JOURNAL OF MECHANICS OF CONTINUA AND MATHEMATICAL SCIENCES
www.journalimcms.org

ISSN (Online) : 2454 -7190 Vol.-15, No.-8, August (2020) pp 151-161 ISSN (Print) 0973-8975

THE COMPARISON OF THE METHODS ESTIMATING THE FRACTIONAL DIFFERENCES OF PARAMETER AND ITS DEPENDENCE ON ESTIMATION THE BEST LINEAR MODEL OF TIME SERIES IN THE ENVIRONMENTAL FIELD

Saad Kazem Hamza[1], Shareen Ali Hussein[2]

[1]Department of statistics, Faculty of Administrative and Economics, University of Baghdad

[2]Department of Quality Assurance and University of Performance

[1]Saad.hamza@coadec.uobaghdad.edu.iq, [2]Shreen.a@uobaghdad.edu.iq

https://doi.org/10.26782/jmcms.2020.08.00014

(Received: June 7, 2020; Accepted: July 29, 2020)

Abstract

This paper exploring the stability to be achieved in the stochastic processes and operations which are called the autoregressive moving average and symbolized by ARMA Model (the roots of the equation should be out of this model circle. Although these models are not stable and become stable after so many conversions and differences. These new models called the autoregressive methods for integrated moving average which is symbolized ARFIMA (p, d, q) and these differences would be integers or fractional numbers. It is worth to be mentioned that the time series which depending on the long term (long memory) so this stability achieved by snapping the fractional differences which are located within the enclosed period (-0.5, 0.5) and are referred shortly ((ARFIMA (p, d, q))). Models which are located within the enclosed period (-0.5, 0.5). This search aims to estimate the parameter of fractional differences (d), three ways by using real data from the Ministry of Environment that include the rates of air pollution in Baghdad City with Nitrogen oxides (NO^2), Ozone (O^3) materials…these ways are: firstly, the way logarithm periodogram chart regression method which is called (Geweke and Porter- Hudak), symbolized (GPH) Secondly, smoothed periodogram regression. Thirdly, the way that called (KASHYAP AND EOM) and it has been used the standard error squares and standard error (SD) as two scale standards among these three ways to estimate the parameter. Akaike standard has been used for choosing the best model of linear models assumed. In this study, we will be dealt with the fractional differences

Keywords : ARFIMA (p, d, q) models, long term memory, smoothed periodogram method, air pollution, spectrum function

I. Introduction

The beginning of concerning the long-term processes was through testing the data of physical science, previously the economists had concerned with and the most famous example was in Water Science included tidal flows and internal flows in tanks that have been documented by Hurts (1951). Articles by Hurts (1951,1956), that analyzed 900-time-series specialized with Earth Nature Science the motivation for this understanding that the river flows and designing of tanks.

Hurts (1951), Mandelbrot and Wallies (1968) they were realized that the time series specialized in temperature and trees belt that shows the long-term memory. It's worth mentioning that the time series which depend on long term memory where the stability has been achieved by taking the fractional differences (d) were being within the enclosed period (-0.5,0.5) by practicing the fractional integrated mixed time series model and referred to it (ARFIMA). Wilkins noted that his research is important because it allows a general class of long memory models to be integral integrally at seasonal frequencies and zero repetitions. "The advantage of this model is that it avoids modifying seasonal data and also allows for" a large range of long memory processes To be integrated into this model and that the maximum potential estimator in the field of replication is easy and appropriate for all parameters in this model. On the practical side, he estimated seasonal ARFIMA models on unemployment rates in the US and UK and found that the data did not show a reflection in the arithmetic mean when taking fractional levels at zero or seasonal frequencies [XIII]. After that [X] presented a new estimate of the semi-linear regression of long-time time series based on the average periodic function, a consistent and unbiased estimate of the spectrum function. The proposed estimate is different from the known method estimate (GPH), and the characteristics of this new estimate have been fixed. The researchers have expanded the application of this methodology to two consistent estimators of the spectrum function, namely the function of the graduated periodic plan and the function of the tapered periodic plan. Monte Carlo's experimental results gave evidence that the methods used had encouraging signs of estimating the fractional difference coefficient. [IV] Presented the self-regression models of the ARFIMA for nominal exchange rates and compared their predictive power with structural models and random walk models. Monthly observations of Canada, France, Germany, Italy, Japan, and the United Kingdom for the period from April 1973 to December 1998. Sowell's most accurate method of estimation was used. The predictive accuracy of long memory models with random walk models and cash models was compared using the Harvey and Lebourance statistical tests, And (Newbold), and the results proved Long memory models are more efficient and accurate models than random walk and cash models.

