi_{12}^{(1)} & \phi_{22}^{(1)} & \phi_{21}^{(2)} & \phi_{22}^{(2)} \end{pmatrix}$$

and

$$\mathcal{C} = \begin{pmatrix} C_{12}(0) & -C_{11}(1) & -C_{12}(1) & -C_{11}(2) & -C_{12}(2)\\ -C_{12}(1) & C_{11}(0) & C_{12}(0) & C_{11}(1) & C_{12}(1)\\ -C_{22}(-1) & C_{21}(0) & C_{22}(0) & C_{21}(1) & C_{22}(1)\\ -C_{12}(-2) & C_{11}(-1) & C_{12}(-1) & C_{11}(0) & C_{12}(0)\\ -C_{22}(-2) & C_{21}(-1) & C_{22}(-1) & C_{12}(0) & C_{22}(0) \end{pmatrix}.$$

We can perform similar calculations using the bivariate log-normal distribution. In this case we start off with the bivariate log-normal density:

$$f(\epsilon_1, \epsilon_2) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\tau^2}\epsilon_1\epsilon_2} exp\left(-\frac{A^2 + B^2 - 2\tau AB}{2(1-\tau^2)}\right),$$

where $A = \frac{ln\epsilon_1 - \mu_1}{\sigma_1}$ and $B = \frac{ln\epsilon_2 - \mu_2}{\sigma_2}$. Hence $\epsilon_{i,t}$'s distributed as Log-normal, $LN(\mu_i, \sigma_i)$. Solving as before for the mean and variances we obtain:

$$\hat{\mu}_i = 2\log\hat{\mu}_{\epsilon_i} - \frac{1}{2}\log(\hat{\sigma}^2_{\epsilon_i} + \hat{\mu}^2_{\epsilon_i})$$

$$\hat{\sigma}_i = \log(\hat{\sigma}^2_{\epsilon_i} + \hat{\mu}^2_{\epsilon_i})) - 2\log\hat{\mu}_{\epsilon_i}$$

and we still have to consider the covariance parameter:

$$\hat{\tau} = \frac{\hat{\sigma}_{\epsilon_1\epsilon_2}}{\hat{\sigma}_{\epsilon_1}\hat{\sigma}_{\epsilon_2}}. \tag{16}$$

3. Model selection

We extend our analysis to model selection. In principle, we should be needing anything new to consider traditional model selection criteria even when the distribution of the innovations is non-normal. However, this may not be necessarily true: in our case there are two dimensions that require order selection, the autoregressive order and, simultaneously, the kind of innovation distribution. Therefore, it might be of advantage to consider traditional model selection criteria along with new ones. The focussed information criterion, which we will be using here, suggests that an optimal model should depend on the parameter under focus, such as the mean, or the variance, or the particular covariance values, etc. Claeskens et al. [11] proposed an extension of the FIC for model-order selection with focus on high predictive accuracy. They assumed that the true time series model is an $AR(\infty)$ and is only approximated by selecting a finite order autoregressive model, AR(p), where $0 < p < p_n$ and the maximal considered AR-order, p_n, may depend on n. They also assumed that the innovations are independent and identically normally distributed, with mean 0 and variance σ^2_ϵ and the autoregressive coefficients ϕ_i's are absolutely sumable. They select the model that yields the best estimate for the focus parameter from the $p_n + 1$ possible AR(p)-models. Best is defined in the sense of having the lowest mean squared forecast error where the parameters are estimated using OLS.

In the following, we extend the FIC for autoregressive models with a constant term where the parameters are estimated using the method of moments under the assumptions of Section 2, that is the marginal distribution of the innovations is non-normal. Consider thus the autoregressive model in (2) where the innovation terms come from some density function f. For time series data $\{x_t\}$, let $x_t = \epsilon_t$ be the smallest model with density function $f(.;\theta)$ where θ is a q-vector of distributional parameters and the largest model is

$$x_t = \phi_0 + \phi_1 x_{t-1} + ... + \phi_{p_n} x_{t-p_n} + \epsilon_t$$

with density function $f(.;\theta,\Phi)$ where Φ is an additional $(p_n + 1)$-vector of parameters. For the smallest model, $\Phi = \Phi_0 = \mathbf{0}$ is fixed and known. For the largest model the parameters (θ, Φ), are estimated using the method of moments as discussed previously. Our goal is then to construct an information criterion aimed at selecting the model yielding the best estimates for the focused parameter from the $p_n + 1$ possible AR(p)-models. Because our goal is to make prediction as accurate as possible, we take as focused parameter as being $\mu_s = \Phi(p_s, h)x(p_s, h)$ where $\Phi(p_s, h) = (\phi_0, \phi_1, ..., \phi_{p_s})^t$,

$x(p_s, h) = (1, x_{n+h-1}, ..., x_{n+h-p_s})$. The results for the FIC apply in the local misspecification framework

$$f_{true} = f(., \theta, \Phi + \frac{\delta}{\sqrt{n}}),$$

where $\delta_1, ..., \delta_{p_n}$ parameters signify the degrees of the model departures in directions $1, ..., p_n$.

Lemma 3.1. *The moment estimator of μ_s in the s^{th} model has limiting distribution of the form:*

$$\sqrt{n}(\hat{\mu}_s - \mu_{true}) \xrightarrow{D} N(\mu_{p_s}, \sigma^2_{p_s}),$$

where

$$\mu_{p_s} = lim_{n\to\infty}\sqrt{n}(\Phi - \Phi_{true})x(p_n, h) = -lim_{n\to\infty}\delta^t x(p_n, h)$$
$$\sigma^2_{p_s} = x(p_s, h)^t R(p_s)^{-1} x(p_s) lim_{n\to\infty}\sigma^2_{p_n},$$

$\Phi = (\Phi(p_s, h), \Phi(p_{s^c}, h))^t$, $\Phi_{true} = \Phi + \frac{\delta}{\sqrt{n}}$, $\sigma^2_{p_n}$ *is variance of the innovations and*

$$R(p_s) = \begin{pmatrix} 1 & \mu_x & \dots & \mu_x \\ \mu_x & \sigma(0) & \dots & \sigma(p_s - 1) \\ \vdots & & & \\ \mu_x & \sigma(p_s - 1) & \dots & \sigma(0) \end{pmatrix}.$$

The limiting distribution has mean squared error:

$$r(p_s) = lim_{n\to\infty} x(p_n, h)^t \delta\delta^t x(p_n, h) + x(p_s, h)^t R(p_s)^{-1} x(p_s, h) lim_{n\to\infty}\sigma^2_{p_n}.$$

