Generalized Entropy Optimization Distributions Dependent on Parameter in Time Series

ALADDIN SHAMILOV* and CIGDEM GIRIFTINOGLU*

*Anadolu University
Faculty of Science
Department of Statistics
26470 Eskisehir
TURKEY
asamilov@anadolu.edu.tr
cgiriftinoglu@anadolu.edu.tr

Abstract - In this paper, we have proposed Generalized Entropy Optimization Problems (GEOP) concerned with parameters which consist of maximizing Entropy Functional subject to constraints dependent on parameters. Therefore, MinMaxEnt and MaxMaxEnt distributions are obtained as solutions of these problems. On bases of MinMaxEnt distribution we have developed a new estimation method of obtaining missing values in time series. Mentioned estimation method is applied to the data generated from autoregressive model for one missing value. The performance of the developed method is evaluated by mean square errors (MSE) calculated from simulation studies and the validity of this method is shown.

Key-words: MinMaxEnt Distribution, MaxEnt Distribution, Time Series with Missing Value, Estimation of Missing value, Autoregressive Modelling

1 Introduction

In our society, we often have to analyse and make inferences using real data that is available for collection. Ideally, we would like to think that the data is carefully collected and has regular patterns with no outliers or missing value. In reality, this does not always happen, so that an important part of the initial examination of the data is to assess the quality of the data and to consider modifications where necessary. A common problem that is frequently encountered is missing observations for time series data. A time series stated as only one realization of a stochastic process is a set of data measured through time. In many areas to economics from engineering is faced with patterns of time series. It is difficult to find a science program not requirement to study with a data set in form of time series. Characteristic property of a time series is that can not be exactly estimated its future behavior. However, missing values is fairly common because of consist in the data depend on time and missing values is a typical distortion of model assumptions in data analysis. Reasons of this problem are listed as the data do not exist at the frequency we wish to observe them; registration errors; deletion of "outliers" [1-3]. Therefore, an important problem in time series analysis is that of estimation of missing values which, for some reason cannot be observed completely. Faced with this situation, several methods of replacement of the missing values were developed in the literature. To include estimations of missing values in time series allow us to forecast better.

One of the key steps in time series analysis is to try to identify and correct obvious errors and fill in any missing observations enabling comprehensive analysis and forecasting. This can sometimes be achieved using simple methods such as eyeballing, or calculating appropriate mean value etc. However, more complex methods may be needed and they may also require a deeper understanding of the time series data. Sometimes, we are required to forecast values beyond, or prior to, the range of known values. To complete this task successfully we need a model which satisfactorily fits the available data even when missing values are present [4].

More complex methods for analyzing time series data will depend on the type of data that we are handling [5,6]. But most of the time, we would use either a deterministic or stochastic approach. Deterministic method assumes the time series data corresponds to an unknown function and we try to fit the function in an appropriate way. The missing observation can be estimated by using the appropriate value of the function at the missing observation. Unlike traditional time series approaches, this method discards any relationship between the variables over time. Another common time series approach for modelling data is to use Box-Jenkins' Autoregressive Integrated Moving. These models use a systematic approach to identify the known data patterns and then select the appropriate formulas that can generate the kind of patterns identified. Once the appropriate model has been obtained, the known time series data can be used to determine appropriate values for the parameters in the model. However, a disadvantage of the Box-Jenkins ARIMA models is that it assumes that data is recorded for every time period. Often time series data with missing values require us to apply some intuitive method or appropriate interpolative technique to estimate those missing values prior to Box-Jenkins's ARIMA approach. Additionally, a new approach to time series analysis is the use of state space modelling. State space modelling emphasises the notion that a time series is a set of distinct components. Essentially, Kalman Filtering and Maximum Likelihood Estimation methods are important procedures for handling state space models. The approach continually performs estimating and smoothing calculations that depend only on output from forward and backward recursions. With modifications on the maximum likelihood procedure, it enables the approach to estimate and forecast for data with missing values [4].

In [4], the application of deterministic and stochastic approaches to modelling time series data with missing value are compared, traditional Box-Jenkins ARIMA and state space models are used to obtain estimates of missing values and for forecasting. Furthermore, in many papers are devoted to incomplete autoregressive time series, the EM-algorithm is used to estimate parameters [7, 8].

