MATHEMATICA APPLICANDA
Vol. 51(1) 2023, p. 121–141
doi: 10.14708/ma.v51i1.7203

Hubert Woszczek (Wrocław)
Agnieszka Wyłomańska (Wrocław)

Autoregressive model with double Pareto distributed noise

Abstract Time series models are a popular tool commonly used to describe time-varying phenomena. One of the most popular models is the Gaussian AR. However, when the data have outlier observations with "large" values, Gaussian models are not a good choice. We therefore abandon the assumption of normality of the data distribution and propose the AR model based on the double Pareto distribution. We introduce the estimators of the model's parameters, obtained by the maximum likelihood method. For this purpose, we use the Maclaurin series expansion and the Chebyshev polynomials expansion of the likelihood function. We compare the results with the Yule-Walker estimator in the finite variance case and with the modified Yule-Walker estimator in the infinite variance case. The accuracy of the results obtained was checked by Monte Carlo simulations.

2010 Mathematics Subject Classification: Primary: 60E05; Secondary: 60G10.

Key words and phrases: autoregressive model, double Pareto distribution, maximum likelihood estimation, Monte Carlo simulations.

1. Introduction.

Time series models are a popular tool commonly used to model time-varying phenomena. The choice of model depends on the application. However, the most popular is the Gaussian autoregressive moving average model (ARMA short) [3, 4, 20, 7]. In practical applications, the most popular is the autoregressive model (AR short), which is a special case of the ARMA model, because of its simplicity and intuitiveness. The AR model is due its popularity to the fact that many other models are considered as its generalizations. One of these extensions is, for example, the AR model with time-dependent coefficients [19]. One significant difference between the standard and time-dependent models is that the extended model is nonstationary in a broad sense. The other interesting extension is the PAR model, in which the parameters of the AR equation are periodic in time [6, 26, 25, 1, 8, 29]. One can also consider the versions of the AR model in continuous time [12, 27].

Another approach to extend the classical AR model is to change the distribution of the noise. In the literature, the large class of distributions that were investigated as noise in the AR model can be found [9, 28, 30]. Due to the possibility of describing the data with stationary-like behavior and large observations, the most popular model is with $\alpha-$stable distributions that belong to the heavy-tailed class of distributions [22, 18]. However, there are significant problems with the statistical analysis of the AR model with the $\alpha-$ stable distributed noise. For instance, lack of explicit form of the probability density function (PDF) of the $\alpha-$stable distribution for almost all α values prevents the use of PDF-based estimation techniques (e.g., maximum likelihood method).

In this paper, instead of considering $\alpha-$stable distribution in the stationary AR model, we introduce a model with the double Pareto distribution [15]. According to our knowledge, this is the first paper in the literature to provide an AR model with this noise distribution. In such a case, the model for describing data with large observations exhibiting stationary properties is still preserved.

In this paper, we introduce the idea of maximum likelihood estimation (MLE short) for the AR model with a double Pareto distribution of the noise. We investigate two cases: a model with finite and infinite variance. In the first case, we compare our results with the classical Yule-Walker approach. In the second case, the autocovariance function is not defined, so we compare MLE estimators with a modification of the Yule-Walker method prepared to handle cases with infinite variance [14]. The general methodology of the maximum likelihood method is based on the likelihood function of the model, which is expressed in the means of the PDF of the noise. The maximum of the likelihood function cannot be obtained analytically in a simple way due to the complicated form of the double Pareto distribution PDF. Here, similarly to [28] for Student's t distributed noise, we propose to approximate it by the Maclaurin series expansion. We also slightly modified this approach and propose an approximation of the likelihood function by Chebyshev polynomials. These approximations make finding the explicit formula for the parameters' estimators possible. The effectiveness of estimators obtained using this methodology is demonstrated on the basis of Monte Carlo simulations.

The rest of the paper is organized as follows. In Section 2 we discuss the AR model with Gaussian noise and with double Pareto noise. We indicate the main properties of such processes. Next, in Section 3 we present how to estimate the parameters of the AR model with the maximum likelihood method, the Yule-Walker method and the modified Yule-Walker method. In Section 4 we provide the computational analysis of the estimators using the Monte Carlo method. In the last section, we present the conclusions of the paper.

2. Autoregressive model. *Autoregressive model* is among the most common time series models. This process $\{X_t\}_{t\in\mathbb{Z}}$ is defined by the following equation [3, 23]:

$$X_t - \phi_1 X_{t-1} - \ldots - \phi_p X_{t-p} = Z_t, \tag{1}$$

where $\{Z_t\}_{t\in\mathbb{Z}}$ is the sequence of uncorrelated random variables with finite variance and mean equal to 0, known as *white noise*. In the classical case, the distribution of $\{Z_t\}_{t\in\mathbb{Z}}$ in (1) is assumed to be Gaussian. This is motivated by the fact that, under this assumption, the estimation methods and testing procedures are well defined and easy to calculate in explicit form. Moreover, assumption of the Gaussian distribution is important due to the Central Limit Theorem. However, when data are heavy-tailed distributed, a Gaussian distribution is inappropriate, and other distributions need to be considered. In this paper, we consider the AR model with a double Pareto distribution, which is an extension of the Pareto distribution [2]. The explicit form of the double Pareto distribution, PDF, enables the use of a wide range of statistical methods. The next subsection contains a short reminder of the basic properties of the standard AR model (1) and then we present the AR model with the double Pareto distribution as well as its main properties.

2.1. AR models with Gaussian distribution. In the Gaussian distributed AR model, it is assumed that noise $\{Z_t\}_{t\in\mathbb{Z}}$ in the model (1) is a sequence of uncorrelated Gaussian distributed random variables with mean $\mu = 0$ and variance σ^2, $\sigma \in \mathbb{R}_+$ ($\mathcal{N}(0,\sigma^2)$). Their distribution PDF is just a PDF of the normal distribution:

$$f_G(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right), \ x \in \mathbb{R}. \tag{2}$$

The autoregressive polynomial of the model (1) is defined by the following equation:

$$\phi(z) = 1 - \phi_1 z - \ldots - \phi_p z^p. \tag{3}$$

There exists a stationary solution of the time series $\{X_t\}_{t\in\mathbb{Z}}$ defined by (1) iff for every $|z| = 1$ polynomial (3) is different from zero. If we additionally assume:

$$\forall_{|z|\leqslant 1} \quad \phi(z) = 1 - \phi_1 z - \ldots - \phi_p z^p \neq 0, \tag{4}$$

then we call the time series (1) causal. Moreover, this assumption gives us the explicit form of the model (1) solution:

$$X_t = \sum_{j=0}^{\infty} \psi_j Z_{t-j}, \tag{5}$$

where $\{\psi_j\}_{j=0}^{\infty}$ is an absolute convergent sequence of constants, defined as $\psi(z) = \sum_{j=0}^{\infty} \psi_j z^j = \frac{1}{\phi(z)}$ [3]. This assumption is needed for the convergence of the sum (5). L^2 convergence is guaranteed by a finite second moment for all Z_t in the model (1).

Other examples of important properties of the AR Gaussian model are finite moments of all order and the existence of an explicit formula of its likelihood function. These properties make a wide range of prediction and estimation methods possible. For instance, the maximum likelihood method or the least squares method [3].

