ON THE AUTOREGRESSIVE FRACTIONAL UNIT INTEGRATED MOVING AVERAGE (ARFUIMA) PROCESS

Olanrewaju I. Shittu and OlaOluwa S. Yaya
Department of Statistics, University of Ibadan, Nigeria

ABSTRACT
This paper presents a nonstationary fractional unit integrated moving average process that can model time series data that are not stationary after the first difference. This is autoregressive fractionally unit integrated moving average, \( ARFUIMA(p,d,q) \) process with \( 0.5 < d < 2.5 \). The classical unit differencing of Box-Jenkins is combined with the semi-parametric approach to estimate the fractional difference parameter. The model when applied on quarterly Nigerian gross domestic products (GDP) series indicates that ARFUIMA model is better than the corresponding autoregressive integrated moving average (ARIMA) model when compared based on diagnostic tests and forecast.

Keywords: Fractional unit integration, ARFIMA, semi-parametric estimation.

INTRODUCTION
It has been observed that several economic and financial data have been modelled on the assumption that differencing parameter is usually an integer. Although models obtained from integer differencing have been found to be fairly adequate and reliable, however, better models with higher forecast performance can be obtained if appropriate fractional difference parameter is used.

When fractional differencing parameter is non-zero, non-stationarity is suspected. This implies there is strong dependence between distant observations. A number of studies have been carried out on this subject and these include studies on real national product (Diebold et al., 1989); consumer and wholesale price (Geweke and Porter-Hudak, 1982) and stock market prices (Lo, 1991).

The motivation for this paper and presentation is derived from Fatoki, et al., (2010) who investigated the annual GDP from 1980 to 2007. The series was found to be integrated of order 2 \( I(2) \) and autoregressive integrated moving average, \( ARIMA(1,2,1) \) was fitted using the usual model identification and order determination tools. We are of the opinion that further study could reveal a possibility of the series being integrated of a fractional order and hence could lead to better forecast performance. It is observed that classical ADF unit root test may not give conclusive remark on fractional difference but a test designed for this purpose such as Kwiatkowski, Phillips, Schmidt and Shin (KPSS) test could. Detail about KPSS test can be found in Kwiatkowski, Phillips, Schmidt and Shin (1992). Following Shittu and Yaya (2009), an improved model in terms of parameter estimates and forecasts can be obtained by modelling the remaining long memory in a series after the first or second difference. This paper therefore proposes a class of
nonstationary \( ARFUIMA(p, d, q) \) where \( 0.5 < d < 2.5 \) process which cater for series that may still be fractionally differenced after first or second differences. The remaining part of this paper is structured as follows: Section 2 presents the model; Section 3 discusses the estimation method and forecast evaluation; Section 4 presents the data analysis results and discussion and Section 5 concludes the work.

THE ARFUIMA PROCESS

Let \( X_t \) be any time series process. Fractionally Unit Integrated (FUI) be defined as

\[
y_t = (1-L)^{d_0} X_t
\]

(1)

where \( X_t \) is the nonstationary time series, \( y_t \) is the covariance stationary process and \((1-L)^{d_0}\) is the difference operator with the fractional unit difference parameter, \( d_0 \).

**Proposition 1**

Suppose a series is nonstationary and can be expressed as \( y_t = (1-L)^{d_0} X_t \). If an ADF test of unit root confirmed that the series is \( I(u) \), if further fractional differencing of \( I(d) \) for \((-0.5 < d < u)\), then the resulting series of \( I(d_0) \) where \( d_0 = d + u \) is a stationary series.

A fractionally integrated series is invertible and stationary when \( d \) is in the interval \((-0.5 < d < 0.5)\) and nonstationary when \( d > 0.5 \) (Sowell, 1992). When the Augmented Dickey Fuller (ADF) test indicates stationarity at \( d = 1 \) or \( d = 2 \), it implies that \( d > 2 \). When this situation arises, Robinson (1995) and Velasco (1999) indicated that the fractional difference parameter can be obtained from the differenced series. Therefore, setting

\[
d = d_0 - u
\]

(2)

where \( u \) is the unit difference parameter, then equation (1) can be re-written as:

\[
y_t = (1-L)^{d_0} X_t \]

(3)

\[
= (1-L)^{u+d} X_t
\]

(4)

\[
= (1-L)^u (1-L)^d X_t
\]

(5)

where \( u \) is the unit difference parameter \((u = 1, 2, \ldots)\) and \( d \) is the fractional difference parameter \((-0.5 < d < 1)\).

