Comparison of Subset and Full Autoregressive Fractional Integrated Moving Average (Arfima) Models

Odusanya, Oluwaseun Ayodeji
Department of Statistics,
D.S Adegbenro ICT Polytechnic, Itori-Ewekoro, Ogun State
E-mail:oloriebiodus@gmail.com
Phone: +2348062739675

Obadina, Olugbenga Gabriel
Department of Mathematical Sciences,
Olabisi Onabanjo University, Ago-Iwoye, Ogun State
E-mail: olugbenga.obadina@yahoo.com
Phone: +2348057235295

ABSTRACT

Several works have been done on Autoregressive Integrated Moving Average (ARIMA) models for capturing short-term behaviour of time series data without considering the long-term occurrences. This study aims at showing that the subset ADFIMA performs better than the full ARFIMA, having removed the redundant parameters. This study therefore, introduces Autoregressive Fractional Integrated Moving Average (ARFIMA) which offers flexibility in the simultaneous modeling of short-term and long-term persistent temporal dependence even between distant observations and compared with its subset. The data used is the Wolfer sunspot data available in Wei (1990). Estimates of the fitted model were used in this study obtained by using Robinson estimator which aid in the estimation of fractional parameter d by obtaining estimates through regression estimation method. The data were analyzed with the use of statistical software called Gretl and OxMatrics 4.10 package respectively. The results also established that subset ARFIMA model behave better and did performed efficiently in residual variance capture as evinced from AIC.
Keyword: Long Memory, ARFIMA Model, Fractional Parameter, Wolfer Sunspot

Keywords: Comparison, Subset and Full Autoregressive Fractional Integrated Moving Average (Arfima) & Models.

iSTEAMS Conference Proceedings Paper Citation Format
Odusanya O.A. & Obadina, O.G. O.O, (2018): Comparison of Subset and Full Autoregressive Fractional Integrated Moving Average (Arfima) Models. Proceedings of the 14th iSTEAMS International Multidisciplinary Conference, AlHikmah University, Ilorin, Nigeria, Vol. 14, No 2, Pp 223-228

1. INTRODUCTION

The Autoregressive Fractional Integrated Moving Average (ARFIMA) (p, d, q) process was first introduced by Granger and Joyeux (1980), and Hosking (1981). The most useful feature for this process is the long memory. This property is reflected by the hyperbolic decay of the autocorrelation function or by the unboundedness of the spectral density function of the process. While in an Autoregressive Moving Average (ARMA) structure, the dependency between observations decays at a geometric rate. The family of autoregressive integrated moving average processes, widely used in time series analysis, is generalized by permitting the degree of differencing to take fractional values. The fractional differencing operator is defined as an infinite binomial series expansion in powers of the backward-shift operator. Fractionally differenced processes exhibit long –term persistence and anti-persistence; the dependence between observations a long time span apart decays much more slowly with time span than is the case with the more commonly studied time series model. Long-term persistent processes have applications in economics and hydrology; compare to existing models of long-term persistence, the family of models introduced here offers much greater flexibility in the simultaneous modeling of short-term and long-term behaviour of a time series.

Long memory, or long term dependence, describes the correlation structure of a series at long lags. If a series exhibits long memory (or the "biased random walk"), there is persistent temporal dependence even between distant observations. Such series are characterized by distinct but nonperiodic cyclical patterns. A stationary stochastic process $\{y_t\}$ is called a long memory process if there exist a real number H and a finite constant C such that the autocorrelation function ρ(τ) has the following rate of decay such that ρ(k) ~ 2 −2 τCH as τ→∞. The parameter H, Hurst Exponent, display the long memory property of the time series. A long memory time series is said fractionally integrated, where the fractional degree of integration d is related to the parameter H, as d = H – 1/2.

Box, Jenkins, and Reinsel (2008) define long memory processes as those with the slower hyperbolic rate of decay, which includes ARFIMA processes with d<0. We follow Box, Jenkins, and Reinsel (2008) and thus call ARFIMA processes for $0.5 < d < 0$ and $0 < d < 0.5$ long memory processes. The primary aim of the research is to show that the subset ARFIMA will perform better than the full ARFIMA, having removed the redundant parameters in the full model.

To achieve the aim, the following objectives are sought:

1. Proposition of ARFIMA model and distribution properties.
2. Estimation of parameters of the proposed full and subset ARFIMA models.
3. Identification of optimal models in the proposed ARFIMA models.