The FIC estimates this risk quantity for each proposed autoregressive model. To estimate $r(p_s)$, we estimate the unknown $\sigma^2_{p_n}$ and $R(p_s)$ by $\hat{\sigma}^2_{p_n}$ and

$$n^{-1}\begin{pmatrix} n & \sum x_{t-1} & \dots & \sum x_{t-p_s} \\ \sum x_{t-1} & \sum x_{t-1}x_{t-1} & \dots & \sum x_{t-1}x_{t-p_s} \\ \vdots & & & \\ \sum x_{t-p_s} & \sum x_{t-p_s}x_{t-1} & \dots & \sum x_{t-p_s}x_{t-p_s} \end{pmatrix}.$$

Also the quantity $\delta\delta^t$ is estimated by:

$$(\hat{\Phi} - \hat{\Phi}(p_n, h))(\hat{\Phi} - \hat{\Phi}(p_n, h))^t + \hat{R}_*(p_s)\hat{\sigma}^2_{p_n} - \hat{R}(p_n)\hat{\sigma}^2_{p_n}$$

where $\hat{\Phi} = (\hat{\Phi}(p_s, h), \Phi(p_{s^c,h}))^t$ and

$$R_*(p_s) = \begin{pmatrix} R(p_s) & 0 \\ 0 & 0 \end{pmatrix}.$$

So we have:

$$\hat{r}(p_s) = x(p_n, h)^t(\hat{\Phi} - \hat{\Phi}(p_n, h))(\hat{\Phi} - \hat{\Phi}(p_n, h))^t x(p_n, h)$$
$$+ 2x(p_s, h)^t R(p_s)^{-1} x(p_s, h)\sigma^2_{p_n} - x(p_n, h)^t R(p_n)^{-1} x(p_n, h)\sigma^2_{p_n}.$$

If we add $x(p_n,h)^t R(p_n)^{-1} x(p_n,h)\sigma^2_{p_n}$, which is independent of p, we arrive at the more compact expression for the FIC:

$$\begin{aligned} FIC &= x(p_n,h)^t(\hat{\Phi}-\hat{\Phi}(p_n,h))(\hat{\Phi}-\hat{\Phi}(p_n,h))^t x(p_n,h) \\ &\quad + 2x(p_s,h)^t R(p_s)^{-1} x(p_s,h)\sigma^2_{p_n}. \end{aligned} \tag{17}$$

Now we consider vector autoregressive model where the innovation terms are non-normal. In this models, the model selection are done in two step based on the FIC. In first step, for time series data $\mathbf{x}_t^{(i)} = (x_{1t},...,x_{it})^t, \quad i = 1,2,...,k$, let $\mathbf{x}_t^{(i)} = \epsilon_{\mathbf{t}}^{(\mathbf{i})}$ be the smallest model and

$$\mathbf{x}_t^{(i)} = \Phi_0 + \Phi_1 \mathbf{x}_{t-1}^{(i)} + ... + \Phi_{p_n}\mathbf{x}_{t-p_n}^{(i)} + \epsilon_{\mathbf{t}}^{(\mathbf{i})}$$

be the largest model where $\epsilon_{\mathbf{t}}^{(\mathbf{i})} = (\epsilon_{\mathbf{1t}} + ... + \epsilon_{\mathbf{it}})^{\mathbf{t}}$ and $\Phi_j = \begin{pmatrix} \phi_{11}^{(j)} & ... & \phi_{1i}^{(j)} \\ \vdots & & \\ \phi_{i1}^{(j)} & ... & \phi_{ii}^{(j)} \end{pmatrix}$ is a $i \times i$ matrix and $\Phi_0 = (\Phi_{10},...,\Phi_{i0})^t$. For $i = 1,...,k$, the focus information criteria is

$$\begin{aligned} FIC_{li} &= \mathbf{x}_i^t(p_n,h)(\hat{\Psi}^{(l)} - \hat{\Psi}^{(l)}(p_n,h))(\hat{\Psi}^{(l)} - \hat{\Psi}^{(l)}(p_n,h))^t \mathbf{x}_i(p_n,h) \\ &\quad + 2\mathbf{x}_i^t(p_s,h)R(p_s)^{-1}\mathbf{x}_i(p_s,h)\Sigma_{p_n}, \end{aligned}$$

where

$$\mathbf{x}_i^t(p_n,h) = (1, x_{1,n+h-1},...,x_{1,n+h-p_n},...,x_{i,n+h-1},...,x_{i,n+h-p_n})^t,$$

$$\Psi^{(l)}(p_n,h) = \left(\phi_{l0}^{(1)}, \phi_{l1}^{(1)},...,\phi_{li}^{(1)},...,\phi_{l1}^{(p_n)},...,\phi_{li}^{(p_n)}\right),$$

$$\Psi^{(l)} = \left(\phi_{l0}^{(1)}, \phi_{l1}^{(1)},...,\phi_{li}^{(1)},...,\phi_{l1}^{(p_s)},...,\phi_{li}^{(p_s)}, 0,...,0\right),$$

$$R^{(i)}(p_s) = \begin{pmatrix} 1 & \mu_x & \dots & \mu_x \\ \mu_x & \Gamma^{(i)}(0) & \dots & \Gamma^{(i)}(p_s-1) \\ \vdots & & & \\ \mu_x & \Gamma^{(i)}(p_s-1) & \dots & \Gamma^{(i)}(0) \end{pmatrix},$$

and

$$\Gamma^{(i)}(k) = E\left\{\mathbf{X}_t^{(i)}\mathbf{X}_t^{(i)^t}\right\}.$$

For $i = 1,...,k$, we select $p_s, 1 < p_s < p_n$ that it has lowliest value FIC_{li}. In second step For $l = 1,...,k$, the model is selected as optimal model that it's FIC is $\min_i FIC_{li}$.

4. **Simulation Analysis**

In this section, we examine by simulation the relative performance of the material presented so far in the paper. In particular, we examine more closely the performance of the method of moments approach for estimation and also the performance of the model selection criterion.