2.2. AR models with double Pareto distribution. The double Pareto distribution is a continuous probability distribution with PDF given by following formula [15]:

$$f(x) = \frac{\theta}{2\beta} \begin{cases} (\frac{x}{\beta})^{\theta-1}, & \text{if } 0 < x < \beta \\ (\frac{\beta}{x})^{\theta+1}, & \text{if } x \geq \beta \end{cases} \tag{6}$$

and CDF given by [15]:

$$F(x) = \begin{cases} \frac{1}{2}(\frac{x}{\beta})^{\theta}, & \text{if } 0 < x < \beta \\ 1 - \frac{1}{2}(\frac{\beta}{x})^{\theta}, & \text{if } x \geq \beta, \end{cases} \tag{7}$$

where $\theta > 0$, $\beta > 0$. This distribution has a few important properties, which we describe below.

Moment of order p for double Pareto distributions exists only for $p < \theta$. This property is important because the existence of moments is an assumption in some methods of parameter estimation. The double Pareto distribution belongs to the domain of attraction of the α-stable distribution. Another important property is that the tails of the double Pareto distribution exhibit a power law behavior.

In this paper, we extend the classical definition of the AR model given in (1) and assume that the noise $\{Z_t\}_{t\in\mathbb{Z}}$ is a sequence of independent identically distributed (i.i.d.) double Pareto-distributed random variables. We consider two cases, $\theta > 2$ and $\theta \leq 2$. In the first case, we can use any methods dedicated for finite-variance distributed models. In the second case, the variance of the noise is infinite, and dedicated algorithms need to be applied. In this case, the assumption of a classically understood lack of correlation of the noise sequence is impossible. However, we can assume that the noise sequence satisfies the independence of the components of the time series $\{Z_t\}_{t\in\mathbb{Z}}$, which is more general than lack of correlation. For any other class of heavy-tailed distributions in the literature, this problem was also investigated; see, for instance, [13, 10, 11]. Let us note that the preliminary investigation whether the data comes from finite- and infinite-variance models is a significant but challenging issue. However, there are methods dedicated to this problem. For instance, in [24, 5] advanced techniques are proposed for testing whether the data are from a finite-variance distribution, without the specification of the distribution type. Moreover, in [16] the problem of discriminating the finite- and infinite-variance models is also discussed.

Similarly to the Gaussian case, the solution of the model is given by formula (5) if condition (4) is satisfied. However, when $\theta \leq 2$, the L^2 convergence cannot be assumed at this time and different norms are analyzed; see, for example, [22].

3. Estimation methodology.

3.1. Maximum likelihood method.

The maximum likelihood method is one of the most widely used methods to estimate model parameters. Its idea is based on the maximization of the likelihood function with respect to the estimated parameters of the given model. The likelihood function is defined as the multidimensional PDF of the given sample from the model. In many cases, the likelihood function might have a sophisticated form. Therefore, finding an analytical solution to this issue is very difficult or sometimes impossible. Then, numerical approximations can be considered as a solution of the problem.

We recall that the log-likelihood function of the Gaussian AR model has the following form [3]:

$$\log L(\sigma;\vec{\phi}) = -\frac{(n-p)}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{t=p+1}^{n}(X_t-\phi_1X_{t-1}-\ldots-\phi_pX_{t-p})^2, \quad (8)$$

where $\vec{\phi}=(\phi_1,\ldots,\phi_p)$.

To find all estimators, one just calculates derivatives with respect to all unknown parameters and equates them all to zero. For instance, in the case of $p=1$, the formula for $\hat{\phi}_1$ as follows::

$$\hat{\phi}_1 = \frac{\sum_{t=2}^{n}X_tX_{t-1}}{\sum_{t=2}^{n}X_{t-1}}. \quad (9)$$

In the case of the considered double Pareto distributed AR(p) model, we apply the same idea.

First, it is necessary to determine the likelihood function. Here, we assume that the residuals are i.i.d. by a double Pareto distribution with the PDF given by formula (6). For simplicity of notation, we take the notation $A_t = (X_t-\phi_1X_{t-1}-\ldots-\phi_pX_{t-p})$, $B_t=(X_t-\phi_1X_{t-1}-\ldots-\phi_pX_{t-p}-1)$. For a given sample, $\{X_t\}_{t=1}^n$ we obtain the following form of the likelihood function:

$$L(\theta,\beta,\vec{\phi}) = \begin{cases}\prod_{t=p+1}^{n}\frac{\theta}{2\beta}(\frac{A_t}{\beta})^{\theta-1}, & \text{for } 0<A_t<\beta\\ \prod_{t=p+1}^{n}\frac{\theta}{2\beta}(\frac{\beta}{A_t})^{\theta+1}, & \text{for } A_t\geq\beta,\end{cases} \quad (10)$$

Thus, the log-likelihood function is given by:

$$\begin{aligned}\log L(\theta,\beta,\vec{\phi}) &= \begin{cases}\sum_{t=p+1}^{n}\log(\frac{\theta}{2\beta})+(\theta-1)(\log A_t-\log\beta), \text{for } 0<A_t<\beta\\ \sum_{t=p+1}^{n}\log(\frac{\theta}{2\beta})+(\theta+1)(\log\beta-\log A_t), \text{for } A_t\geq\beta\end{cases}\\ &= \begin{cases}(n-p)(\log\theta-\theta\log\beta-\log 2)+(\theta-1)\sum_{t=p+1}^{n}\log A_t, \text{for } 0<A_t<\beta\\ (n-p)(\log\theta+\theta\log\beta-\log 2)-(\theta+1)\sum_{t=p+1}^{n}\log A_t, \text{for } A_t\geq\beta.\end{cases}\end{aligned} \quad (11)$$

Finding the maximum of the function (11) is a challenging issue due to the $\log(X_t-\phi_1X_{t-1}-\ldots-\phi_pX_{t-p})$ component. Calculating any of the derivatives with respect to ϕ_i $i=1,2,\cdots,p$ leads to the equation with a sum of

$n-p$ elements in the form of fractions with different denominators. All denominators contain parameters that are unknown and have to be estimated. This fact causes that if we calculate derivatives with respect to unknown parameters, which equates them all to zero, we obtain a complicated set of equations, computationally expensive to solve. For this reason, we decided to use two expansions of the log-likelihood function (11): the Maclarurin series expansion and the Chebyshev polynomial expansion. We assume that the condition $|X_t - \phi_1 X_{t-1} - \ldots - \phi_p X_{t-p} - 1| < 1$ is satisfied. If this condition is not satisfied in practical applications, we divide the real data by a sufficiently large number.