Lately, researchers [II]. Used the bootstrap method of the periodic scheme to compute the logarithm of the average sample of the spectrum bootstrap to estimate the parameter (d) in ARFIMA models (p, d, q) , This method follows the semi-linear regression method. The Monte Carlo studies have been used to compare the specific sample of the estimated estimator with the semi-formal regression based on the periodic chart function. The researchers estimated that the proposed estimate has an average of error boxes less than the conventional method.

According to that the long term memory models (AFRIMA) which allow to be a value parameter, fractional value they're a part of wide models for the series time compared with short long memory (ARIMA) that would be the value of (d) exemplified by the parameter integers differences, so the models (ARFIMA) that dealt with continuity of data is the best alternative for the (ARIMA) models because the latter being unable to practice the fractional differences. The Fractional integrates producing to the researcher's tools with high accuracy to describe their series time by data processing generally that they are integrated fractional while modeling their data.

II. Theoretical

There are so many methods could be used in valuing the differences of a parameter (d) and we will mention them down below:

First Method (The Method of Logarithm - Periodical Regression)

[VIII]. Suggested this method and they cleared that this method which is called (GPH) could be used also in the cases of small samples and to describe this method the process will be $\{ x_t \}$ subjected to the process ARIMA (p, d, and q) as the following:

$$\emptyset(\mathrm{B})(1-B)^d x_t = \theta(B) a_t$$

$$B_1 x_t = u_t$$

$$u_t = \emptyset^{-1}(B)\theta(B) \in_t \tag{2.1}$$

So, $\theta(\mathrm{B}), \emptyset(\mathrm{B})$ are two- dimensional for the parameters of autoregressive and moving average respectively so the:

$$\emptyset(\mathrm{B}) = 1 - \emptyset_1 \mathrm{B} - \emptyset_2 \mathrm{B}^2 - \dots \emptyset_p \mathrm{B}^p$$

$$\theta(\mathrm{B}) = 1 - \theta_1 B_1 - \theta_2 B_2 - \dots \theta_q B_q$$

$$(1-B)^d = \sum_{j=0}^{\infty} \frac{\Gamma(j-d)}{\Gamma(\mathrm{j}+1)\Gamma(-\mathrm{d})} \mathrm{B}^j$$

$1 - \mathrm{B}$ Is the factor of fractional difference

$(1-B)^d = \sum_{j=0}^{\infty} \frac{\Gamma(j-d)}{\Gamma(\mathrm{j}+1)\Gamma(-\mathrm{d})} \mathrm{B}^j$

(d) is the parameter of fractional differences

$\{\epsilon_t\}$ Is a white noise method with independent symmetric variables (i.i.d) and follow distribution

Where $\{u_t\}$ represents stable $.N\ (0, \sigma_\epsilon^2)$ process with average (O) and variation (σ_u^2) and $f_u(w)$ has spectrum dense when O<w< π

$$f(\omega) = fu(\omega)\left\{2\sin\left(\frac{\omega}{2}\right)\right\}^{-2d} \tag{2.2}$$

The Logarithm takes the spectrum dense f (w)

$$ln\{f(\omega_j)\} = ln\{fu(\omega_j)\} - dln\left\{2\sin\left(\frac{\omega_j}{2}\right)\right\}^2 \tag{2.3}$$