4.1. **Parameters Estimation.** Consider a first order autoregressive model. The observations from the autoregressive model are generated with $\phi = 0$ and $\phi = 0.5$ where the innovation terms are assumed to be identically and independently distributed as Gamma, G(2,2) and Weibull, W(2,2). Here we assume that the true model is known and only the parameters need to be estimated. The results for all estimation procedures and their mean square error are given for different sample sizes, of n=50, 100, 500, 1000 and are summarized in Table 1 that follows. In the table we present the average, across replications, estimates of the parameters (the distributional parameters and the autoregressive parameter) and also the corresponding mean squared error vis-a-vis the true parameters. The results are very interesting and provide a clear practical recommendation: if one knows the underlying innovation distribution then the modified MLE approach works best, whereas the MLE approach exhibits performance similar to the MME; in small samples the MME might be less accurate in estimating the distributional parameters but its accuracy grows with the sample size, as expected. If one is to choose between the three methods the MME would be ranked second. However, there is an important point not to be missed here: the modified MLE cannot be extended (not easily at least) to autoregressive models of order greater than one; the MME on the other hand has no such problems and is usable under any AR order. This is an advantage of the MME which should be considered in practice, when models of order higher than 1 are required.

The results in Table 1 are for the case of a correctly specified model, in terms of its innovation distribution. But what happens when we consider model with a misspecified distribution? Consider therefore that the true model is a first order autoregressive model with log-normal, LN(1,0.5) innovations with $\phi = 0.6$. We will ignore the true model and estimate models assuming a Weibull, Gamma and Normal distributions to examine the impact of misspecification on estimation methods. In Table 2, we report the results for the mean-squared error of the estimation of the one-step ahead prediction. We do this to illustrate the potential problems that will arise in a misspecified model when parameters that do not belong to the true model are estimated and then used to make predictions. The results are supportive of the conclusions in Table 1, that is that the MME becomes highly competitive to the modified MLE approach as the sample size increases and that the MLE approach is not really suited to the estimation of models with non-normal innovations.

TABLE 1. The estimation of the autoregressive and distributional parameters

	50		100		500		1000	
$\phi = 0$	G(2,2)	W(2,2)	G(2,2)	W(2,2)	G(2,2)	W(2,2)	G(2,2)	W(2,2)
MME								
$\hat{\alpha}_1$	2.5544	2.3308	2.2591	2.1625	2.0527	2.0286	2.0251	2.0184
MSE	(133.3093)	(39.4343)	(44.3280)	(12.2425)	(6.1632)	(1.6947)	(2.9708)	(0.8161)
$\hat{\alpha}_2$	1.8367	2.1830	1.9138	2.0978	1.9829	2.0208	1.9896	2.0128
MSE	(37.2942)	(15.9633)	(18.6295)	(6.5779)	(3.7242)	(1.1844)	(1.7654)	(0.5306)
$\hat{\phi}$	-0.0560	-0.0945	-0.0268	-0.0503	-0.0081	-0.0097	-0.0030	-0.0053
MSE	(2.3632)	(3.2307)	(1.1403)	(1.3074)	(0.1960)	(0.2189)	(0.1010)	(0.1011)
MLE								
$\hat{\alpha}_1$	1.4819	1.8327	1.5240	1.9041	1.7266	1.9667	1.8349	1.9807
MSE	(52.6420)	(19.5295)	(42.4959)	(7.9703)	(21.6020)	(1.7777)	(12.1586)	(1.0907)
$\hat{\alpha}_2$	2.2003	1.7485	2.2091	1.8630	2.1380	1.9638	2.0855	1.9775
MSE	(44.7582)	(19.3109)	(30.3388)	(8.3835)	(9.1159)	(1.9100)	(4.5295)	(1.2075)
$\hat{\phi}$	0.3546	0.1374	0.3292	0.0741	0.1707	0.0212	0.0961	0.0143
MSE	(20.9110)	(7.2629)	(19.9818)	(3.5291)	(10.2571)	(0.9108)	(5.5911)	(0.6302)
MMLE								
$\hat{\alpha}_1$	1.8174	1.8924	1.8081	1.931	1.8989	1.9667	1.9404	1.9766
MSE	(12.0693)	(3.6552)	(8.5106)	(1.4467)	(2.0341)	(0.2849)	(0.7439)	(0.1528)
$\hat{\alpha}_2$	1.9082	1.9259	2.0040	1.9644	1.9853	1.9897	1.9564	1.994
MSE	(14.3293)	(2.3548)	(7.4028)	(0.6757)	(1.1743)	(0.1007)	(0.5524)	(0.0467)
$\hat{\phi}$	0.1267	0.0388	0.0886	0.0183	0.0421	0.0074	0.0335	0.0051
MSE	(1.8449)	(0.8256)	(0.8950)	(0.2220)	(1.1743)	(0.0258)	(0.1225)	(0.0126)
	50		100		500		1000	
$\phi = 0.5$	G(2,2)	W(2,2)	G(2,2)	W(2,2)	G(2,2)	W(2,2)	G(2,2)	W(2,2)
MME								
$\hat{\alpha}_1$	3.9078	3.2155	2.8505	2.5238	2.1701	2.0982	2.0733	2.0519
MSE	(980.3710)	(342.0374)	(196.1582)	(62.3287)	(17.7324)	(4.7718)	(6.9471)	(2.1376)
$\hat{\alpha}_2$	1.7246	3.1552	1.8212	2.5586	1.956	2.1090	1.9810	2.0576
MSE	(54.4863)	(195.7433)	(26.4507)	(53.3745)	(5.3779)	(4.1151)	(2.5516)	(1.7618)
$\hat{\phi}$	0.3137	0.1960	0.4051	0.3568	0.4792	0.4721	0.4909	0.4855
MSE	(6.0401)	(13.2699)	(1.8244)	(3.3768)	(0.2067)	(0.2463)	(0.0868)	(0.0998)
MLE								
$\hat{\alpha}_1$	1.4184	1.7580	1.496	1.8540	1.6801	1.9632	1.7918	1.9791
MSE	(56.8922)	(22.3784)	(41.5087)	(10.7607)	(25.0036)	(1.5800)	(15.5799)	(0.8075)
$\hat{\alpha}_2$	2.0827	1.6714	2.0510	1.8106	2.0468	1.9579	2.0366	1.9770
MSE	(48.2775)	(26.8801)	(26.7853)	(13.5807)	(7.3110)	(1.9268)	(3.6749)	(0.9064)
$\hat{\phi}$	0.7648	0.5950	0.7348	0.5540	0.6376	0.5109	0.5837	0.5060
MSE	(10.0438)	(3.0466)	(9.2065)	(1.5736)	(5.2730)	(0.2228)	(3.1433)	(0.1046)
MMLE								
$\hat{\alpha}_1$	1.7551	1.8984	1.7816	1.9248	1.9292	1.9659	1.9479	1.9762
MSE	(14.6434)	(3.0126)	(9.3592)	(1.6088)	(0.9889)	(0.3129)	(0.4649)	(0.1509)
$\hat{\alpha}_2$	1.9108	1.9424	1.9570	1.9666	1.8873	1.9872	1.9139	1.9920
MSE	(14.2072)	(1.0639)	(7.2119)	(0.3947)	(1.6780)	(0.0834)	(0.8893)	(0.0398)
$\hat{\phi}$	0.5768	0.5147	0.5561	0.5095	0.5292	0.5036	0.5211	0.5026
MSE	(0.6474)	(0.1200)	(0.3424)	(0.0445)	(0.0929)	(0.0076)	(0.0493)	(0.0033)