3.1.1. Maclaurin series expansion. Taking the first three terms of the Maclaurin series [21] gives the following approximation of the log-likelihood function (11):

$$\log L(\theta, \beta, \vec{\phi}) \approx \begin{cases} (n-p)(\log\theta - \theta\log\beta - \log 2)+ \\ (\theta-1)(\sum_{t=p+1}^{n} B_t - \frac{1}{2}\sum_{t=p+1}^{n} B_t^2 + \\ +\frac{1}{3}\sum_{t=p+1}^{n} B_t^3), \quad \text{for } 0 < A_t < \beta \\ (n-p)(\log\theta + \theta\log\beta - \log 2)+ \\ (\theta+1)(\sum_{t=p+1}^{n} B_t - \frac{1}{2}\sum_{t=p+1}^{n} B_t^2 + \\ +\frac{1}{3}\sum_{i=2}^{n} B_t^3), \quad \text{for } A_t \geq \beta. \end{cases} \tag{12}$$

Now, it is easy to find partial derivatives with respect to the unknown parameters:

$$\frac{\partial \log L\left(\theta, \beta, \vec{\phi}\right)}{\partial \phi_j} = \begin{cases} (\theta-1)\,(-\sum_{t=p+1}^{n} [X_{t-j}(1 - B_t + B_t^2)]), & \text{for } 0 < A_t < \beta \\ (\theta+1)(-\sum_{t=p+1}^{n} [X_{t-j}(1 - B_t + B_t^2)]), & \text{for } A_t \geq \beta, \end{cases} \tag{13}$$

where $j = 1, 2, \ldots$.

$$\frac{\partial \log L(\theta, \beta, \vec{\phi})}{\partial \beta} = \frac{\theta}{\beta} \tag{14}$$

and

$$\frac{\partial \log L(\theta,\beta,\vec{\phi})}{\partial\theta} = \begin{cases} (n-p)(\frac{1}{\theta} - \log\beta) + (\sum_{t=p+1}^{n} B_t \\ -\frac{1}{2}\sum_{t=p+1}^{n} B_t^2 + \frac{1}{3}\sum_{i=2}^{n} B_t^3), & \text{for } 0 < A_t < \beta \\ (n-p)(\frac{1}{\theta} + \log\beta) + (\sum_{t=p+1}^{n} B_t \\ -\frac{1}{2}\sum_{t=p+1}^{n} B_t^2 + \frac{1}{3}\sum_{i=2}^{n} B_t^3), & \text{for } A_t \geq \beta. \end{cases} \tag{15}$$

To find the maximum of the function (12) one has to set the partial derivatives equal to zero and solve the resulting system of equations. The derivative of the log-likelihood function of the double Pareto distribution with respect to β is always positive, and thus the function is always positive, making it impossible to find the maximum of the function (12) with respect to the parameter β. However, for a fixed value of β, it is possible to find the estimates of $\phi_j, j = 1, 2, \ldots, p$ parameters. The formula obtained using the first two terms of the Maclaurin expansion for $p = 1$ is given by:

$$\hat{\phi_1} = \frac{\sum_{t=2}^{n} X_{t-1}(X_t - 2)}{\sum_{t=2}^{n} X_{t-1}^2}. \tag{16}$$

3.1.2. Chebyshev polynomials expansion. Taking the first four Chebyshev polynomials of the first kind [17] gives the following approximation of the log-likelihood function (11):

$$\log L(\theta,\beta,\vec{\phi}) \approx \begin{cases} (n-p)(\log\theta - \theta\log\beta - \log 2) + \\ (\theta - 1)(\sum_{t=p+1}^{n}(1-\log 2) - 2\sum_{t=p+1}^{n} B_t^2 + \\ +\frac{8}{3}\sum_{t=p+1}^{n} B_t^3), \quad \text{for } 0 < A_t < \beta \\ (n-p)(\log\theta + \theta\log\beta - \log 2) + \\ (\theta + 1)(\sum_{t=p+1}^{n}(1-\log 2) - 2\sum_{t=p+1}^{n} B_t^2 + \\ +\frac{8}{3}\sum_{t=p+1}^{n} B_t^3), \quad \text{for } A_t \geq \beta \end{cases} \tag{17}$$

Now, it is easy to find partial derivatives with respect to the unknown parameters:

$$\frac{\partial \log L\left(\theta,\beta,\vec{\phi}\right)}{\partial \phi_j} = \begin{cases} (\theta-1)\,(-\sum_{t=p+1}^{n}[X_{t-j}(-4B+8B_t^2)]), & \text{for } 0<A_t<\beta \\ (\theta+1)(-\sum_{t=p+1}^{n}[X_{t-j}(-4B_t+8B_t^2)]), & \text{for } A_t\geq\beta, \end{cases} \tag{18}$$

where $j = 1, 2, \dots, p$.

$$\frac{\partial \log L(\theta,\beta,\vec{\phi})}{\partial \beta} = \frac{\theta}{\beta} \tag{19}$$

and

$$\frac{\partial \log L(\theta,\beta,\vec{\phi})}{\partial \theta} = \begin{cases} (n-p)(\frac{1}{\theta}-\log\beta)+ & \\ +(\sum_{t=p+1}^{n}(1-\log 2)-2\sum_{t=p+1}^{n}B_t^2+ & \\ +\frac{8}{3}\sum_{t=p+1}^{n}B_t^3), & \text{for } 0<A_t<\beta \\ (n-p)(\frac{1}{\theta}+\log\beta)+ & \\ +(\sum_{t=p+1}^{n}(1-\log 2)-2\sum_{t=p+1}^{n}B_t^2+ & \\ +\frac{8}{3}\sum_{t=p+1}^{n}B_t^3), & \text{for } A_t\geq\beta. \end{cases} \tag{20}$$

To find the maximum of the function (17) one has to set the partial derivatives equal to zero and solve the resulting system of equations. The derivative of the logarithmic likelihood function of the double Pareto distribution with respect to β is always positive, and thus the function is always positive, making it impossible to find the maximum of the function (17) with respect to the parameter β. However, for a fixed value of β, it is possible to find the estimates of $\phi_j, j = 1, 2, \dots, p$ parameters. The formula obtained using the first three terms of the Chebyshev polynomials of the first kind for $p = 1$ is given by:

$$\hat{\phi}_1 = \frac{\sum_{t=2}^{n} X_{t-1}\,(X_t-1)}{\sum_{t=2}^{n} X_{t-1}^2}. \tag{21}$$

3.2. Yule-Walker method and modified Yule-Walker method. Considering the general AR(p) model given by(1) and assuming that it is causal, we can establish the so-called Yule-Walker method for ϕ_j, $j = 1, 2, \dots$ parameters estimation [3]. We note that the Yule-Walker algorithm does not assume any specific distribution of the residuals. The only assumption is related to their finite-variance distribution.

In the case of AR(1) model with finite variance, we obtain:

$$\hat{\phi}_1 = \frac{\sum_{t=1}^{n-1}\left(X_{t+1}-\bar{X}\right)\left(X_t-\bar{X}\right)}{\sum_{t=1}^{n}\left(X_t-\bar{X}\right)^2}, \tag{22}$$

where $\bar{X}$ is the corresponding sample mean.

When the variance of the noise distribution is infinite, the classical Yule-Walker approach cannot be used. However, there exist extensions of this method, where covariance is replaced by other measures of dependence. One of the possible approaches is to use fractional lower order covariance (FLOC), which is described in [13]. The estimator obtained using this measure for the AR(1) model is given by:

$$\hat{\phi}_1 = \frac{\sum_{t=2}^{n}|X_{t-1}|^{\gamma}|X_t|^{\delta}\, sign\left[X_{t-1}X_t\right]}{\sum_{t=2}^{n}|X_{t-1}|^{\gamma-1}|X_t|^{\delta+1}}, \tag{23}$$

where γ and δ are parameters that satisfy $0<\gamma+\delta<\theta$ in the case of double Pareto distribution.