2. METHODOLOGY

2.1 Fractional ARIMA Processes

Given a time series process y_t with autocorrelation function ρ_j at lag j, the process possesses long memory if the quantity $\lim_{n\to\infty} \sum_{j=-n}^{n} |\rho_j|$ is nonfinite (McLeod and Hipel, 1978). In other words, the spectral density $f(\omega)$ is bounded at low frequencies (Baillie, 1996). Long memory models were first applied to econometrics by Granger and Joyeux (1980) and Hosking (1981). Robinson (2004) and Baillie (1996) have provided comprehensive reviews of this topic.

A long memory process is characterized by a slow decay of the autocovariance function of the type $\gamma(\tau) \sim C\tau^{-\alpha}$ with $C > 0$ a constant depending on the process, $0 < \alpha < 1$ and τ large, Fractional ARIMA (p, d, q), are well known examples of long memory processes. Let us first define the fractional difference operator $(1 - B)^d$, with $d \,\epsilon(-0.5, 0.5)$, as the Binomial series expansion:

$$(1 - B)^d \equiv \sum_{j=0}^{\infty} \binom{d}{j} (-1)^j B^j \quad (2.1)$$

With B the backshift operator and with square summable coefficients:

$$\binom{d}{j} (-1)^j = \frac{\Gamma(d+1)(-1)^j}{\Gamma(d-j+1)\Gamma(j+1)} = \frac{\Gamma(-d+j)}{\Gamma(-d)\Gamma(j+1)} \quad (2.2)$$

Here $\Gamma(.)$ denote the Gamma function. A fractional ARIMA (p, d, q) process with p and q nonnegative integers is defined as the stationary solution of the equation

$$\Phi(B)(1 - B)^d(y_t - \mu) = \Theta(B)\varepsilon_t \quad (2.3)$$

With polynomials $\Phi(B) = 1 + \phi_1 B + \phi_2 B^2 + \cdots + \phi_p B^p$ and $\theta(B) = 1 + \theta_1 B + \theta_2 B^2 + \cdots + \theta_q B^q$ and where μ is the process finite mean and ε_t is a Gaussian white noise process with zero mean and variance σ_ε^2. Here we assume $\mu = 0$ without loss of generality. Differencing d times the process produces an ARMA (p, q) model. Fractional ARIMA processes are stationary and invertible. They exhibit positive dependency between distant observations for $0 < d < 0.5$ (long memory), negative dependency for $-0.5 < d < 0$ (intermediate memory) and reduce to short memory ARMA (p, q) processes for $d = 0$. A special class of processes is the fractionally integrated obtained for $p = 0$and $q = 0$, also called fractionally differenced white noise, or $I(d)$, in that differencing d times produces a white noise process.

2.2 Algorithm For Fitting Full And Subset Autoregressive Integrated Moving Average Models

We fit full autoregressive integrated moving average models of various orders and choose that model for which Akaike Information Criterion (AIC) is minimum. Let the order of this full autoregressive integrated moving average model be p+d+q and let the model be $y_t = \varphi_1 y_{t-1} + \cdots + \varphi_{p+d} y_{t-p-d} + e_t - \theta_1 e_{t-1} - \cdots - \theta_q e_{t-q}$, denoted by ARIMA (p,d,q) respectively.

Let the mean sum of squares of the residuals be $\hat{\sigma}_e^{2(1)}$ and its Akaike Information Criterion (AIC) be equal to AIC(1). The estimation of models is done by using Malquardt algorithm and Newton-Raphson iterative method. Having fitted the full model, we can now fit the best subset models by considering the $2^k - 1$ subsets using the fitted models with minimum AIC. We consider the elements of $2^k - 1$ subsets using the approach of Hagan and Oyetunji (1980) and Ojo (2007) and choose that model for which AIC is minimum.

Let the best subset autoregressive integrated moving average model be

$$y_t = \varphi_{n1} y_{t-n1} + \cdots + \varphi_{nl+d} y_{t-nl-d} + e_t - \theta_{k1} e_{t-k1} - \cdots - \theta_{kq} e_{t-kq},$$

Where
$n_1, n_2, \dots, n_{l+d}; k_1, \dots, k_q$ are subsets of the integers (1,2,…,p+d+q). Let the mean sum of squares of the residuals be $\hat{\sigma}_e^{2(2)}$ and AIC value be AIC(2); $\text{AIC}(2) \leq \text{AIC}(1)$. This is our final subset autoregressive integrated moving average models

3. RESULTS AND DISCUSSION

3.1 Data Used

To present the application of the models proposed, we will use a real time series dataset, the Wolfer sunspot, available in Wei (1990). As the Wolfer sunspot dataset represents a non-stationary series. The research work considered $t = 250$.