1: $\hat{\alpha}_1$ is the first parameter of innovation distribution and $\hat{\alpha}_2$ is the second parameter of innovation distribution

4.2. **Order Selection.** We next present the results of a simulation study where we examine the performance of the FIC compared to AIC. The two most commonly used penalized time series model selection criteria, the FIC and AIC examine and compare. Their performance in estimating the quantities is computed. Despite their different foundations, some similarities between the amounts of the two statistics can be observed. Now, we consider the data

Table 2. The Impact of Misspecification in Estimation - values of predictive MSE

n	Model	MME	MLE	MMLE
50	Normal	344.3900	517.8836	567.6493
	Gamma	470.4266	49539.5813	503.4276
	Weibull	455.6002	701533.6373	450.5488
150	Normal	243.1524	222.2343	222.2343
	Gamma	290.8515	98198.7034	304.8580
	Weibull	251.6783	43729.4071	288.0663
250	Normal	287.9399	317.7273	317.7273
	Gamma	328.0730	54642.9172	388.4967
	Weibull	338.7549	720837.2269	348.8783
500	Normal	354.1267	376.6668	376.6668
	Gamma	390.0438	24075.5040	396.7021
	Weibull	350.8262	41910.6018	365.2664
1000	Normal	393.1417	392.8771	392.8771
	Gamma	397.4847	505618.1406	460.0824
	Weibull	441.0899	31079.0240	498.6948

generating model as

$$x_t = \phi_1 x_{t-1} + ... + \phi_4 x_{t-4} + \epsilon_t,$$

where the innovations are independent and identically distributed as Gamma, G(2,2), and Weibull, W(2,2), and the autoregressive parameters take the values of (0.7,0.2,-0.5,-0.1). Recall that we estimate the parameters using the method of moments. we generate series of lengths $n = 50, 150, 250, 500, 1000$, which we use for both model order selection and parameter estimation. Then, for each of the M = 1000 simulation runs, we choose a maximal order $p_n = 4$ and select the optimal order by AIC and FIC. Table 3 shows the relative frequency of correct order selection for both criteria.

As expected, when the sample size increases the relative frequency of both criteria also increases. We observe however that the relative frequency of AIC is always smaller than the relative frequency FIC.

Table 3. Relative Frequency of Order Selection-FIC and AIC

n	Model	FIC	AIC
50	Gamma	0.7410	0.4014
	Weibull	0.7590	0.4210
150	Gamma	0.8130	0.5400
	Weibull	0.8040	0.7750
250	Gamma	0.8360	0.6660
	Weibull	0.8350	0.7864
500	Gamma	0.9390	0.7870
	Weibull	0.8980	0.8620
1000	Gamma	0.9720	0.9160
	Weibull	0.9480	0.9050

Now consider an autoregressive moving average model, ARMA(1,1),

$$x_t = \phi_1 x_{t-1} + \epsilon_t + \eta_1 \epsilon_{t-1}$$

as the true model, where ϵ_t's are independent and identically distributed as G(2,2) and both ϕ and η take values in $\{0, 0.1, ..., 0.9\}$. The stationarity and invertibility conditions on the parameters in this model reduce to $\phi < 1$ and $\eta < 1$. The ARMA(1,1) has an $AR(\infty)$ representation. Hence, we select the optimal model among the candidate autoregressive models.
We generate series $\{x_t\}$ of length n+h which we use $x_t \;\; t = 1, ..., n$ for both parameter estimation and model order selection and $x_t \;\; t = n+1, ..., n+h$ to estimation of the prediction accuracy of the h-step ahead forecast of $\{x_t\}$. For i^{th} simulation run, we select the optimal order, p_i, and compute the h-step ahead forecast value. Define the h-step ahead forecast as

$$\hat{x}^{(i)}_{n+h} = \hat{\phi}_0 + \hat{\phi}_1 x_{n+h-1} + ... + \hat{\phi}_{p_i} x_{n+h-p_i}$$

and the mean squared error, MSE, of the h-step ahead prediction of the series $\{x_t\}$ as

$$MSE = \frac{1}{M} \sum_{i=1}^{n} \left(\hat{x}^{(i)}_{n+h} - x^{(i)}_{n+h} \right)^2 .$$

We choose the maximal order $p_n = 14$ and $h = 1$ here. The values of mean squared error of the h-step prediction, where the prediction is performed using the selected models by AIC and FIC are given in Table 4. The results in the table broadly support our earlier results in Table 4, that is, the FIC produces more frequently the anticipated correct results than the AIC, the two criteria converge in large samples but also note that the FIC is less sensitive to higher values in the parameters while the AIC is.

Table 4. The values of MSE of the h-step prediction of FIC and AIC

		FIC				
n	η/ϕ	0	0.2	0.5	0.7	0.9
250	0	18.1289	16.8586	17.9867	19.1204	30.3211
	0.2	17.21678	18.2691	18.9431	23.2991	40.7544
	0.5	17.8127	19.0254	21.0825	30.3303	60.3201
	0.7	18. 7237	20.5237	24.5411	39.2334	75.0953
	0.9	21.5743	24.2550	30.7749	50.7959	99.1254
1000	0	16.3969	16.5251	16.4303	18.1744	29.5830
	0.2	16.4195	16.5566	16.5715	18.5829	35.4526
	0.5	16.7074	17.1094	17.5742	20.7895	54.2658
	0.7	17.1750	17.7521	18.9475	23.6494	73.7971
	0.9	18.8841	19.7040	22.4179	28.9334	50.7091
		AIC				
250	0	18.3378	16.9949	18.0002	19.8161	33.0101
	0.2	17.2693	18.4164	18.9974	23.5140	42.3378
	0.5	17.6687	19.0625	21.4720	31.7102	65.9157
	0.7	18.7905	20.98380	25.2788	41.4197	105.4085
	0.9	22.2617	24.6024	32.3855	55.4085	70.5387
1000	0	16.4460	16.7363	16.6048	18.3777	28.0392
	0.2	16.4790	16.7928	16.8876	19.1919	40.9344
	0.5	16.7486	17.15773	17.8400	21.4454	73.9933
	0.7	17.2100	17.9437	19.4258	24.8616	81.2232
	0.9	18.7553	23.0707	25.3354	31.0527	60.1722