3.3. Estimation of θ and β parameters. The estimator of the θ parameter can be determined without using the Maclaurin expansion and has the following form with a known β parameter [15]:

$$\hat{\theta} = \frac{n-p}{\left|\sum_{t=p+1}^{n}\log A_t - \log\beta\right|}. \tag{24}$$

In our case, we estimate the model parameters multiple times for different possible values of β. In more detail, for each possible, β we introduce the methodology to estimate the parameters of the AR model. Then, we calculate the residuals from the model, and finally we check if they are double Pareto distributed. This problem can be addressed with the theoretical cumulative distribution function (CDF) of the double Pareto distribution by comparing it with the empirical cumulative distribution function (ECDF) of residuals for each selected value of β. These two functions can be compared, e.g. by calculating the mean square error between them.

$$\mathrm{MSE} = \sum_{i=p+1}^{n}\left(F_S(z_i)-\hat{F}_S(z_i)\right)^2, \tag{25}$$

where $F_S(\cdot)$ is the CDF of the double Pareto distribution (7) and $\hat{F}_S(\cdot)$ is the ECDF of the residuals. By $\{z_t\}_{t=1}^{n}$ we mark a sequence of realizations of the residuals of the AR(p) model. We selected a value of β, which leads to the lowest value of MSE (25). However, the proposed method has the disadvantage that we cannot limit the range of parameters β to be considered. The number of terms in a set should be selected following the desired precision of the β estimator.

3.4. Estimation of the p parameter. In the previous subsection, we assumed that the parameter p is known. Now, we investigate the problem of how to select the optimal value of p. If the noise distribution is Gaussian, *information criteria*, such as the Akaike Information Criterion, Bayesian Information Criterion or the final prediction error criterion [3] are natural choices to solve this issue. To use the first two, the existence of a global maximum of the likelihood function (10) is necessarily. The variance estimator is the basis for the third one, but in our case it is not properly defined for the infinite-variance case. Due to these reasons, the classical information criteria are useless for the double Pareto distributed AR(p) model case.

To solve this problem, we proceed analogously to the estimation of the β parameter. That is, for all p from a given set of possible values, we estimate all parameters and then calculate the MSE (25) between the ECDF of residuals and the theoretical CDF of the double Pareto distribution. This approach is computationally expensive, but fortunately most of the possible values of the p parameter are relatively small.

4. Monte Carlo simulations. We check the precision of the methods described in Section 3 to estimate the model parameters by performing Monte Carlo simulations using the first three terms of the Maclaurin expansion and the first four Chebyshev polynomials of the first kind. Here, we consider the AR(1) time series with the double Pareto distribution with parameters $\phi_1 = 0.7$, $\beta = 0.8$ and $\theta = 2.5$ in the finite variance case and $\phi_1 = 0.7$, $\beta = 0.8$ and $\theta = 1.3$ in the infinite variance case. Such a selection of parameters ensures that the model is stationary. We choose four lengths of the simulated trajectories, namely $n = 100$, $n = 500$, $n = 1000$, and $n = 5000$. For each case, the trajectories were simulated 1000 times to verify the efficiency of the proposed techniques.

4.1. Finite variance case. Single trajectory of series with finite variance and $n = 500$ is presented in Figure 1. First, we compare the estimators obtained with the Yule-Walker method, properly defined for the finite-variance case. In Figures 3a, 3b, 3c, 3d, we see that in the finite variance case both Maclaurin series and Chebyshev polynomials expansions tend to underestimate the value of ϕ_1 and in this case the Yule-Walker estimators are significantly better.

4.2. Infinite variance case. Single trajectory of series with infinite variance and $n = 500$ is presented in Figure 2. In this case, expansions of MLE estimators outperform modified Yule-Walker algorithm, what we can

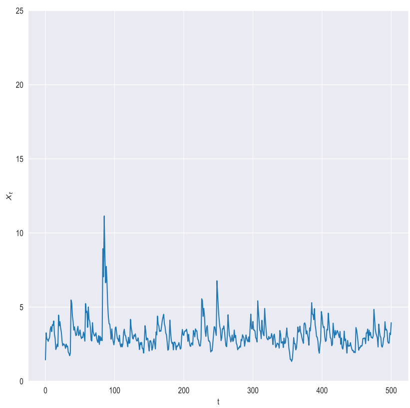


Figure 1: Sample trajectory for AR(1) with double Pareto distributed noise with finite variance

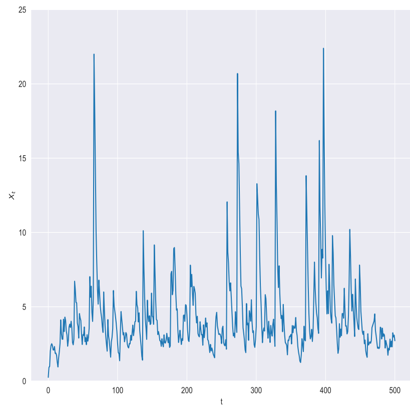


Figure 2: Sample trajectory for AR(1) with double Pareto distributed noise with infinite variance

see in Figures 4a, 4b, 4c, 4d, so we recommend using them in infinite variance case. We also observe that the median of the Taylor series expansion is closer to the theoretical value of the estimated parameter; however, it clearly tends to underestimate.

Due to the usefulness of the estimators obtained, we decided to investigate them more in the infinite variance case. As the set of possible values of β, we chose a sequence from 0.6 to 1 with a step equal to 0.02. We consider the possible p equal to 1 or 2. If we consider the higher values of p, the computational cost will increase.

First, we tested the results of the estimation of all parameters of the model using the Taylor series expansion. The results obtained from the estimates of p are as follows: for $n = 100$ the algorithm points 507 times to $\hat{p} = 1$; for $n = 500$ the algorithm points 509 times to $\hat{p} = 1$; for $n = 1000$ the algorithm points 525 times to $\hat{p} = 1$; for $n = 5000$ the algorithm points 504 times to $\hat{p} = 1$. The results of the estimation of ϕ_1, β, and θ for the simulated time series trajectories with $p = 1$ together with the estimation results of β for all trajectory lengths are presented in Figures 5a, 5b, 5c and 5d, respectively, for four considered trajectory lengths. We can see that in all cases, the actual value of the parameter β coincides with the median of the parameter determined by the simulation. The parameter θ is very close to the median in all cases. For the theoretical value of the parameter ϕ_1 for a trajectory length of 100 is almost equal to the third quartile, for trajectory lengths of 500 and 1000, the estimated parameter is between the third quartile and median determined for the simulation-calculated parameter, and for a trajectory with length of 5000 it almost coincides with the median for the simulation value of the parameter. For the parameters β and θ, we see that the interquartile range does not change with increasing trajectory length. For ϕ_1, we see that the interquartile range decreases with increasing n. Thus, the results are acceptable. Table 1 shows the results of the estimation of parameters ϕ_1 and ϕ_2 as in the Definition 1 of the AR(p) model for $p = 2$ for the data generated from the AR(1) process. We treated the data as if they were from the AR(2) process to validate the estimation of the parameter p, that is, we expected that the median simulation value of the parameter ϕ_2 was close to 0. For each n, we observe that the median of the simulation-determined $\hat{\phi_1}$ is close to 0.7. For this estimator, we also observe that the interquartile decreases with increasing n. The median $\hat{\phi_2}$ determined from the simulation is close to 0 and approaches it as n increases. In contrast, its interquartile range decreases with increasing n and already for $n = 5000$ it reaches a very small value. For both estimators, we see that the mean square error between the theoretical and simulation values is small for all values of n.