In this study, firstly MaxEnt distribution of an observed time series is determined as a multivariate normal distribution that its dimension equal to number of observations. Then the entropy of distribution that has maximum entropy is obtained as a functional which is called entropy optimization functional. Furthermore Generalized Entropy Optimization distributions (GEOD) introduced in [8] by A. Shamilov are defined when moment functions and moment values depend on parameter. Out of GEOD, MinMaxEnt and MaxMaxEnt for MaxEnt Distributions with moment functions and moment values dependent on parameter are defined for the time series with missing values by minimization and maximization of entropy functional. The definition of mentioned distributions concerned with problem of finding Lagrange multipliers as solution of corresponding system of equations. But from mentioned system of equations the determination of Lagrange multipliers is very difficult. This difficulty is overcome asymptotically by applying methods of complex analysis in [9].

By means of relations established between entropy values of MaxEnt, MinMaxEnt and MaxMaxEnt distributions, it is proved that the distribution containing the largest information is MinMaxEnt distribution. Through the result, estimating of missing values is considered respectively as a problem and a method based on MinMaxEnt distribution is developed for solving these problems. Thus, MinMaxEnt Distribution is used for estimation of missing value in time series. In order to use the developed method, the computations are performed by the programmes written in Matlab according to the number of missing values. Performance of estimation method based on MinMaxEnt distribution is evaluated by mean square errors (MSE) calculated on the time series generated from autoregressive model AR(4), by assumption that the value in each position in time series is missing.

Furthermore, MSE's are calculated from simulation studies by using mentioned method to estimate one missing value in time series generated by autoregressive models with lag four, the validity of the new method is shown.

2 MaxEnt Distribution for Observed Time Series

In observed time series represented by the sequence of observations $\left[y_0\, y_1\, y_2 \ldots y_T\right]^T$ is assumed that range of observation is equal. Another assumption for taken time series is that the mean of the observations is zero. This assumption is not restrictive at all, because if the mean

$$\bar{y} = \frac{1}{N}\sum_{j=1}^{N} y_j \tag{1}$$

is non-zero, then the series can be transformed via

$$y_i' = y_i - \bar{y} \tag{2}$$

to another time series with zero mean. Furthermore, it is assumed that stochastic process generating time series is stationary. This assumption is not also restrictive, because non-stationarity of the time series may be removed by a suitable transformation such as data differencing,

$$y_j'' = y_{j+1} - y_j.\ j = 1,2,\ldots,N \tag{3}$$

The differencing operation can be repeated until the data is stationary. In summary then, given a discrete time series y, it should posses the following properties:

- The samples are equispaced in time;
- The series has a zero mean value;
- The stochastic process yielding observations is stationary.

Only when these conditions are satisfied is the following mathematical study of a time series applicable [9].

The important property of a stochastic process is given as up to lag m autocovariances r_k which is covariance between y_j and y_{j+k}:

$$r_k = \mathrm{cov}[y_j, y_{j+k}],\ k = 0,1,\ldots,m$$

$$r_k = E[(y_j^* - \mu_y^*)(y_{j+k} - \mu_y)). \tag{4}$$

Because the process has a zero mean and is stationary, equation (4) is written as the following:

$$r_k = E[y_j y_{j+k}]. \tag{5}$$

For the real time series can be written as

$$r_{-k} = r_k. \tag{6}$$

For the discrete time series up to lag m autocovariances are obtained from the data as follows:

$$\hat{r}_k = \frac{1}{N}\sum_{j=0}^{N-k} y_j y_{j+k}. \tag{7}$$

When the information about the time series is given as autocovariances up to lag m obtained by equation (7), MaxEnt distribution of the reel time series has form as follows:

$$p(\mathbf{y}) = \frac{1}{(2\pi)^{\frac{N}{2}}\left|\mathbf{\Lambda}^{-1}\right|^{\frac{1}{2}}} e^{-\frac{1}{2}\mathbf{y}'\mathbf{\Lambda}\mathbf{y}} \tag{8}$$