$\omega_j = 2\pi k/n$,k=0,1,….$n/2$ will represented a group of (Fourier harmonic repetitions), and (n) represents the size of sample

And by adding and subtracting the amount $\ln\{fu(0)\}$) to the right side of quotation (2-3) then we have:

$$\ln\{f(\omega_j)\} = \ln\{fu(0)\} - dln\left\{2\sin\left(\frac{\omega_j}{2}\right)\right\}^2 + \ln\{fu(\omega_j)\} - \ln\{fu(0)\} \tag{2.4}$$

$$\ln\{f(\omega_j)\} = \ln\{fu(0)\} - dln\left\{2\sin\left(\frac{\omega_j}{2}\right)\right\}^2 + \ln\{\frac{fu(\omega_j)}{fu(0)})\}. \tag{2.5}$$

After that we add $ln\{I(\omega_k)\}$ to both side of the previous quotation while $I_x(w)$ known as periodical planning for the sample:

$$I_x(\omega\ ) = \frac{1}{2\pi\{R(0)+2\sum_{s=1}^{n-1} R(s)\cos(s\omega\ )\}} \qquad \omega \in [-\pi, \pi] \tag{2.6}$$

And the R(s) defines as the auto variation function of the sample (Reisen, 1993)

By

$$R(s) = 1/n\sum_{i=1}^{n-s}(x_i - \overline{x})(x_{i+s} - \bar{x}) \quad s = 0, \mp1, \dots. \mp(n-1) \tag{2.7}$$

Offsetting the periodical scheme of sample, the quotation (2-6) in the quotation (2-5) we got- [XI]:

$$\boldsymbol{ln\{I(\omega_j)\} + ln\{f(\omega_j)\} = ln\{fu(0)\} - dln\left\{2sin\left(\frac{\omega_j}{2}\right)\right\}^2 + ln\left\{\frac{fu(\omega_j)}{fu(0)}\right\} + ln\{I(\omega_j)\}} \tag{2.8}$$

$$\boldsymbol{ln\{I(\omega_k)\} = ln\{fu(0)\} - dln\left\{2\sin\left(\frac{\omega_j}{2}\right)\right\}^2 + ln\left\{\frac{fu(\omega_j)}{fu(0)}\right\} + ln\{I(\omega_j)\} - \ln\{f(w_j)\}} \tag{2.9}$$

$$\ln\{I(\boldsymbol{\omega_j})\} = \ln\{fu(0)\} - dln\left\{2\sin\left(\frac{\boldsymbol{\omega_j}}{2}\right)\right\}^2 + \ln\left\{\frac{fu(\boldsymbol{\omega_j})}{fu(0)}\right\} + \ln\left\{\frac{I(\boldsymbol{\omega_j})}{f(\boldsymbol{\omega_j})}\right\} \tag{2.10}$$

While the highest rate for j when g (n) the so g (n)/n→0 when ∞→n

W^j is too close to zero $\boldsymbol{\omega_j} \leq \omega_{g(n)}$ when should small so the lowest rate) ωg(n

$\ln\left\{\frac{fu(\omega_j)}{fu(0)}\right\}$ will be too small (negligible) comparing with the other limits in the right side of quotation so the quotation finally approach to the :

$$ln\{I(\boldsymbol{\omega_j})\} \approx ln\{fu(0)\} - dln\left\{2\sin\left(\frac{\omega_j}{2}\right)\right\}^2 + \ln\left\{\frac{I(\omega_j)}{f(\omega_j)}\right\} \qquad (2.11)$$

And it has been known the formula regression equation approaching to:

$$y_j = a + bx_j + e_j \qquad j = 1,2, \dots . g(n) \qquad (2.12)$$

By comparing to the quotation no. (2-12) with quotation No. (2-11) that we obtained note below:

$$y_j = ln\{I(\boldsymbol{\omega_j})\} \quad and \quad x_j = ln\left\{2\sin\left(\frac{\omega_j}{2}\right)\right\}^2 and \quad b = -d \quad e_j = \ln\left\{\frac{I(\omega_j)}{f(\omega_j)}\right\} + c$$

$$a = ln\{fu(0)\} - c$$

$$c = E\left[-\ln\left\{{I(\boldsymbol{\omega_j})}/{f(\boldsymbol{\omega_j})}\right\}\right]$$