n	ϕ_1	ϕ_2	$med(\hat{\phi_1})$	$med(\hat{\phi_2})$	$IQR_{\hat{\phi_1}}$	$IQR_{\hat{\phi_2}}$	$MSE_{\hat{\phi_1}}$	$MSE_{\hat{\phi_2}}$
100	0.7	0	0.662	$-7.771 \cdot 10^{-6}$	0.693	$2.929 \cdot 10^{-6}$	0.155	0.421
500	0.7	0	0.693	$-8.341 \cdot 10^{-7}$	0.025	$2.929 \cdot 10^{-6}$	0.090	0.465
1000	0.7	0	0.697	$-1.949 \cdot 10^{-7}$	0.012	$7.123 \cdot 10^{-7}$	0.066	0.476
5000	0.7	0	0.700	$-7.169 \cdot 10^{-9}$	0.004	$2.685 \cdot 10^{-8}$	0.049	0.487

Table 1: Table with median, interquartile range and mean squared error for estimates: $\hat{\phi_1}$ and $\hat{\phi_2}$ using the Taylor series approximation for the AR (2) model with double Pareto distributed noise with infinite variance for data sampled from the AR (1) model with double Pareto distributed noise with infinite variance for trajectories lengths: 100, 500, 1000 and 5000

Secondly, we test results of the estimation of all parameters of the model using Chebyshev polynomials of the first kind expansion. The results obtained from the estimates of p are as follows: for $n = 100$ the algorithm points 522 times to $\hat{p} = 1$; for $n = 500$ the algorithm points 525 times to $\hat{p} = 1$; for $n = 1000$ the algorithm points 502 times to $\hat{p} = 1$; for $n = 5000$ the algorithm points 512 times to $\hat{p} = 1$. The results of parameter estimation ϕ_1, β, and θ for the simulated time series trajectories with $p = 1$ along with the estimation results of β for all trajectories are presented in Figures 6a, 6b, 6c and 6d, respectively, for four considered trajectory lengths. We can see that in all cases, the actual value of the parameter β coincides with the median of the parameter determined by the simulation. The parameter θ is very close to the median in all cases. For the theoretical value of the parameter ϕ_1 for a trajectory length of 100 the estimated parameter is between the third quartile and median for trajectory lengths of 500, 1000 and 5000, the estimated parameter is between the first quartile and median determined for the simulation-calculated parameter. For the parameters β and θ, we see that the interquartile range does not change with increasing trajectory length. For ϕ_1, we see that the interquartile range decreases with increasing n. Thus, the results are acceptable. Table 2 shows the results of the estimation of parameters ϕ_1 and ϕ_2 as in the Definition 1 of the AR(p) model for $p = 2$ for the data generated from the AR(1) process. We treated the data as if they were from the AR(2) process to validate the estimation of the parameter p, that is, we expected that the median simulation value of the parameter ϕ_2 was close to 0. For each n, we observe that the median of the simulation-determined $\hat{\phi}_1$ is close to 0.7. For this estimator, we also observe that the interquartile decreases with an increase of n. The median $\hat{\phi}_2$ determined from the simulation is close to 0 and approaches it as n increases. In contrast, its interquartile range decreases with increasing n and already for $n = 5000$ it reaches a very small value. For both estimators, we see that the mean square error between the theoretical and simulation values is small for all values of n.

n	ϕ_1	ϕ_2	$med(\hat{\phi}_1)$	$med(\hat{\phi}_2)$	$IQR_{\hat{\phi}_1}$	$IQR_{\hat{\phi}_2}$	$MSE_{\hat{\phi}_1}$	$MSE_{\hat{\phi}_2}$
100	0.7	0	0.692	$-4.152 \cdot 10^{-6}$	0.710	$1.790 \cdot 10^{-5}$	0.134	0.427
500	0.7	0	0.701	$-3.666 \cdot 10^{-7}$	0.019	$1.256 \cdot 10^{-6}$	0.077	0.465
1000	0.7	0	0.702	$-9.125 \cdot 10^{-8}$	0.012	$3.463 \cdot 10^{-7}$	0.061	0.476
5000	0.7	0	0.701	$-3.736 \cdot 10^{-9}$	0.005	$1.450 \cdot 10^{-8}$	0.051	0.485

Table 2: Table with median, interquartile range and mean squared error for estimates: $\hat{\phi}_1$ and $\hat{\phi}_2$ using the Chebyshev polynomials approximation for the AR(2) model with double Pareto distributed noise with infinite variance for data sampled from the AR(1) model with double Pareto distributed noise with infinite variance for trajectories lengths: 100, 500, 1000 and 5000

5. Conclusions. In this paper, we have considered the AR model with double Pareto distribution. This model can be useful for modeling data with stationary-like behavior but with large observations. We propose here the maximum likelihood-based approach to estimate the model's parameters. For the double Pareto distribution, the PDF is given in explicit form; however, it is expressed by means of complicated and special functions. Thus, it is impossible to find the explicit formula for the parameters' estimators. However, the application of the Maclaurin series expansion and Chebyshev polynomials expansion gives the possibility of obtaining exact formulas for the estimated parameters. We compare the obtained estimators with the Yule-Walker method in the finite variance case and with the modified Yule-Walker method in the infinite variance case. In the first case, the Yule-Walker method gives better results; however, when the variance is infinite, the proposed estimators perform better. Using the Monte Carlo method, we demonstrate the effectiveness of the proposed estimators.

Author Contributions: Conceptualization: [A. Wyłomańska, H. Woszczek], Methodology: [A. Wyłomańska, H. Woszczek], Formal analysis and investigation: [H. Woszczek], Writing - original draft preparation: [A. Wyłomańska, H. Woszczek]; Writing - review and editing: [A. Wyłomańska, H. Woszczek], Supervision: [A. Wyłomańska].

Funding: The work of AW was supported by National Center of Science under Opus Grant 2020/37/B/HS4/00120 "Market risk model identification and validation using novel statistical, probabilistic, and machine learning tools".

Conflicts of Interest: The authors declare no conflict of interest.

References

[1] P. L. Anderson and M. M. Meerschaert. Parameter estimation for periodically stationary time series. *Journal of Time Series Analysis*, 26(4): 489–518, 2005. doi: 10.1111/j.1467-9892.2005.00428.x. Cited on p. 121.

[2] B. C. Arnold. Pareto distribution. *Wiley StatsRef: Statistics Reference Online*, pages 1–10, 2014. Cited on p. 123.

[3] P. J. Brockwell and R. A. Davis. *Introduction to time series and forecasting, Second Edition*. Springer, 2002. Zbl 1355.62001. Cited on pp. 121, 122, 123, 125, 128, and 130.

[4] P. J. Brockwell and R. A. Davis. *Time series: theory and methods*. Springer Science & Business Media, 2013. Zbl 1169.62074. Cited on p. 121.

[5] I. Fedotenkov. A bootstrap method to test for the existence of finite moments. *Journal of Nonparametric Statistics*, 25(2):315–322, jun 2013. doi: 10.1080/10485252.2012.752487. Cited on p. 124.

[6] P. Franses. *Periodicity and Stochastic Trends in Economic Time Series*. Oxford University Press, Oxford, 1996. Zbl 0868.62088. Cited on p. 121.

[7] J. Gao, Z.-Y. Xu, and L.-T. Zhang. Approximating long-memory DNA sequences by short-memory process. *Physica A: Statistical Mechanics and its Applications*, 388(17):3475 – 3485, 2009. doi: 10.1016/j.physa.2005.06.099. Cited on p. 121.