Where $p(\mathbf{y})$ is a multivariate normal distribution and $\mathbf{\Lambda}$ is a toeplitz matrix consisting Lagrange multipliers. These multipliers are asymptotically obtained by solution of the following matrix form:

$$\begin{bmatrix} \hat{r}_0 & \hat{r}_1 & \hat{r}_2 & \ldots & \hat{r}_m \\ \hat{r}_1 & \hat{r}_0 & \hat{r}_1 & \ldots & \hat{r}_{m-1} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ \vdots & \vdots & \vdots & \ddots & \hat{r}_1 \\ \hat{r}_m & \hat{r}_{m-1} & \hat{r}_{m-2} & \ldots & \hat{r}_0 \end{bmatrix} \begin{bmatrix} 1 \\ a_1 \\ a_2 \\ \vdots \\ a_m \end{bmatrix} = \begin{bmatrix} 1/\left|g_0\right|^2 \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \tag{9}$$

Furthermore, Lagrange multipliers are determined as the following:

$$\lambda_k = \left|g_0\right|^2 \sum_{j=0}^{m-k} a_j a_{j+k},\ k = 0,1,\ldots,m. \tag{10}$$

Then $\left|\mathbf{\Lambda}\right|$ is depicted as follows:

$$\left|\mathbf{\Lambda}\right| = \prod_{j=1}^{N}\left(\sum_{k=1}^{m} 2\lambda_k \cos\left(\frac{2\pi jk}{T+1}\right) + \lambda_0\right). \tag{11}$$

Therefore, MaxEnt distribution is constituted as the form (8) via equations (10) and (11) [9].

It is known that the entropy value of MaxEnt distribution is

$$H\left(p(y)\right) = \ln\frac{(2\pi e)^{\frac{N}{2}}}{\left|\mathbf{\Lambda}\right|^{\frac{1}{2}}}. \tag{12}$$

where $\left|\mathbf{\Lambda}\right|$ is obtained by formula (11). If we take into account the definition of entropy optimization (EO) functional U, then the equation (12) is considered as EO functional which will be required to define Generalized Entropy Optimization Distributions (GEOD) in Section 3.

3 Generalized Entropy Optimization Distributions

In this section, firstly GEOD formulated in [9] are introduced. Then, special cases of GEOD with moment functions and moment values depend on parameter are investigated. Moreover, MinMaxEnt distribution out of GEO distributions is defined for observed time series with missing value.

In [10], giving set of moment vector functions $g(x)$, according to each measure L special functional $U(g)$ is defined. Via the moment vector functions giving the least and the greatest values to $U(g)$, GEO distributions in the form MinMaxEnt and MaxMaxEnt are obtained. Each of these distributions is closest (or furthest) to the given (from the given) a priori distribution in the corresponding measure.

Let us consider the problem of optimizing entropy optimization measure L subject to constraints,

$$\int_a^b f(x) g_j(x) dx = \mu_j \ , \ j = 0,1,\ldots,m \qquad (13)$$

where $\mu_o = 1$, $g_0(x) = 1$, $1, g_1(x),\ldots,g_m(x)$ are linearly independent moment functions, L is given functional on probability density functions $f(x)$.

If the distribution $f^{(0)}(x)$ and moment vector functions $g(x) = (1, g_1(x),\ldots,g_m(x))$ are given, then one can obtain moment value $\mu = (1, \mu_1,\ldots,\mu_m)$ for this moment vector function. Consequently, optimum value of L can be considered as a functional dependent on moment vector function $g(x)$. We call this functional entropy optimization functional and will use the notation $U(f)$ or $U(g)$ [11].

GEOD: Let $f^{(0)}(x)$ be given probability density function of random variable X, L be an entropy optimization measure and K be a set of given moment vector functions. It is required to choose moment vector functions $g^{(1)}, g^{(2)} \in K$ such that $g^{(1)}$ defines entropy optimization distribution $f^{(1)}(x)$ closest to $f^{(0)}(x)$, $g^{(2)}$ defines entropy optimization distribution $f^{(2)}(x)$ furthest from $f^{(0)}(x)$ with respect to entropy optimization measure L. In other words moment vector functions $g^{(1)}, g^{(2)}$ which give to U the least and the greatest values respectively play important role in definition of special distributions.

Let

$$\min_{g \in K} U(g) = U\left(f^{(1)}(x)\right) = U\left(g^{(1)}\right); \qquad (14)$$

$$\max_{g \in K} U(g) = U\left(f^{(2)}(x)\right) = U\left(g^{(2)}\right). \qquad (15)$$

In fact, If L is chosen as Shannon entropy measure, then $U(g)$ is maximum value of H subject to constraints generated by $g(x)$, in other words is $H_{\max}$. For this reason $f^{(1)}(x)$ is called MinMaxEnt distribution and $f^{(2)}(x)$ is called p.d.f. of MaxMaxEnt distribution.