5. **A real-data example**

In this section, we consider an example using real data. Our datasets consist of the Europe oil prices, Brent and American stock market index based on the market capitalizations of 500 large companies having common stock listed on the NYSE or NASDAQ , $S\&P500$. The Brent and $S\&P500$ datasets consist of the daily returns with the sample extending from May 1992 to December 2020 for a total of n = 7467 observations. These data can be found in

$$“https://fred.stlouisfed.org/series/DCOILBRENTEU”$$

and

$$“http://finance.yahoo.com/quote/5EGSPC/components?p = 5EGSPC”$$

respectively. For dataset $S\&P500$, consider

$$R.S\&P500 = 100(\nabla(log(s\&p500))),$$

where $\nabla(x)$ denotes the first order differences operator applied to a time series $\{x\}$,$\nabla(x) = x_t - x_{t-1}$. The associated datasets of Brent and $R.S\&P500$ were constructed by summing daily squared returns. Specially, if we denote by $r_{t,i}$ the i^{th} daily return for month then the monthly dataset is denoted as

$$Data \xrightarrow{def} \left(\Sigma_{i=1}^{m}(r_{t,i} - \mu_t)^2\right)^{\frac{1}{2}}, \tag{18}$$

where m is the number of days and μ_t is mean of months. So, we define two groups of data, group1={VB} and group2={VSP}, by substituting the series Brent and $R.S\&P500$ in Definition 18 , respectively. The datasets in group 1 describe the information of growth in oil price. The information of economic growth is in group 2.

Descriptive statistics of the returns for all two of our datasets are given in Table 5. All series have unconditional means that are statistically different from zero. Also, VSP and VB have positive skewness. Finally, all series are characterized by heavy tails since they have positive the sample excess kurtosis. The hypothesis of normality is strongly rejected for all series where $P-value < 0.05$.

Table 5. Descriptive Statistics for Empirical Series

series	n	$\bar{x}$	$\hat{\sigma}$	$\mathcal{S}$	$\mathcal{K}$	$\mathcal{P}$
VSP	92	7.9679	19.1491	2.7922	11.7045	2.754e-11
VB	92	23.8491	579.1778	2.1900	5.3055	2.213e-11

Notes:
1. n denotes the number of observations, $\bar{x}$ denotes the sample mean, $\hat{\sigma}$ denotes the sample standard deviation, $\mathcal{S}$ denotes the sample skewness, $\mathcal{K}$ denotes the sample excess kurtosis.
2. $\mathcal{P}$ is the p-value of the Shapiro test for normality of the underlying series.

In Figures 1 to 2 we present graph of the return series in two defined group. Figure 1 is for the associated dataset of Brent and Figure 2 is for the associated dataset of $S\&P500$ data.

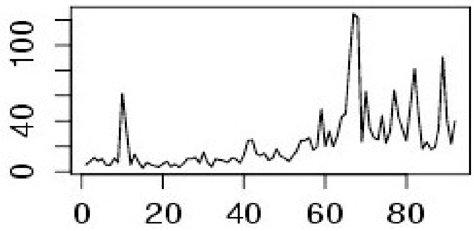

FIGURE 1. The curve of the VB series.

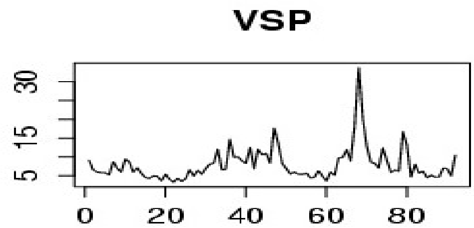

FIGURE 2. The curve of the VSP series.

The rolling correlation analysis is performed for correlations between variable in group 2 and variable in group 1. The rolling correlations are computed for R=80. The monthly rolling correlation analysis are reported in Table 6. All results confirm that variable VB is related to VSP.

The monthly rolling causality for the first order autoregressive model of series are considered. The results are given in Table 7. In the this Table, the significant causality at the 5% level is observed. In otherworld VB does Granger-cause VSP but the variables VSP do not Granger-cause VB.

To work with a series that has positive values and can be modelled with the methods presented earlier in the paper, we consider VSP. The sample autocorrelation and partial sample autocorrelation functions, Figure 3, suggests that a low-order autoregressive model might provide a reasonable description for the data. We will work with a first order model and the results are as follows.

Table 6. The Rolling Correlation

Time	VB-VSP
2016-12-31	0.5736
2017-03-31	0.5559
2017-06-30	0.5330
2017-09-30	0.5214
2017-12-31	0.5199
2018-09-30	0.5160
2018-12-31	0.5135
2019-06-30	0.5166
2019-09-30	0.5096
2019-12-31	0.4777
2020-03-31	0.4759
2020-06-30	0.4754
2020-12-31	0.4754

Table 7. The Rolling Causality Testing when AR(1) is Fitting

	VSP does not GC	VB no IC between VSP	VB does not GC VSP
2016-12-31	0.1180	0.0028	4.9224 e-05
2017-03-31	0.2050	0.0109	4.8085 e-05
2017-06-30	0.1001	0.0158	4.6842 e-05
2017-09-30	0.1565	0.0092	4.0362 e-04
2017-12-31	0.2166	0.0077	5.5153 e-04
2018-09-30	0.2065	0.0084	5.9767 e-04
2018-12-31	0.2177	0.0070	5.1945 e-04
2019-06-30	0.2199	0.0069	5.9159 e-04
2019-09-30	0.1993	0.0080	6.5899 e-04
2019-12-31	0.1043	0.0190	4.8215 e-04
2020-03-31	0.1890	0.0126	1.1705 e-03
2020-06-30	0.2044	0.0133	1.4410 e-03
2020-12-31	0.1841	0.0110	1.3467 e-03

The value of the estimated autoregressive coefficient based on the method of moments is 0.5614, so the residuals are computed by $\epsilon_t = x_t - 0.5614x_{t1}$. Note that we are not using a constant term because the mean of the residuals will be used for distributional fitting. Therefore, since all of observation are non-negative, we can use one of the distributions discussed earlier to see how it fits. Consider autoregressive model with Gamma, Log-Normal, Normal innovations