[8] P. Giri, S. Sundar, and A. Wyłomańska. Fractional lower-order covariance (floc)-based estimation for multidimensional PAR(1) model with $\alpha-$stable noise. *International Journal of Advances in Engineering Sciences and Applied Mathematics*, 13(2):215–235, 2021. ISSN 0975-5616. doi: 10.1007/s12572-021-00301-0. Cited on p. 121.

[9] A. Grzesiek and A. Wyłomańska. Asymptotic behavior of the cross-dependence measures for bidimensional AR(1) model with $\alpha-$stable noise. *Banach Center Publications*, 122:133–157, 2020. doi: 10.4064/bc122-8. Cited on p. 121.

[10] A. Grzesiek, M. Teuerle, and A. Wyłomańska. Cross-codifference for bidimensional VAR(1) time series with infinite variance. *Communications in Statistics - Simulation and Computation*, pages 1–26, 2019. doi: 10.1080/03610918.2019.1670840. Cited on p. 124.

[11] A. Grzesiek, P. Giri, S. Sundar, and A. Wyłomańska. Measures of cross-dependence for bidimensional periodic AR(1) model with alpha-stable distribution. *Journal of Time Series Analysis*, 41(6):785–807, 2020. doi: 10.1111/jtsa.12548. Cited on p. 124.

[12] A. C. Harvey and J. H. Stock. Continuous time autoregressive models with common stochastic trends. *Journal of Economic Dynamics and Control*, 12(2):365–384, 1988. ISSN 0165-1889. doi: 10.1016/0165-1889(88)90046-2. Cited on p. 121.

[13] P. Kruczek, A. Wyłomańska, M. Teuerle, and J. Gajda. The modified Yule-Walker method for alpha-stable time series models. *Physica A*, 469:588–603, 2017. doi: 10.1016/j.physa.2016.11.037. Cited on pp. 124 and 129.

[14] P. Kruczek, W. Żuławiński, P. Pagacz, and A. Wyłomańska. Fractional lower order covariance based-estimator for Ornstein-Uhlenbeck process with stable distribution. *Mathematica Applicanda*, 47(2):259–292, 2019. doi: 10.14708/ma.v47i2.6506. Cited on p. 122.

[15] F. M. Al-Athari. Parameter estimation for the double pareto distribution. *Journal of Mathematics and Statistics*, 7(4):289–294, oct 2011. doi: 10.3844/jmssp.2011.289.294. Cited on pp. 122, 124, and 129.

[16] K. Maraj-Zygmąt, G. Sikora, M. Pitera, and A. Wyłomańska. Goodness-of-fit test for stochastic processes using even empirical moments statistic. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 33(1):013128, jan 2023. doi: 10.1063/5.0111505. Cited on p. 124.

[17] J. Mason and D. C. Handscomb. *Chebyshev Polynomials*. Chapman and Hall/CRC, sep 2002. doi: 10.1201/9781420036114. Cited on p. 127.

[18] C. L. Nikias and M. Shao. *Signal processing with alpha-stable distributions and applications*. Wiley-Interscience, 1995. URL `https://dl.acm.org/doi/10.5555/210666`. Cited on p. 121.

[19] J. Nowicka-Zagrajek and A. Wyłomańska. Measures of dependence for stable AR(1) models with time-varying coefficients. *Stochastic Models*, 24(1):58–70, 2008. doi: 10.1080/15326340701826906. Cited on p. 121.

[20] M. M. Rounaghi and F. Nassir Zadeh. Investigation of market efficiency and financial stability between S&P 500 and London Stock Exchange: Monthly and yearly forecasting of time series stock returns using ARMA model. *Physica A: Statistical Mechanics and its Applications*, 456:10 –

21, 2016. doi: 10.1016/j.physa.2016.03.006. Cited on p. 121.

[21] W. Rudin. *Principles of Mathematical Analysis*. International series in pure and applied mathematics. McGraw-Hill, 1976. ISBN 9780070856134. URL https://books.google.pl/books?id=kwqzPAAACAAJ. Cited on p. 126.

[22] G. Samorodnitsky and M. Taqqu. *Stable Non-Gaussian Random Processes: Stochastic Models with Infinite Variance*. Chapman and Hall, 1994. Zbl 0925.60027. Cited on pp. 121 and 124.

[23] R. Shumway and D. Stoffer. *Time Series Analysis and Its Applications With R Examples, Third Edition*. Springer, 2016. Zbl 1367.62004. Cited on p. 122.

[24] L. Trapani. Testing for (in)finite moments. *Journal of Econometrics*, 191(1):57–68, mar 2016. doi: 10.1016/j.jeconom.2015.08.006. Cited on p. 124.

[25] B. Troutman. Some results in periodic autoregression. *Biometrika*, 66: 219–228, 1979. doi: 10.2307/2335652. Cited on p. 121.

[26] E. Ursu and K. F. Turkman. Periodic autoregressive model identification using genetic algorithms. *Journal of Time Series Analysis*, 33(3):398–405, 2012. doi: 10.1111/j.1467-9892.2011.00772.x. Cited on p. 121.

[27] A. Wyłomańska and J. Gajda. Stable continuous-time autoregressive process driven by stable subordinator. *Physica A: Statistical Mechanics and its Applications*, 444:1012–1026, feb 2016. doi: 10.1016/j.physa.2015.10.081. Cited on p. 121.

[28] A. Wyłomańska and P. Stępniak. The maximum likelihood method for student's t-distributed autoregressive model with infinite variance. *Mathematica Applicanda*, 48(2), mar 2021. doi: 10.14708/ma.v48i2.7059. Cited on pp. 121 and 122.

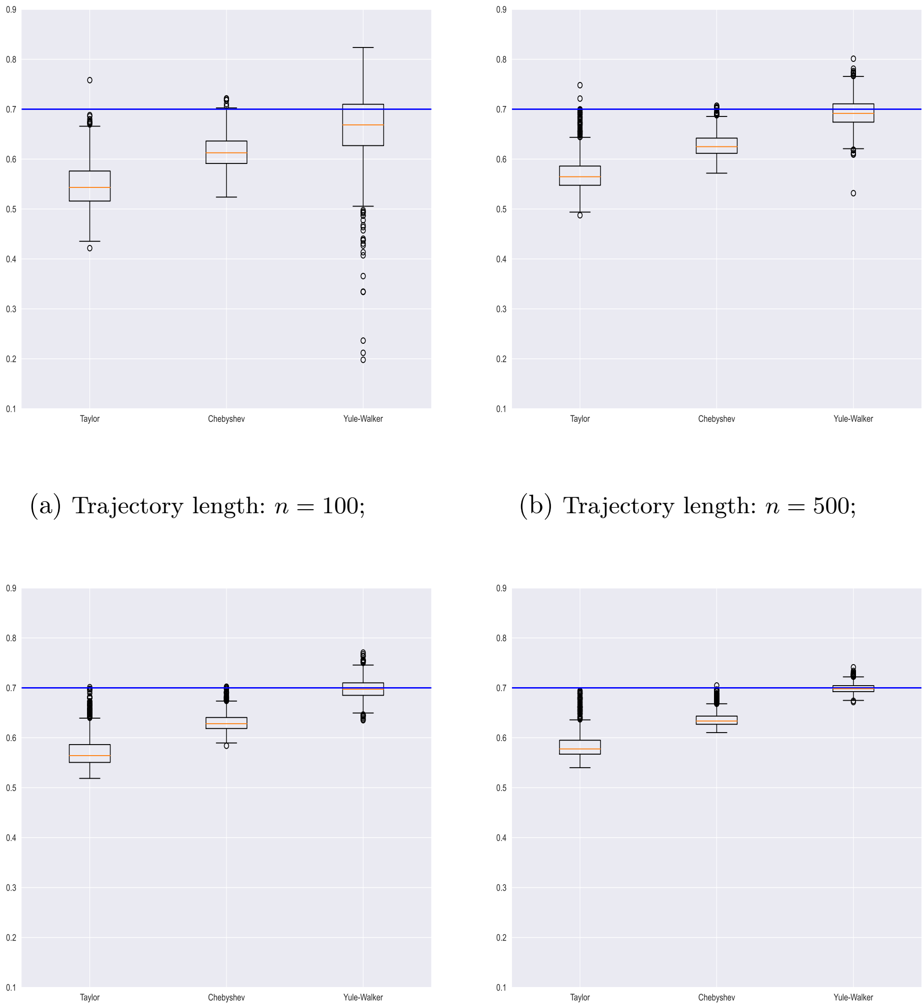
[29] A. Wyłomańska, J. Obuchowski, R. Zimroz, and H. Hurd. Periodic autoregressive modeling of vibration time series from planetary gearbox used in bucket wheel excavator. In F. Chaari, J. Leśkow, A. Napolitano, and A. Sanchez-Ramirez, editors, *Cyclostationarity: Theory and Methods*, pages 171–186. Springer International Publishing, Cham, 2014. doi: 10.1007/978-3-319-04187-2_12. Cited on p. 121.