3.2 Results

The examination of the time plot of the Wolfer sunspot shows that there is some kind of non-stationarity in the observed data.

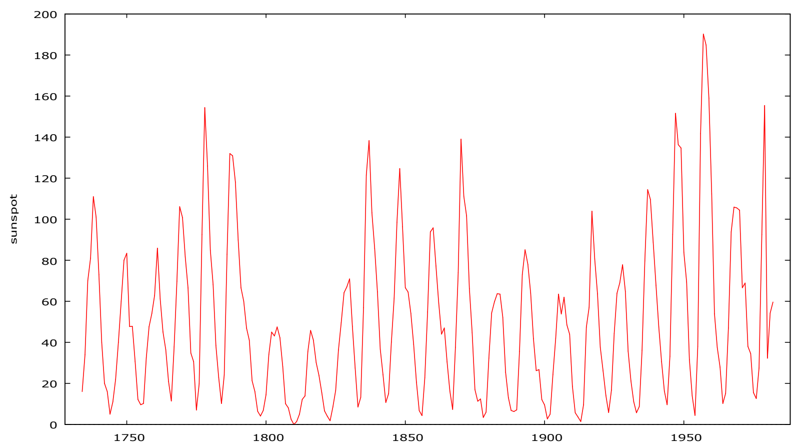

Figure 1:Time Plot of Wolfer sunspot, 1734-1983

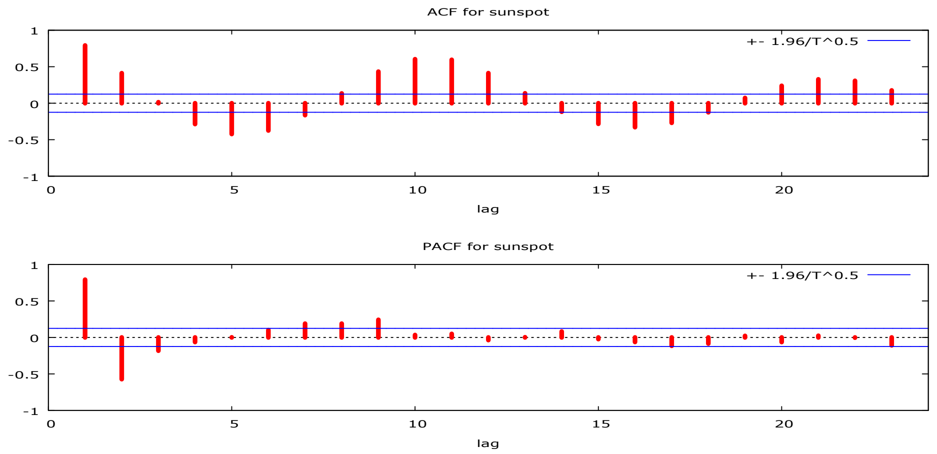

Figure 2: The Correlogram of Wolfer sunspot

Assuming the order of integration d is an integer, the ADF test was carried out on the series as shown in Table 1. The test was unable to detect the non-stationarity in the series at 1%, 5% and 10% level of significance.

Table 1: ADF test for Wolfer sunspot

ADF Test Statistic	-7.889056	1% Critical Value*	-3.9987
		5% Critical Value	-3.4294
		10% Critical Value	-3.1379

Table 2: KPSS test for Wolfer sunspot

KPSs Test Statistic	0.00978058	10% Critical Value*	0.347
		5% Critical Value	0.463
		2.5% Critical Value	0.574
		1% Critical Value	0.739

Both ADF and KPSS test showed that the data is stationary.
A two dimemsional grid values of (p,q) was set up with maximum values (p,q)= (10,10) and a search over all the constituent models was undertaken using the AIC to select the best fitting model. The best model AFRIMA (3,d,2) model with parameters as shown in the table below written as $\Delta^d X_t = \varphi_1 \Delta^d X_{t-1} + \varphi_2 \Delta^d X_{t-2} + \varphi_3 \Delta^d X_{t-3} + \varepsilon_t + \theta_1 \Delta^d \varepsilon_{-1} + \theta_2 \Delta^d \varepsilon_{t-2}$