When d_ϵ $(-0.5,0.0)$ the factors of series

$$\left[\ln\left\{{I(\boldsymbol{\omega_j})}/{f(\boldsymbol{\omega_j})}\right\}\right] \qquad j = 1, 2, \quad . \ . \quad ., g(n)$$

The both researchers [VII]. Cleared that these factors are distributed as independent random variables for Gumbel with an arithmetic mean of =-0.577261 and variance $\pi^2/6$ by using the small squares method (OLS)

$$\hat{d} = -\frac{\sum_{j=1}^{g(n)}(x_j-\bar{x})y_j}{\sum_{j=1}^{g(n)}(x_j-\bar{x})^2} \qquad (2.13)$$

And variance

$$var(\hat{d}) = \frac{\pi^2}{6\sum_{j=1}^{g(n)}(x_j-\bar{x})^2} \qquad (2.14)$$

Second Method: (Smoothed Periodogram)

We will use in this method an estimator to coordinate the spectrum function which is the smoothed period gram by using the parzen lag window to estimate the basic differences in the equation regression and according to the smoothed periodogram scheme as the following down below:

$$\ln\{f_s(\boldsymbol{\omega_j})\} = \ln\{fu(0)\} - d\ln\left\{2\sin\left(\frac{\omega_j}{2}\right)\right\}^2 + ln\{f_s(\boldsymbol{\omega_j})/f(\boldsymbol{\omega_j}\} + \ln\left\{\frac{fu(\omega_j)}{fu(0)}\right\} \qquad (2.15)$$

While $f_s(\omega_j)$ defined as the smoothed periodogram and the scheme will be:

$$f_s(\boldsymbol{\omega_j}) = \frac{1}{2\pi}\sum_{s=-m}^{m} k\left(\frac{s}{m}\right) R(s)\cos(s\boldsymbol{\omega_j}) \qquad (2.16)$$

And $k(u)=k\left(\frac{s}{M}\right)$ represents the clear slowdown window and this is odd continuous function in the range $-1 < u < 1$ and k(0)=1 and k(-u)=(u) and the parameter (m) that indicates to it with function for n (the size of sample) and choose $n\to\infty, m\to\infty, (m/n)\to 0$

$0 < \beta < 1, M = n^{\beta}$ And to limit the range for j $1 \le j \le g(n)$ and choose g (n) as mentioned in the first method the regression equation as follow:

$$ln\{f_s(\omega_j)\} \approx ln\{fu(0)\} - dln\left\{2sin\left(\frac{\omega_j}{2}\right)\right\}^2 + ln\{f_s(\omega_j)/f(\omega_j\} \qquad (2.17)$$

And to compare the equation (2-17) with equation of simple linear regression with formula (2-12) we find the:

$y_j = ln\{f_s(w_j)\}$ و b=-d و $x_j = ln\left\{2sin\left(\frac{w_j}{2}\right)\right\}^2$

$e_k = ln\{f_s(w_j)/f(w_j\}$ و a= $ln\{fu(0)\}$

By using the method of small squares (OLS) the estimated formula of

Differences of parameter $\hat{d}$ we can have it from the (2-13) and the variation will be in:

$$\mathrm{var}(\hat{d}) \approx 0.53928\frac{M}{n\sum_{j=1}^{g(n)}(x_j - \bar{x})^2} \qquad w \neq 0, \pi$$

When (-0.5, 0.0) d€ $(\{f_s(w_i)/f(w_i\}$ has normal distribution with an arithmetic mean of zero and variation as follow :

$$var[ln\{f_s(w_j)/f(w_j\}] \approx \left[\begin{matrix} 0.539285\left(\frac{m}{n}\right) & w\neq 0,\pi \\ 1.07856\left(\frac{m}{n}\right) & w=0,\pi \end{matrix}\right.$$