[30] A. Zeevi and P. W. Glynn. Recurrence properties of autoregressive processes with super-heavy-tailed innovations. *Journal of Applied Probability*, 41(3):639–653, 2004. ISSN 00219002. URL http://www.jstor.org/stable/4141343. Cited on p. 121.

Appendices

A. Suplementary figures.

A.1. Estimates of ϕ_1 for the AR(1) model I. The Figures 3 show boxplots of comparison of estimates ϕ_1 for the AR(1) model with double Pareto distributed noise with finite variance using the Taylor series approximation for the MLE, Chebyshev polynomials of the first kind approximation for the MLE, and the Yule-Walker method.



(a) Trajectory length: $n = 100$;

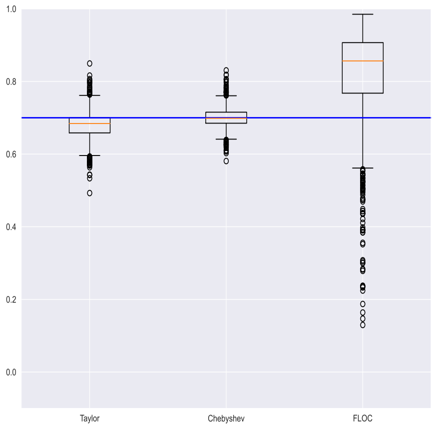
(b) Trajectory length: $n = 500$;

(c) Trajectory length: $n = 1000$;

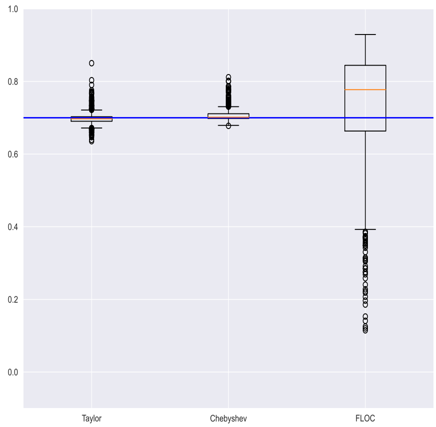
(d) Trajectory length: $n = 5000$;

Figure 3: The theoretical value of the parameter: $\phi_1 = 0.7$.

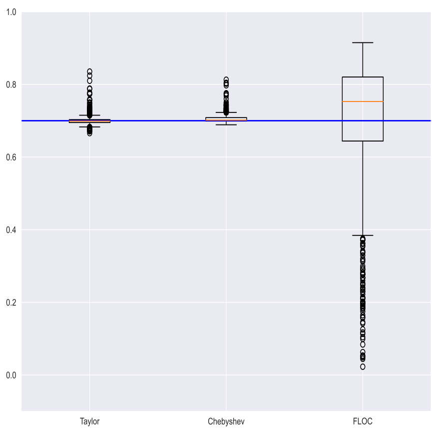
A.2. Estimates of ϕ_1 for the AR(1) model II. The figure 4 boxplots of comparison of estimates ϕ_1 for AR(1) model with double Pareto distributed noise with infinite variance using the Taylor series approximation of the MLE, Chebyshev polynomials of the first kind approximation for the MLE and the modified Yule-Walker method.



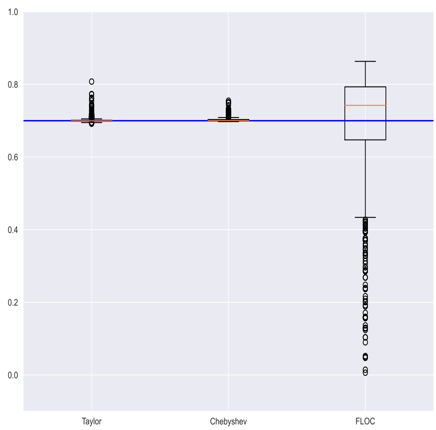
(a) Trajectory length: $n = 100$;



(b) Trajectory length: $n = 500$;



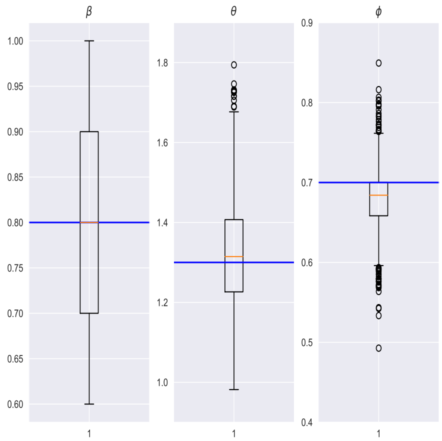
(c) Trajectory length: $n = 1000$;



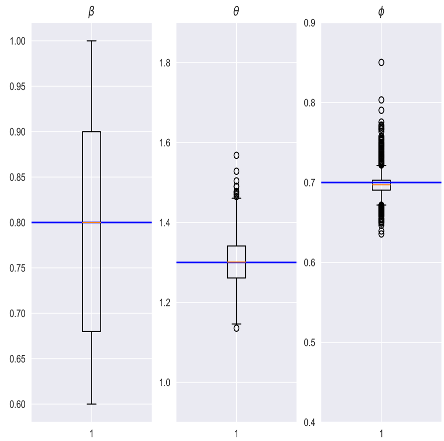
(d) Trajectory length: $n = 5000$;

Figure 4: The theoretical value of the parameter: $\phi_1 = 0.7$.

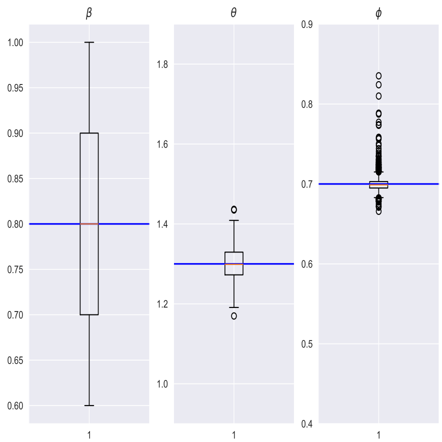
A.3. Estimates of ϕ_1 for the AR(1) model III. The figure 5 boxplots of the estimates ϕ_1, β and θ for the AR(1) model with double Pareto distributed noise with infinite variance using Taylor series approximation.