Let us consider GEOD with finite number of given moment functions and moment values dependent on parameters. In this case condition (13) takes the form

$$\int_a^b f(x) g_j(x,\nu) dx = \mu_j(\nu), \quad j = 0,1,\ldots,m \quad (16)$$

where $g_0(x) \equiv 1, \mu_0 = 1$, ν is scalar or vector parameter, $f(x)$ is unknown p.d.f.

It is required to obtain extremum value of EO measure L subject to constraints (16).

Let $L = H = -\int_a^b f(x) \ln f(x) dx$.

Then EO functional U defined by (12) turns into function on ν:

$$U(\nu) \equiv -\sum_{j=0}^{m} \lambda_j\left(g(x,\nu), \mu(\nu)\right) \mu_j(\nu) \qquad (17)$$

MaxMaxEnt and MinMaxEnt distributions defined **in section 3** can be obtained depending on the greatest and the least values of $U(\nu)$. The continuity of $\lambda_j\left(g(x,\nu), \mu(\nu)\right)$ on ν can be proved by global implicit function theorem [12,13].

4 GEO Distributions for the Time Series with Missing Value

From (9) and (10) follows that $\lambda_k, k = 0,1,\ldots,m$ can be considered as functions of $\hat{r}_0, \hat{r}_1, \ldots, \hat{r}_m$. According to (7) $\hat{r}_0, \hat{r}_1, \ldots, \hat{r}_m$ depend on $y_0, y_1, \ldots, y_T$. For this reason EO functional U obtained by (12) according to (11) and (10) is a function of $y_0, y_1, \ldots, y_T$. This function allows us to define new distributions by consideration as unknown parameters when one or several of $y_0, y_1, \ldots, y_T$ are missing values. Firstly, we assume that one of $y_0, y_1, \ldots, y_T$ is parameter, this parameter denoted as ν. Then EO functional is function of ν, in other words U can be write as $U(\nu)$. We assume that $\nu \in [\alpha, \beta]$. It is easy to see that $U(\nu)$ is continuous with respect to ν. Therefore, the following statements are valid:

$$\min_{\nu\in[\alpha,\beta]} U(\nu) = U(\nu_0); \qquad \max_{\nu\in[\alpha,\beta]} U(\nu) = U(\nu_1). \tag{18}$$

If we consider one of $y_0, y_1, \ldots, y_T$ as parameter $\nu(y_j = \nu)$, then the probability distribution obtained by formula (8) is a function of ν. This function is denoted by $p(\nu)$. Any probability distribution satisfying the same constraints as MaxEnt distribution is denoted by $p_0(\nu)$ or $p_0(y)$.

Therefore, MinMaxEnt distribution for the time series with missing value can be considered as MinMaxEnt distribution with moment functions and moment values dependent on parameter. Because missing value ν involve in moment constraints, entropy optimization functional (12) depends on missing value ν. Thus, entropy optimization functional denoted $U(\nu)$ is minimized with respect to ν and obtained ν_0 generate MinMaxEnt distribution and is maximized with respect to ν and obtained ν_1 generate MaxMaxEnt distribution. The following theorem assert that ν_0 is estimation for the missing value ν.

Theorem 1. Let we are given the real valued stochastic time series $y_0, y_1, \ldots, y_T$; involving unknown element: $\nu = y_k$; $0 \le k \le T$ and $U(\nu), \nu \in [\alpha, \beta]$, $\alpha = \min_{0\le k\le T} y_k; \beta = \max_{0\le k\le T} y_k$, is Entropy Optimization Function defined by (12), where $|\Lambda|$ is defined by (11), λ_k is defined by (10).