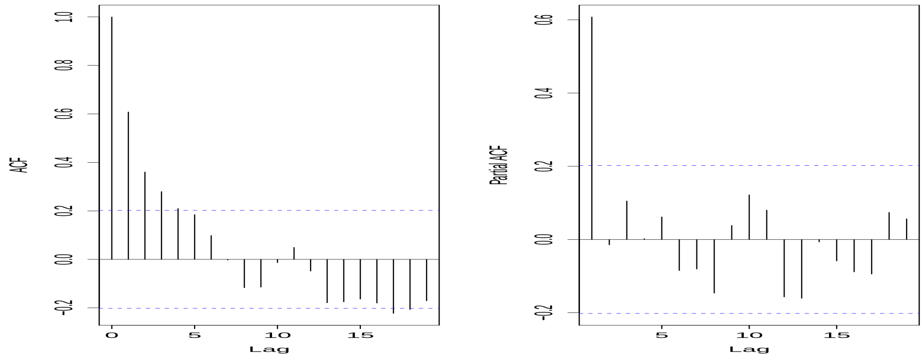

FIGURE 3. The sample autocorrelation function and partial autocorrelation function of VSP data set .

and vector autoregressive model with bivariate Log-Normal, LN-LN, bivariate Log-Normal and Normal, LN-N, and bivariate Normal, N-N, as competing models. Table 8 shows the estimated value of parameters (autoregressive and distributional), of the AIC and of the p-value of the Kolmogorov-Smirnov test. The estimated value of parameters of vector autoregressive and of the AIC are given in Table 9. Because the first order autoregressive with Log-Normal innovation has least value of AIC, 696.5973, then the first order autoregressive with Log-Normal distribution is selected as a suitable model for the innovations. The Kolmogorov-Smirnov test confirms this result. The p-value of the Kolmogorov-Smirnov test of autoregressive model with Log-Normal distribution is larger than 0.05.

Table 8. Distributional Fitting of the VSP series

		$\hat{\phi}$	$\hat{\alpha}_1$	$\hat{\alpha}_2$	AIC	K.S
AR(1)	Normal	0.5614	3.4971	12.7059	3913.6720	1.1489e-12
	Gammma	0.5614	0.9625	3.6332	1152.5410	1.1102e-16
	Log-Normal	0.5614	0.8957	0.7124	696.5973	0.6372
AR(2)	Normal	$\begin{pmatrix} 0.5855 \\ -0.0430 \end{pmatrix}$	3.6476	12.7260	3862.4380	7.8714e-13
	Gammma	$\begin{pmatrix} 0.5855 \\ -0.0430 \end{pmatrix}$	1.0455	3.4888	1146.2160	3.3306e-16
	Log-Normal	$\begin{pmatrix} 0.5855 \\ -0.0430 \end{pmatrix}$	0.9585	0.6711	697.3780	0.6549
AR(3)	Normal	$\begin{pmatrix} 0.5870 \\ -0.0626 \\ 0.0335 \end{pmatrix}$	3.5252	12.6809	3894.8000	1.9425e-12
	Gammma	$\begin{pmatrix} 0.5870 \\ -0.0626 \\ 0.0335 \end{pmatrix}$	0.9800	3.5971	1155.0950	3.3306e-16
	Log-Normal	$\begin{pmatrix} 0.5870 \\ -0.0626 \\ 0.0335 \end{pmatrix}$	0.9082	0.7032	700.4134	0.7167
AR(4)	Normal	$\begin{pmatrix} 0.5877 \\ -0.0640 \\ 0.0468 \\ -0.0226 \end{pmatrix}$	3.6051	12.7098	3876.5150	1.9022e-12
	Gammma	$\begin{pmatrix} 0.5877 \\ -0.0640 \\ 0.0468 \\ -0.0226 \end{pmatrix}$	1.0226	3.5254	1152.6090	6.6613e-16
	Log-Normal	$\begin{pmatrix} 0.5877 \\ -0.0640 \\ 0.0468 \\ -0.0226 \end{pmatrix}$	0.9413	0.6820	701.7065	0.7521
AR(5)	Normal	$\begin{pmatrix} 0.5876 \\ -0.0639 \\ 0.04663 \\ -0.0206 \\ -0.0034 \end{pmatrix}$	3.6177	12.7139	3875.2100	2.4281e-12 1
	Gammma	$\begin{pmatrix} 0.5876 \\ -0.0639 \\ 0.04663 \\ -0.0206 \\ -0.0034 \end{pmatrix}$	1.0294	3.5143	1153.9030	1.1102e-16
	Log-Normal	$\begin{pmatrix} 0.5876 \\ -0.0639 \\ 0.04663 \\ -0.0206 \\ -0.0034 \end{pmatrix}$	0.9464	0.6787	703.6077	0.6612

1: $\hat{\alpha}_1$ is the first parameter of innovation distribution and $\hat{\alpha}_2$ is the second parameter of innovation distribution