Third Method

This method called (Kashyap and Eom), those scientists made this method in 1988 by using this method in the regression equation:

$$lnI_x(\omega_j) = ln\sigma^2 - dln\left\{4sin^2\left(\frac{\omega_j}{2}\right)\right\} + \ln\left\{\frac{I_u(\omega_j)}{\sigma^2}\right\} \tag{2.18}$$

While $\ln I_x(\omega_j)$ represents the periodical scheme to reduce the error and similar to this equation of the simple regression earlier (2-12)

And when we compare the two equations we have the:

$$y_j = lnI_x(\omega_j)$$

$$a = ln\sigma^2$$

$$x_j = ln\left\{4sin^2\left(\frac{\omega_j}{2}\right)\right\}$$

$$b = -d$$

And the output ($\hat{d}$) is extracted from the equation by (OLS) formula (2-13)

III. Experimental Study

Describe the Sample of Research

The special data of the research which is concerned of the environmental pollution in Baghdad City has been taken from the Environment Ministry of the dioxide (NO2) and Ozone material (O3). In 2016-2018 the data of environmental pollution has been recorded more than time in one day a number of views for each series reached to 5000 observations (reading), and at a rate of five readings per day [VII]

Estimation of Fractional Parameter Differences:

By using the statistical program (R) through the statistical packages in the (packages program) so the parameter estimated (d) by three methods mentioned in the second chapter (by noticing the charters of the tow time series search topic before and after taking the differences depending on the fractional differences of parameter that have been estimated by three methods which mentioned in the annex, the results would be as the following:

The Time Series of Dioxide (NO2)

The table No.1 shows the results of the fractional differences of parameter ($\hat{d}$) for time series (NO2) by using the three estimation methods mentioned in the theoretical side, through noticing the results we found the best estimation of smoothed periodogram method that depending on the standard of tradeoff the regression standard (SD) and the error standard (SE) they were the lowest value:

Table No: 1

Methods	GPH	Smoothed Periodogram	Kashyap And Eom
$\hat{d}$ value	0.3331063	0.27051	0.212062
SD	0.8485939	0.3176126	0.05402317
SE	0.07633678	0.04075286	0.0485975

The Time Series of Ozone (O3)

The table No.2 shows the results of the fractional differences of parameter ($\hat{d}$) for time series of (NO3) by using the three estimation methods mentioned in the theoretical side, through noticing the results we found the best estimation of smoothed periodogram method on depending on the standard of tradeoff the regression standard (SD) and the error standard (SE) they were the lowest value:

Table No: 2

Methods	GPH	Smoothed periodogram	Kashyap and Eom
$\hat{d}$ value	0.4498419	0.5118457	0.2863783
SD	0.0848592	0.03174049	0.05402304
SE	0.08949818	0.04501605	0.05697631

The Estimation of the Best Linear Model to Represent the two Time Series (NO2), (O3)

After we reached that the best method to estimate the parameter (d) is (the smoothed periodogram method) for your knowledge the point of bitrate is 0.5 because the regression and the middle curve of error squares in this method lowest than the regression and middle curve of error squares in the previous methods , so it has been taken the basic differences parameter to stable the two time series (NO2),(O3) to choose the best statistical model for (AFRIMA) (p, d, q) the Akaike's information criteria has been used to compare between the supposed statistical models and the results as its shown down below:

Table (No.3) represent the results of two time series by using the Akaike's information criteria

Table No: 3

The degree of auto regression model P The degree of moving media model q	**Series NO2**	**Series O3**
	Akaikla value	
*(p=1,q=1)	14255.19	29691.47
(p=1,q=2)	14260.66	29698.8
(P=1,q=3)	14273.29	29996.19
(p=2,q=2)	14298.11	30069.38
(p=2,q=3)	14299.04	30082.86
(p=3,q=3)	14297.3	30082.86
(p=3,q=1)	14298.11	30066.25
(P=3,q=2)	14296.46	30086.86