Moreover, let $p(\gamma_0)$, $p(\gamma_1)$ are respectively MinMaxEnt and MaxMaxEnt distributions, where

$$p(\nu_0) = p(y)\Big|_{\nu=\nu_0}; \; p(\nu_1) = p(y)\Big|_{\nu=\nu_1}$$

$$\min_{\nu\in[\alpha,\beta]} U(\nu) = U(\nu_0); \; \max_{\nu\in[\alpha,\beta]} U(\nu) = U(\nu_1)$$

Then between entropy values of MinMaxEnt-$p(\nu_0)$, MaxEnt-$p(\nu)$ and MaxMaxEnt-$p(\nu_1)$ distributions the inequalities

$$H(p(\nu_0)) \le H(p(\nu)) \le H(p(\nu_1)) \tag{19}$$

hold.

Proof: $p(y)$ is defined by formula (8), where $|\Lambda|$ and λ_k are obtained by (11) and (10) respectively, if we consider $y_k = \nu$ as parameter, where $\nu \in [\alpha, \beta]$, then $p(y)$ and $U(\nu)$ defined by (12) are continuous functions of ν. Consequently continuous function $U(\nu)$ on the interval $\nu \in [\alpha, \beta]$ reaches its greatest and least values. For this reason, values of γ_0 and γ_1 satisfying (18) exist. In other words the inequalities

$$U(\gamma_0) \le U(\gamma) \le U(\gamma_1) \tag{20}$$

hold.

If we take into account the definition (12) of U, then inequalities (19) are result of inequalities (20). Theorem is proved.

Theorem 2. Let we are given the real time series $y_0, y_1, \ldots, y_T$; involving missing value: $\nu = y_k$, $0 \le k \le T$ and $U(\nu)$, $\nu \in [\alpha, \beta]$, $\alpha = \min_{0\le k\le T} y_k; \beta = \max_{0\le k\le T} y_k$, is Entropy Optimization Function defined by (12), where $|\Lambda|$ is defined by (11), λ_k is obtained by (10).

Moreover, Let $p(y)$ or $p(\nu)$ be MaxEnt distribution depend on ν, where $\nu \in [\alpha, \beta]$,

$\alpha = \min_i y_i$, $\beta = \max_i y_i$. Let $p(\nu_1)$ be MaxMaxEnt and $p(\nu_0)$ be MinMaxEnt distributions such that

$$\max_{\nu\in[\alpha,\beta]} H(p(\nu)) = \max_{\nu\in[\alpha,\beta]} U(\nu) = H(p(\nu_1)) = U(\nu_1)$$

$$\min_{\nu\in[\alpha,\beta]} H(p(\nu)) = \min_{\nu\in[\alpha,\beta]} U(\nu) = H(p(\nu_0)) = U(\nu_0)$$

where $U(\nu)$ is EO functional: $U(\nu) = H(p(\nu))$.

Proof: Let $p_0(\nu)$ be probability distribution of given time series which have entropy $H(p_0(\nu))$. According to Jaynes's concentration theorem, if we take any probability distribution $p_0(\nu)$ satisfying the same constraints as MaxEnt distribution $p(\nu)$, there is a 95% chance that its entropy $H(p_0(\nu))$ greater than $H(p(\nu)) - \frac{\chi^2_\vartheta(0.95)}{2(T+1)}$, and there is a 99% chance that its entropy greater than $H(p(\nu)) - \frac{\chi^2_\vartheta(0.99)}{2(T+1)}$, where $\vartheta = T - m$ is degrees of freedom.

In other words for large values of N entropies of most of the probability distributions satisfying a given set of constraints will be concentrated near the maximum value:

$$H(p(\nu)) - \frac{\chi^2_\vartheta(0.95)}{2(T+1)} \le H(p_0(\nu)) \le H(p(\nu))$$

$$H(p(\nu)) - \frac{\chi^2_\vartheta(0.99)}{2(T+1)} \le H(p_0(\nu)) \le H(p(\nu))$$

Consequently by virtue of these inequalities $H(p_0(\nu))$ asymptotically can be considered as $H(p(\nu))$.

The amount of our uncertainty we need to resolve or information we need to gain about the observed time series $y_0, y_1, \ldots, y_T$ before $p(\nu)$ reflects over state of knowledge is

$$\Delta H = H(p(\nu_1)) - H(p(\nu))$$

or

$$\Delta U = U(\nu_1) - U(\nu)$$

From the Theorem 1 it follows that $U(\nu_0) \le U(\nu) \le U(\nu_1)$. Therefore, according to this inequality, ΔU reaches its maximum value, when $\nu = \nu_0$. Consequently if $\nu = \nu_0$ then $p(\nu_0)$ represents MinMaxEnt distribution and as unknown term of observed time series $y_0, y_1, \ldots, y_T$ can be taken $\nu = \nu_0$ generating the MinMaxEnt distribution $p(\nu_0)$.