TABLE 9. Distributional Fitting of the VSP series

	model	$\hat{\phi}$	$\hat{\alpha}_1$	$\hat{\alpha}_2$	AIC
VAR(1)	N-N	$\begin{pmatrix} 0.5571 & 1.1733 \\ -0.0131 & 0.5223 \end{pmatrix}$	$\begin{pmatrix} 3.8468 \\ 2.1466 \end{pmatrix}$	$\begin{pmatrix} 18.7390 & 0.4784 \\ 0.4784 & 569.5636 \end{pmatrix}$	8013.7250
	LN-LN	$\begin{pmatrix} 0.5571 & 1.1733 \\ -0.0131 & 0.5223 \end{pmatrix}$	$\begin{pmatrix} 1.9919 \\ 0.2571 \end{pmatrix}$	$\begin{pmatrix} 0.8181 & 0.4784 \\ 0.4784 & 4.8251 \end{pmatrix}$	3037.0580
	LN-N	$\begin{pmatrix} 0.5571 & 1.1733 \\ -0.0131 & 0.5223 \end{pmatrix}$	$\begin{pmatrix} 1.9919 \\ 0.8181 \end{pmatrix}$	$\begin{pmatrix} 2.1466 & 0.4784 \\ 0.4784 & 569.5636 \end{pmatrix}$	2821.9640
VAR(2)	N-N	$\begin{pmatrix} 0.0990 & 0.1818 \\ -0.0588 & 0.1954 \\ 0.0942 & 0.1696 \\ 0.1106 & 0.3622 \end{pmatrix}$	$\begin{pmatrix} 5.2211 \\ 8.1987 \end{pmatrix}$	$\begin{pmatrix} 18.7523 & 0.4782 \\ 0.4782 & 569.5634 \end{pmatrix}$	7740.5530
	LN-LN	$\begin{pmatrix} 0.0990 & 0.1818 \\ -0.0588 & 0.1954 \\ 0.0942 & 0.1696 \\ 0.1106 & 0.3622 \end{pmatrix}$	$\begin{pmatrix} 2.5396 \\ 2.9166 \end{pmatrix}$	$\begin{pmatrix} 0.5234 & 0.4782 \\ 0.4782 & 2.2484 \end{pmatrix}$	3019.3750
	LN-N	$\begin{pmatrix} 0.0990 & 0.1818 \\ -0.0588 & 0.1954 \\ 0.0942 & 0.1696 \\ 0.1106 & 0.3622 \end{pmatrix}$	$\begin{pmatrix} 2.5396 \\ 0.5234 \end{pmatrix}$	$\begin{pmatrix} 8.1987 & 0.4784 \\ 0.4784 & 569.5634 \end{pmatrix}$	2745.5050
VAR(3)	N-N	$\begin{pmatrix} 0.0273 & -0.0289 \\ -0.0268 & 0.2630 \\ 0.0260 & -0.0299 \\ 0.0255 & 0.0951 \\ 0.0204 & -0.0154 \\ 0.0447 & 0.02536 \end{pmatrix}$	$\begin{pmatrix} 6.4517 \\ 15.8800 \end{pmatrix}$	$\begin{pmatrix} 8.7529 & 0.4782 \\ 0.4782 & 569.5638 \end{pmatrix}$	7583.8210
	LN-LN	$\begin{pmatrix} 0.0273 & -0.0289 \\ -0.0268 & 0.2630 \\ 0.0260 & -0.0299 \\ 0.0255 & 0.0951 \\ 0.0204 & -0.0154 \\ 0.0447 & 0.02536 \end{pmatrix}$	$\begin{pmatrix} 2.9085 \\ 4.1878 \end{pmatrix}$	$\begin{pmatrix} 0.3719 & 0.4782 \\ 0.4782 & 1.1812 \end{pmatrix}$	3156.6950
	LN-N	$\begin{pmatrix} 0.0273 & -0.0289 \\ -0.0268 & 0.2630 \\ 0.0260 & -0.0299 \\ 0.0255 & 0.0951 \\ 0.0204 & -0.0154 \\ 0.0447 & 0.02536 \end{pmatrix}$	$\begin{pmatrix} 2.9085 \\ 0.3719 \end{pmatrix}$	$\begin{pmatrix} 15.8800 & 0.4784 \\ 0.4784 & 569.5638 \end{pmatrix}$	3148.5010
VAR(4)	N-N	$\begin{pmatrix} 0.0016 & -0.1369 \\ -0.0038 & 0.3663 \\ 0.0015 & -0.1322 \\ 0.0044 & -0.0085 \\ 0.0013 & -0.0950 \\ 0.0044 & -0.1540 \\ 0.0013 & -0.0873 \\ 0.0040 & -0.0946 \end{pmatrix}$	$\begin{pmatrix} 8.7798 \\ 25.7318 \end{pmatrix}$	$\begin{pmatrix} 18.7529 & 0.4782 \\ 0.4782 & 569.5641 \end{pmatrix}$	7284.932
	LN-LN	$\begin{pmatrix} 0.0016 & -0.1369 \\ -0.0038 & 0.3663 \\ 0.0015 & -0.1322 \\ 0.0044 & -0.0085 \\ 0.0013 & -0.0950 \\ 0.0044 & -0.1540 \\ 0.0013 & -0.0873 \\ 0.0040 & -0.0946 \end{pmatrix}$	$\begin{pmatrix} 3.4323 \\ 5.0722 \end{pmatrix}$	$\begin{pmatrix} 0.2177 & 0.4782 \\ 0.4782 & 0.6206 \end{pmatrix}$	4932.952
	LN-N	$\begin{pmatrix} 0.0016 & -0.1369 \\ -0.0038 & 0.3663 \\ 0.0015 & -0.1322 \\ 0.0044 & -0.0085 \\ 0.0013 & -0.0950 \\ 0.0044 & -0.1540 \\ 0.0013 & -0.0873 \\ 0.0040 & -0.0946 \end{pmatrix}$	$\begin{pmatrix} 3.4323 \\ 0.2177 \end{pmatrix}$	$\begin{pmatrix} 25.7318 & 0.4784 \\ 0.4784 & 569.5641 \end{pmatrix}$	3252.095

1: $\hat{\alpha}_1$ is the first parameter of innovation distribution and $\hat{\alpha}_2$ is the second parameter of innovation distribution

Next we set the maximal order of the autoregressive model with Log-Normal residuals equal to $p_n = 4$ and compute the FIC. The values are given in Table 10. The FIC identifies a first order autoregressive model, a non-surprising result as the FIC can (as we already demonstrated in the simulations) provide a better model. Also the values of FIC of the vector autoregressive model are computed and given in Table 10. In this case, The FIC select a first order vector autoregressive model a better model.

Table 10. The values of FIC

model	p=1	p=2	p=3	p=4
AR	24.7927	30.0477	42.9851	95.5676
VAR	69.25248	355.6759	400.1593	238.4142

The rolling forecasts of proposed models are computed and the Mean Squared Error, MSE, and Mean Absolute Error, MAE, are given in Table 11. The presented results of 80 rolling prediction show that the first-order autoregressive model has least MSE and MAE.

Table 11. The values of MSE and MAE of Competing models

	model	p=1		p=2		p=3		p=4	
		MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
AR	Normal	6.1328	5.6553	5.8433	5.3344	5.9733	5.4882	5.7715	5.2872
	Gamma	2.5893	2.1059	2.5740	2.1711	2.5320	2.1330	2.5022	2.1471
	Log-Normal	2.0164	2.0988	2.0521	2.1661	2.0665	2.1392	2.0724	2.1234
VAR	LN-LN	16.3824	11.5185	24.9598	23.3628	17.0413	13.3752	17.4496	13.6709
	LN-N	22.2171	18.2381	24.5332	22.6790	18.0111	13.5896	19.1299	16.8423
	N-N	39.9340	34.5262	27.6216	26.3851	25.3044	23.4468	23.2369	21.4277

6. Concluding remarks

In this paper we make a number of contributions to the literature that relates to autoregressive models with non-normal innovation terms. First, we propose a method of moments estimation approach for both univariate and bivariate series that have non-normal innovations and show how the estimators can be obtained for any autoregressive order; this is important because going beyond the first order model we cannot easily obtain or compute maximum likelihood or modified maximum likelihood estimators. The method of moments estimators are show to be consistent and asymptotically normally distributed. Second, we show how one can estimate the distributional parameters of the obtained non-normal innovations again with the method of moments and for a variety of example distributions. We provide distributional estimator for both the univariate and bivariate cases. Third, we propose model selection using the FIC and explain why it might be a better model selection criterion in the cases examined in the paper. Fourth, we examine the properties of the estimators and model selection criteria using simulations which validate the earlier theoretical results. In summary, the theoretical derivations and the simulations support the use of the method of moments estimation in larger samples and the use of the FIC as a final order and model selection criterion. In cases, therefore, that a researcher is faced with time series data that have non-normal innovations the methods presented in this paper should be of immediate use. We leave for future research empirical applications that can further illustrate the use of the presented methods.