And the best supposed model after taking the differences of parameter (d) of the two time series when p=1,q=1 ARFIMA (1,d,1), the best linear model for the time series of dioxide NO2 is (p=1, d=027051, q=1)

While the best linear model of time series O3 is (p=1, d=0.5118457, q=1)

IV. Conclusions

We Found that the best method of smoothed periodical regression which is the most better estimation method among the three ones of fractional estimation of parameter (d) in the model ARFIMA (p, d, q). And It has been cleared that the best linear model of time series NO2 when p=1, q=1, d=0.27051, the linear model is (p=1,d=0.5118457,q=1). Also cleared that dioxide NO2, Ozone O3 has the character in its data of long memory and this is dangerous indicator on the long way for those two materials and their negative impact on environment.

V. Recommendations:

I Estimating the long and unstable memory that time-varying long memory parameter.

II Estimating the long memory in the multivariate –long memory models.

III Putting devices to measure the environment in the public area to indicate the rates of weather pollution.

IV Practicing some of measurements that have been used in some developed countries in environment field by enacting law (limiting the highest level of pollution after that, announcing a formal holiday as similar in the cities where the temperature at highest level.

References

I. Baillie, R. T. (1996). Long memory processes and fractional integration in econometrics. Journal of Econometrics, 73 (1), 5-59. doi:10.1016/0304-4076 (95)01732-1

II. Franco, G. C., & Reisen, V. A. (2007). Bootstrap approaches and confidence intervals for stationary and non-stationary long-range dependence processes. Physica A: Statistical Mechanics and Its Applications, 375 (2), 546-562. doi:10.1016/j.physa.2006.08.027

III. Hassler, U. (1993). Regression of Spectral Estimators with Fractionally Integrated Time Series. Journal of Time Series Analysis, 14 (4), 369-380. doi:10.1111/j.1467-9892.1993.tb00151.

IV. Karemera, D., & Kim, B. J. (2006). Assessing the forecasting accuracy of alternative nominal exchange rate models: The case of long memory. Journal of Forecasting, 25 (5), 369-380. doi:10.1002/for.994

V. KASHYAP, R.L. and EOM, B. (1988) Estimation in long memory time series models J. time series Anal.9, 35-41 long-memory parameter". Department of Statistics, UFES, ES, UFMG, MG Brazil.

VI. Mandelbrot, B. B., & Wallis, J. R. (1968). Noah, Joseph, and Operational Hydrology. Water Resources Research, 4(5), 909-918. doi:10.1029/wr004i005p00909

VII. Ministry of Environment, records Waziriya station, daily readings of the station during the days of actual work for two years 2017.2018, Iraq - Baghdad 0.2019.

VIII. Porter, Hudak, S. (1982). Long – Term Memory: Modelling A simplified Spectral Approach. Unpublished Ph.D Dissertation, University of Wisconsin

IX. Reisen, V. A., Cribari-Neto, F., & Jensen, M. J. (2003). Long Memory Inflationary Dynamics: The Case of Brazil. Studies in Nonlinear Dynamics & Econometrics, 7 (3). doi:10.2202/1558-3708.1157

X. Reisen.V.A. (1993) "Estimation of the fractional difference parameter in the ARIMA (p, d, q) model using the smoothed periodogram". Journal of time series analysis .Vol.15, No.3.

XI. Reisen.V.A. & Franco.G.C. (2006). "Log average sample spectral estimators of long-memory parameter". Department of Statistics, UFES, ES, UFMG, MGBrazil.

XII. Richard T. Baillie (1996) "long memory processes and fractional integration in econometric" department of Economics, Michigan state university, USA.

XIII. Wei, William W. S. – 1990 – (Time Series Analysis) – Addison – Wesley publishing Company p: 278.

XIV. Wilkins, N. (2003). Fractional Integration at a Seasonal Frequency with an Application to Quarterly Unemployment Rates. School of Finance and Economics University of Technology, Sydney, 1-32.