Theorem is proved. Then missing value ν can be estimated as the value of $\nu = \nu_0$ generating MinMaxEnt distribution $p(\nu_0)$. Similarly, instead one unknown parameter it is possible to consider vector parameter $\mathbf{v} = (\nu_1, \nu_2, \ldots, \nu_k)$ and define MinMaxEnt and MaxMaxEnt distributions. This theorem holds also when $\mathbf{v}$ is a unknown vector. Therefore, by virtue of MinMaxEnt distribution it is possible to estimate several missing values in time series.

5 An Application of Estimation Method Based on MinMaxEnt Distribution for Missing Value

In application of estimation method based on MinMaxEnt distribution for missing value is used the data set generated from autoregressive model. In autoregressive process developed by Box and Jenkins [14], each observation is made up of a random error component and a linear combination of prior observations. $AR(p)$ models are depicted as follows:

$$x_t = \xi + \phi_1 x_{(t-1)} + \phi_2 x_{(t-2)} + \ldots + \phi_p x_{(t-p)} + \varepsilon$$

where ξ is a constant and $\phi_1, \phi_2, \ldots, \phi_p$ are the autoregressive model parameters, $x_{(t-1)}, x_{(t-2)}, \ldots, x_{(t-p)}$ are prior observations [15].

The used data set generated from $AR(4)$ process as follows:

$$X_t = 2.76X_{t-1} - 3.81X_{t-2}$$
$$+2.65X_{t-3} - 0.92X_{t-4} + \varepsilon_t,\ \varepsilon_t : N(0,1)$$

is given in Table 1.

Table 1 The data generated from AR(4) and MinMaxEnt estimations of missing values in each position

t	$X(t)$	$(MinMaxEnt)_2$
1	-1.9172	-
2	8.8041	-
3	5.8994	5.6387
4	-4.9992	-5.4169
5	-11.4250	-11.8323
6	-3.6937	-3.5559
7	16.4869	16.1702
8	33.3703	32.9922
9	30.1568	29.3102
10	2.4457	2.5585
11	-34.1657	-34.6224
12	-56.8119	-56.7917
13	-47.0279	-46.9790
14	-6.0415	-6.4423
15	43.5992	43.2731
16	72.2938	71.6972
17	60.5811	60.2642
18	12.9017	12.8657
19	-44.9481	-44.9732
20	-80.9485	-80.9485
21	-74.6286	-74.4715
22	-28.4727	-28.3051
23	31.9604	31.2631
24	72.1508	71.8033
25	71.8607	71.3803
26	33.8458	33.5386
27	-18.4783	-18.5549
28	-55.9572	-56.1478
29	-60.8923	-
30	-34.7771	-
	$(MSE)_2$	**0.1294**

Firstly, by using the data in Table 1, estimations based on MinMaxEnt distribution are obtained for missing values in each position given constraints up to lag two autocovariances. These estimations are denoted by $(\text{MinMaxEnt})_2$. It is seen that calculated mean square error (MSE) is quite small. The MSE is the second moment (about the origin) of the error, and thus incorporates both the variance of the estimator and its bias. Thus, we can performance of estimation method based on MinMaxEnt distribution is good.

Secondly, by using the same data set in Table 1, estimations based on MinMaxEnt distribution are obtained for missing values in each position given constraints up to lag three autocovariances. These estimations are denoted by $(\text{MinMaxEnt})_3$.