Appendix.

Proof of Lemma 3.1. For h-step ahead prediction

$$\begin{aligned}\sqrt{n}\,(\hat{\mu}_s - \mu_{true}) &= \sqrt{n}\,(\hat{\mu}_s - \mu_s + \mu_s - \mu_{true}) \\ &= \sqrt{n}\left(\hat{\Phi}(p_s,h) - \Phi(p_s,h)\right)^t x(p_s,h) \\ &\quad + \sqrt{n}\left(\Phi(p_s,h)^t x(p_s,h) - \Phi^t_{true} x(p_n,h)\right) \qquad (19)\end{aligned}$$

For the first term (19), using Slutsky's Theorem, we have asymptotic Normal distribution $N(0, x(p_s,h)^t R(p_s)^{-1} x(p_s,h)\sigma^2_\epsilon(p_s))$. Also, for the second term (19), we have

$$\begin{aligned}\sqrt{n}\left(\Phi(p_s,h)^t x(p_s,h) - \Phi^t_{true} x(p_n,h)\right) &= \sqrt{n}\left((\Phi(p_s,h), \Phi(p_{s^c},h))^t - \Phi^t_{true}\right) x(p_n,h) \\ &= \sqrt{n}(\Phi - \Phi_{true})^t x(p_n,h) \\ &= -\delta^t x(p_n,h).\end{aligned}$$

Then

$$\begin{aligned}\sqrt{n}\,(\hat{\mu}_s - \mu_{true}) &\xrightarrow{D} N(-lim_{n\to\infty}\delta^t x(p_n,h), x(p_s,h)^t R(p_s)^{-1} x(p_s) lim_{n\to\infty}\sigma^2_{p_n} \\ &= N(\mu_{p_s}, \sigma^2_{p_s})\end{aligned}$$

where $\mu_{p_s} = -lim_{n\to\infty}\delta^t x(p_n,h)$ and $\sigma^2_{p_s} = x(p_s,h)^t R(p_s)^{-1} x(p_s) lim_{n\to\infty}\sigma^2_{p_n}$. □

References

[1] H. Akaike, *Fitting autoregressive models for prediction*, Ann. Inst. Stat. Math 21: 243-247, (1969).

[2] H. Akaike, *Information theory and an extension of the maximum likelihood principle*, in the Second International Symposium on Information Theory, eds. by B.N. Petrov and F. Csake, Akademiai Kiado, Hungary, 267-281, (1973).

[3] A.D. Akkaya, and M.L. Tiku, *Estimating parameters in autoregressive models in non-normal situations: asymmetric innovations*, Communications in Statistics Theory and Methods 30: 517-536, (2001a).

[4] A.D. Akkaya, and M.L. Tiku, *Corrigendum: time series models with asymmetric innovations*, Communications in Statistics Theory and Methods 30 (10): 2227-2230, (2001b).

[5] A.D. Akkaya, and M.L. Tiku, *Time series AR(1) model for short tail distribution*, Statistics 39 (2): 117-132, (2005).

[6] N. Balakrishna, *Non-Gaussian autoregressive - type time Series*, Springer Nature (2022).

[7] O.T. Bayrak, and A.D. Akkaya, *Estimating parameters of a multiple autoregressive model by the modified maximum likelihood method*, Journal of Computational and Applied Mathematics 233: 1763-1772, (2010).

[8] P. Bondon, *Estimation of autoregressive models with epsilon-skew-normal innovations*, Journal of Multivariate Analysis 100, 1761-1776, (2009).

[9] G. E. P. Box and G. M. Jenkins, *Time series analysis forecasting and control*, 2nd ed., Holden-Day, San Franscisco, (1976).

[10] G. Claeskens, and N. L. Hjort, *The focused information criterion with discussion*, Journal of the American Statistical Association 98: 900-916, (2003).

[11] G. Claeskens, C. Croux, and J. Van Kerckhoven, *"Prediction focussed model selection for autoregressive models*, Australian & New Zealand Journal of Statistics 49 (4): 359-379, (2007).

[12] M. Freimer, G.S. Mudholkar, G. Kollia, and C.T. Lin, *A study of the generalized Tukey lambda family*, Communications in Statistics: Theory and Methods 17 (10): 3547-3567, (1988).

[13] D. P. Gaver and P. A. Lewis, *First order autoregressive Gamma sequences and point processes*, Adv. Appl. Prob 2 : 727-745, (1980).

[14] C. Gourieroux, and J. Jasiak, *Autoregressive gamma processes*, Journal of forecasting 25(2), 129 -152, (2006).

[15] S. Kullback and R. A. Leibler, *Information and sufficiency modelling*, Annals of Mathematical Statistics 22: 79-86, (1951).

[16] R. Shibata, *Approximate efficiency of a selection procedure for the number of regression variables*, Biometrika 71 (1): 43-49, (1984).

[17] A. G. W. Steyn, *Estimating parameters in an autoregressive process by combining the methods of moments and of least squares*, South African Statistical Journal 14(2), 103 -120, (1980).

[18] M. L. Tiku, *Estimating the mean and standard deviation from a censored Normal sample*, Biometrica 54 : 155-165, (1967).

[19] G. Walker, *On periodicity in series of related terms*, Proceedings of the Royal Society of London, Ser. A. 131: 518-532, (1931).

[20] G. Yule, *On a method of investigating periodicities in disturbed series with special reference to wolfer's sunspot numbers*, Philosophical Transactions of the Royal Society of London, Ser. A. 226: 267-298, (1927).

Sedigheh Zamani Mehreyan
Orcid number: 0000-0001-9953-8203
Department of Statistics
Imam Khomeini International University
Qazvin, Iran
Email address: zamani@sci.ikiu.ac.ir

Abdolreza Sayyareh
Orcid number: 0000-0002-8585-9900
Department of Computer Science and Statistics, Faculty of Mathematics
K.N. Toosi University of Technology
Tehran, Iran
Email address: asayyareh@kntu.ac.ir