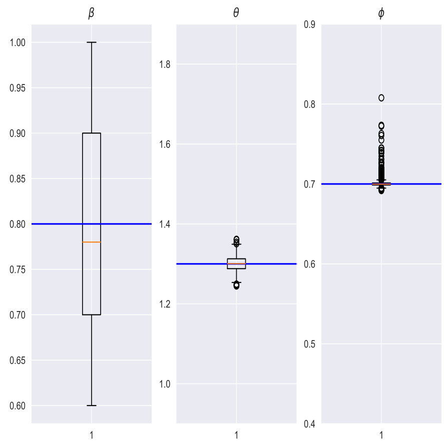
(a) Trajectory length: $n = 100$;



(b) Trajectory length: $n = 500$;



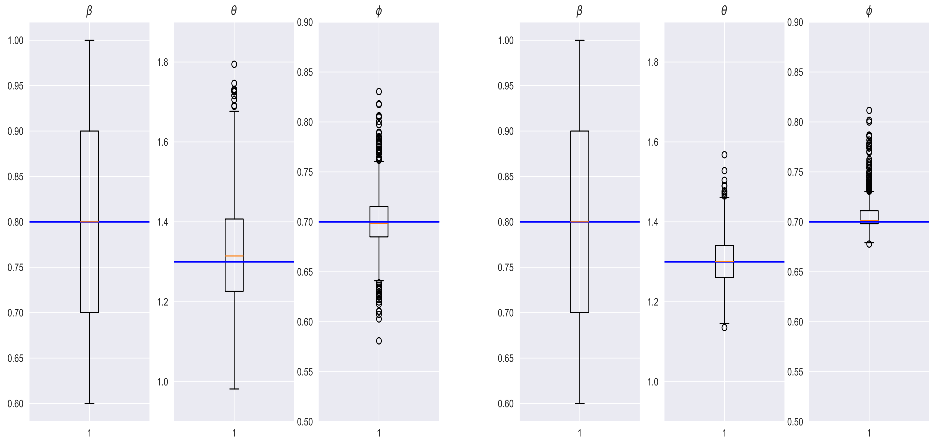
(c) Trajectory length: $n = 1000$;



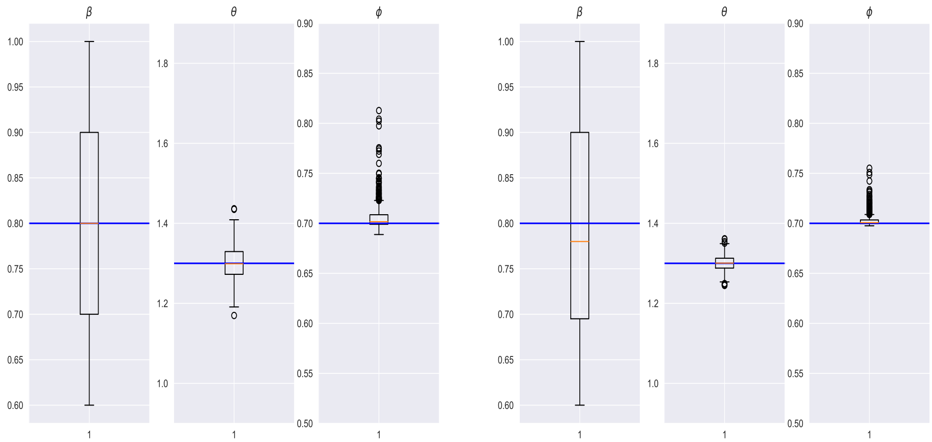
(d) Trajectory length: $n = 5000$;

Figure 5: The theoretical values of parameters: $\phi_1 = 0.7$, $\beta = 0.8$ and $\theta = 1.3$.

A.4. Estimates of ϕ_1 for the AR(1) model IV. The figure 6 the boxplots of the estimates ϕ_1, β and θ for the AR(1) model with double Pareto distributed noise with infinite variance using the Chebyshev polynomials approximation.



(a) Trajectory length: $n = 100$; (b) Trajectory length: $n = 500$;



(c) Trajectory length: $n = 1000$; (d) Trajectory length: $n = 5000$;

Figure 6: The theoretical values of parameters: $\phi_1 = 0.7$, $\beta = 0.8$ and $\theta = 1.3$.

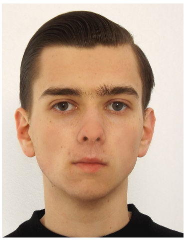
Model autoregresyjny z szumem o podwójnym rozkładzie Pareto

Hubert Woszczek, Agnieszka Wyłomańska

Streszczenie Modele szeregów czasowych to popularne narzędzie powszechnie stosowane do modelowania zjawisk zmiennych w czasie. Najpopularniejszym modelem jest gaussowski model AR, który jest stacjonarny. Jednak gdy w danych występują obserwacje odstające o „dużych“ wartościach, modele gaussowskie nie są odpowiednim narzędziem do ich modelowania. Odchodzimy zatem od założenia o normalności rozkładu danych i proponujemy model AR oparty na podwójnym rozkładzie Pareto. Przedstawiamy estymatory parametrów modelu, uzyskane metodą największej wiarogodności. W tym celu wykorzystujemy rozwinięcie funkcji warogodności w szereg Maclaurina oraz rozwinięcie za pomocą wielomianów Czebyszewa. Wyniki porównaliśmy z estymatorem Yule-Walkera w przypadku o skończonej wariancji oraz ze zmodyfikowanym estymatorem Yule-Walkera w przypadku nieskończonej wariancji. Poprwaność otrzymanych wyników została sprawdzona za pomocą symulacji Monte Carlo.

Klasyfikacja tematyczna AMS (2010): 60E05; 60G10.

Słowa kluczowe: model autoregresyjny, podwójny rozkład Pareto, symulacje MC.



Hubert Woszczek is MSc student of Applied Mathematics at Wroclaw University of Science and Technology (WUST). He received BCs degree in Applied Mathematics in 2023. His area of interests relates to time series analysis and stochastic processes.

Agnieszka Wyłomańska is a Professor of Wroclaw University of Science and Technology (WUST) at the Faculty of Pure and Applied Mathematics and a member of the Hugo Steinhaus Center for Stochastic Processes. Her area of interest relates to time series analysis, stochastic modeling and statistical analysis of real data. She is the author of more than 100 research papers in the area of the applied and industrial mathematics. She cooperates with industrial companies, especially the mining industry. References to her research papers can be found in zbMath under ai:wylomanska.agnieszka and in MathSciNet under ID: 738846.

Hubert Woszczek
Wrocław University of Science and Technology
Wybrzeze Wyspianskiego 27, 50-370 Wrocław, Poland
E-mail: 255722@student.pwr.edu.pl

Agnieszka Wyłomańska
Faculty of Pure and Applied Mathematics,
Wrocław University of Science and Technology,
Wybrzeże Wyspiańskiego 27, Wrocław 50-370
E-mail: Agnieszka.Wylomanska@pwr.edu.pl

Communicated by: Krzysztof Szajowski

(Received: 23rd of February 2023; revised: 20th of June 2023)