Table 2 The data generated from AR(4) and MinMaxEnt estimations of missing values in each position

t	$X(t)$	$(MinMaxEnt)_3$
1	-1.9172	-
2	8.8041	-
3	5.8994	-
4	-4.9992	-5.2914
5	-11.4250	-11.8093
6	-3.6937	-3.6118
7	16.4869	16.1592
8	33.3703	33.2251
9	30.1568	29.5001
10	2.4457	2.6745
11	-34.1657	-34.6565
12	-56.8119	-56.8670
13	-47.0279	-47.0769
14	-6.0415	-6.4176
15	43.5992	43.4434
16	72.2938	71.8809
17	60.5811	60.3586
18	12.9017	12.8027
19	-44.9481	-45.0940
20	-80.9485	-80.9485
21	-74.6286	-74.7012
22	-28.4727	-28.3925
23	31.9604	31.3760
24	72.1508	72.0291
25	71.8607	71.5250
26	33.8458	33.5998
27	-18.4783	-18.5836
28	-55.9572	-
29	-60.8923	-
30	-34.7771	-
	$(MSE)_3$	**0.0861**

In these MSE's comparison in Table 1 and Table 2, we can see that MSE of MinMaxEnt estimations with lag three autocovariances $(MSE)_3$ is smaller than $(MSE)_2$ of MinMaxEnt estimations with lag two autocovariances $(MSE)_2$. Thus, we think that this result comes from the fact that $(\text{MinMaxEnt})_3$ shows better performance than $(\text{MinMaxEnt})_2$.

6 Simulation Study

In this section, in order to test effectiveness of this estimation method, we apply mentioned method to simulated data sets derived from same time series model $AR(4)$. One way to obtain such a large amount of data is by simulation using the computer. However, in this study, we have written the programme to estimate of missing values in each position and to calculate the MSE. Estimation process based on MinMaxEnt Distribution with lag two autocovariances is realized one hundred times for the generated data sets of which sample size are 50 under the assumption that the value in each position is taken (accepted) as missing value.

The simulation was carried out by following a number of steps. These were:

1. Generate specific time series model data sets for testing.
2. Take out a single value from the data set and store it at another cell for comparison.
3. Apply the MinMaxEnt estimation method to each value in data set one by one.
4. Calculate the mean square error.
5. Repeat the process one hundred times, determine the MSE for each data set.

The results of simulation study are as follows in Table 3 and Table 4.

According to our results in Table 3 and Table 4, the performance of the developed method is pretty good due to very small MSE values.

Table 3 Calculated MSE Values of 100 Time series generated from AR(4) Model

No	MSE	No	MSE
1	0.0699	**51**	0.2270
2	0.3352	**52**	0.1176
3	0.0906	**53**	0.2740
4	0.1815	**54**	0.3734
5	0.2406	**55**	0.1158
6	0.1940	**56**	0.0735
7	0.1394	**57**	0.1205
8	0.1156	**58**	0.3967
9	0.5051	**59**	0.2581
10	0.0865	**60**	0.0951

Table 4 Calculated MSE Values of 100 Time series generated from AR(4) Model (continue)

No	MSE	No	MSE
11	0.0775	**61**	0.0692
12	0.1982	**62**	0.1343
13	0.0743	**63**	0.0883
14	0.1398	**64**	0.0998
15	0.5759	**65**	0.1356
16	0.1246	**66**	0.5838
17	0.1907	**67**	0.3519
18	0.0688	**68**	0.1029
19	0.1121	**69**	0.1418
20	0.1085	**70**	0.0707
21	0.0699	**71**	0.0925
22	0.0872	**72**	0.2429
23	0.2822	**73**	0.1568
24	0.0793	**74**	0.1428
25	0.0452	**75**	0.2707
26	0.0763	**76**	0.1224
27	0.1029	**77**	0.1871
28	0.1099	**78**	0.0910
29	0.3705	**79**	0.1458
30	0.1619	**80**	0.0605
31	0.0780	**81**	0.1445
32	0.5575	**82**	0.0974
33	0.1527	**83**	0.1649
34	0.1021	**84**	0.1004
35	0.1921	**85**	0.0804
36	0.0978	**86**	0.3792
37	0.5068	**87**	0.1383
38	0.1007	**88**	0.4664
39	0.0805	**89**	0.1780
40	0.0904	**90**	0.1030
41	0.0699	**91**	0.1060
42	0.1066	**92**	0.1581
43	0.1057	**93**	0.0947
44	0.1204	**94**	0.1138
45	0.2319	**95**	0.1410
46	0.2078	**96**	0.1156
47	0.2491	**97**	0.1092
48	0.1083	**98**	0.0939
49	0.1085	**99**	0.1581
50	0.1080	**100**	0.0997

Moreover, It is seen that determined MSE values is very much smaller than 1, while standard deviation of random noise is 1.