Table 3 Full ARFIMA Model Estimation for Wolfer sunspot

	Estimates	Std. Error	t-value	Prob.
d parameter	0.0935565	0.1577	0.593	0.554
AR-1	2.50148	0.07494	33.4	0.000
AR-2	-2.35775	0.1261	-18.7	0.000
AR-3	0.823482	0.07188	11.5	0.000
MA-1	-1.62721	0.08015	-20.3	0.000
MA-2	0.752462	0.07189	10.5	0.00
Constant	50.6535	6,903	7.34	0.000
Log-Likelihood	-1077.28737		AIC.T	2170.57474
No. Of observation	250		AIC	8.68229898
No. Of parameters	8		Mean	49.9176
Var	1523.94		Std. dev.	17.8872

Residual var	319.95			

The full ARFIMA model is given as:

$$\Delta^{0.0935}X_t = 2.50148\Delta^{0.0935}X_{t-1} - 2.35775\Delta^{0.0935}X_{t-2} + 0.823482\Delta^{0.0935}X_{t-3} - 1.62721\Delta^{0.0935}\varepsilon_{-1} + 0.752462\Delta^{0.0935}\varepsilon_{t-2}$$

In fitting the subset AFRIMA models to the dataset, there are $2^4 - 1 = 15$ possible subsets. The choice of the order is made on the basis of the minimum AIC having fitted the 15 possible subsets, it was found that the AIC is minimum in the model below:

Table 4: Subset ARFIMA Model Estimation for Wolfer sunspot

	Estimates	Std. Error	t-value	Prob.
d parameter	0.09243	0.2177	0.483	0.423
AR-1	2.6281	0.0494	32.6	0.000
AR-3	0.71962	0.06188	12.8	0.000
MA-1	-1.71345	0.0915	-22.8	0.000
MA-2	0.85862	0.0639	9.35	0.00
Constant	48.36535	7.036	7.46	0.000
No. Of observation	250			
No. Of parameters	8			
AIC	8.673276			
Residual var	317.76			

The Subset ARFIMA model is given as:

$$\Delta^{0.09243}X_t = 2.6281\Delta^{0.09243}X_{t-1} + 0.71962\Delta^{0.09243}X_{t-3} - 1.71345\varepsilon_{-1} + 0.85862\Delta^{0.09243}\varepsilon_{t-2}$$

4. CONCLUSION

It is quite clear that the subset AFRIMA model performed better than the full AFRIMA model having removed the redundant parameter in the full model. Though with the testing of all the subsets of the best model, as long as this will reduce the residual variance then it is not too much to consider the subset models especially when dealing with non-stationary series. Finally, the real data used in this study is a type well known data that is non-stationary. Since there is an improvement in our model using this data, then when a non-stationary data is encountered in different fields, we can as well go ahead and estimate the subset model. The residual variance of the subset ARFIMA model is smaller than the full model.

REFERENCES

1. Box, G. E. P., G. M. Jenkins, and G. C. Reinsel. (2008). Time Series Analysis: Forecasting and Control. 4th ed. Hoboken, NJ: Wiley.
2. Granger C.W .J. and R. Joyeux, (1980). An introduction to long-memory time series models and fractional differencing- J. Time Series Anal., 15-29.
3. Hagan, V. and Oyetunji, O.B. (1980).On the Selection of Subset Autoregressive Time Series Models.UMIST Technical Report No. 124 (Dept. of Mathematics).
4. Hosking, J.R.M. (1981). Fractional Differencing. Biometrika 68(1), 165-176.
5. McLeod, A. &Hipel, K.W. (1978). Preservation of the rescaled adjusted range: a reassessment of the Hurst phenomenon, Water Resources Research 14(3), 491-508.
6. Ojo, J.F. (2007) On the Estimation and Performance of Subset Autoregressive Moving Average Models. European Journal of Scientific Research, Vol 18, 14: 700-706.
7. Robinson, P.M. (2004). Efficient tests of nonstationary hypotheses. Journal of American Statistical Association, 80(428), 1420-1437.
8. Wei, W.S (1990). Time Series Analysis.Univariate and Multivariate Methods. Addison-Wesley Publishing Co. Inc.