6 Conclusion

MinMaxEnt distribution out of GEOD is very important in modeling statistical data. In the present study, by virtue of MaxEnt Distribution for a stationary time series giving autocovariance constraints up to lag m, Generalized Entropy Optimization Distributions are introduced. Then Generalized Entropy Optimization Distributions with moment constraints dependent on parameter are investigated. Furthermore, we have showed that the MinMaxEnt distribution out of GEOD successfully can be applied also to estimating problems for observed time series with missing value.

In this study, mentioned method is applied to the data set generated by $AR(4)$ process under the assumption that the value in each position is accepted as missing value.

However, proposed method can be also used to interpolate and forecast more than one missing values in time series. Furthermore, A comparison of this method with other estimation and interpolation methods can be realized.

References:

[1] R. Little and D. Rubin, Statistical Analysis with Missing Data. Wiley, New York, 1987.

[2] J.L. Shafer, Analysis of Incomplete Data. Chapmen & Hall, London, 1997.

[3] W.H. Greene, Econometric Analysis. Macmillan, New York, 2000.

[4] Chi Fung D. S. "Methods for the Estimation of Missing Values in Time Series," Master Thesis, Cowan University, Western, Australia, 2006.

[5] Shamilov, A., Giriftinoglu, C., "Estimation by combining MinxEnt principle and time series analysis" *WSEAS Transactions on Mathematics*, 5, 2006, 677-683.

[6] A. Shamilov A., Giriftinoglu Ç.,Usta I., Kantar Y.M., "The Applications of Construction of MinxEnt Probability Distributions",WSEAS Transactions on Business and Economics, 5, 2006, 3.

[7] R. Little and D. Rubin, Statistical Analysis with Missing Data. Wiley, New York, 1987.

[8] M. Pourahmadi "Estimation and interpolation of missing values of a stationary time series" *Journal of Time Series Analysis,* 10(2), 1989,149–169.

[9] Kapur, J.N. and Kesavan, H.K., Entropy Optimization Principles with Applications, Academic Press, New York, USA, 1992.

[10] Shamilov, A., "Generalized entropy optimization problems and the existence of their solutions" *Phsica A.,* 382, 2007, 465-472.

[11] Shamilov, A., "A development of entropy optimization methods" *WSEAS Trans. Math.,* 5, 2006,568-575.

[12] Shub, M., "The implicit function theorem revisited". *IBM J. RES. DEVELOP* 38, 1994, 259-264.

[13] Rheinboldt, W. C., Local mapping relations and global implicit function theorems. University of Maryland, 1968.

[14] Box, G. E. P., and Jenkins, G. , *Time Series Analysis: Forecasting and Control*, Holden-Day, 1976.

[15] http://www.statsoft.com/textbook/sttimser html#arima.

[16] Wu, X. "Calculation of maximum entropy densities with application to income distribution", *Journal of Econometrics*, 115, 2003 347-354.

[17] A. Shamilov, Y. Mert Kantar, I. Usta, "On a Functional Defined by Means of Kullback-Leibler Measure and its Statistical Applications", *WSEAS Transactions on Mathematics*, 6, 5, 2006, 671-676.

[18] A. Shamilov, S. Yolacan, "Statistical Structure of Printed Turkish, English, German, French, Russian and Spanish", *WSEAS Transactions on Mathematics*, 6, 5, 756-762,

[19] Kullback, S., *Information Theory and Statistics*, Dover Publications, New York, USA, 1968.

[20] Papoulis, A., *Probability, Random variables, and stochastic processes*, McGraw-Hill, Singapore, USE, 1991.

[21] Cover, T.M. and Thomas, J.A., *Elements of Information Theory*, A Wiley-Interscience Publication, New York, USA, 1992.

[22] Wu, N.,*The Maximum Entropy Method*, Springer, Berlin, Germany 1997.

[23] Priestley M.B., *Spectral Analysis and Time Series,* ACADEMIC PRESS, INC., London, 1981.

[24] Cramer H., *Mathematical Methods of Statistics*, PRINCETON UNIVERSITY PRESS, Princeton, (1966).

[25] Hamaz, A. ve Ibazizen, M., “Comparison of Two Estimation Methods of Missing Values Using Pitman-Closeness Criterion,” *Communications in Statistics-Theory and Methods,* 38,2009, 2210-2213.