From (3), additivity of difference parameters holds and binomial theorem applied on the difference operator results into complex function. Using the Binomial expansion, the fractional difference operator in equation (1) becomes,

\[
(1-L)^{d_0} = \sum_{k=0}^{\infty} \binom{d_0}{k} (-1)^k L^k
\]
\[ y_i = \sum_{k=0}^{\infty} \frac{\Gamma[-(u + d) + k]}{\Gamma[-(u + d)] \Gamma(k + 1)} X_{t-i} \] (6)

The above expression will hold under the assumption that the fractionally differenced series is still stationary in the interval \((-0.5 < d_0 < u + d)\). The process in (4) can then be re-written as,

\[ y_i = \sum_{k=0}^{\infty} \frac{\Gamma[-(u + d) + k]}{\Gamma[-(u + d)] \Gamma(k + 1)} X_{t-i} \] (7)

When \(u = 1\),

\[ y_i = \sum_{k=0}^{\infty} \frac{\Gamma[-(1 + d) + k]}{\Gamma[-(1 + d)] \Gamma(k + 1)} X_{t-i} \] (8)

When \(u = 2\),

\[ y_i = \sum_{k=0}^{\infty} \frac{\Gamma[-(2 + d) + k]}{\Gamma[-(2 + d)] \Gamma(k + 1)} X_{t-i} \] (9)

Thus \(u\) can take values 1, 2, 3, 4, ..., however it rarely gets beyond 2.

**METHODOLOGY**

We consider estimation of fractional difference parameter, \(d_0\) in the \(ARFIMA(p, d_0, q)\) model using the frequency domain approach described in Geweke and Porter-Hudak (1983) (GPH) and applied in Yaya and Shittu (2010). The definition used in (4) above still applies as 'differencing and adding back' method (Velasco, 2005). The nonstationary series, \(X_t\) is then differenced, \(u\) times in order to guarantee that the true \(d\) is in \(-0.5 < d < u\) according to proposition 1.

Once the value of \(d\) is determined, \(y_i\) is approximated by using \(\hat{d}_0 = \hat{u} + \hat{d}\) in,

\[ \tilde{y}_i = \sum_{j=0}^{\infty} \frac{\Gamma[-(\hat{u} + \hat{d}) + j]}{\Gamma[-(\hat{u} + \hat{d})] \Gamma(j + 1)} X_{t-i} \]
\[ = \sum_{k=0}^{\infty} \frac{\Gamma[-(\hat{u} + \hat{d}) + t + k]}{\Gamma[-(\hat{u} + \hat{d})] \Gamma(t + k + 1)} X_k \] (10)
Forecasts Evaluation

Evaluation of forecast performance of the models will be judged on Meese and Rogoff (1988), MR statistic. The criterion uses the ratio of the root mean square prediction error (RMSPE) of one of the models to the model to the given base model to check the statistical significance. The MR statistic defined as:

\[
MR = \frac{\bar{S}_{UV}}{\sqrt{\frac{1}{n^2} \sum_{j=1}^{n} u_j^2 v_j^2}}
\]

is asymptotically normally distributed with mean zero and variance one where \( n \) is the number of forecasts generated, \( u \) and \( v \) are transformed functions of forecast errors of the two models; \( \bar{S}_{UV} \) is the sample covariance of the means of \( u \) and \( v \), approximated by \( \frac{1}{n} \sum_{j=1}^{n} (u_j - \bar{u})(v_j - \bar{v}) \) where \( \bar{u} = \frac{1}{n} \sum_{j=1}^{n} u_j \) and \( \bar{v} = \frac{1}{n} \sum_{j=1}^{n} v_j \) with \( u_j = e_{1j} - e_{2j} \) and \( v_j = e_{1j} + e_{2j} \) in which \( e_{2j}, i = 1, 2 \) is the \( j^{th} \) forecast error of the model \( i \) and \( n \) is the number of forecasts. The null hypothesis of MR statistic is \( \text{cov}(U, V) = 0 \). When \( MR > Z_{\alpha/2} \), null hypothesis is rejected. This implies that forecast accuracy in the first model is significantly better than the second model. The test is most reliable when \( n \) is large.

RESULTS AND DISCUSSION

Quarterly Nigeria Gross Domestic Products (GDP) data were used to illustrate the proposed model. The data span from 1960 to 2008 with 196 data points. Even though Fatoki, et al., (2010) used annual data on GDP (1980-2007) i.e. 28 data points. Preliminary analysis given in Table 1 below shows that GDP is positively skewed (Skewness=2.5249) and heavy tailed (Kurtosis= 8.5559).

Table 1: Descriptive analysis on GDP series

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>Median</th>
<th>Maximum</th>
<th>Minimum</th>
<th>Std. Dev</th>
<th>Skewness</th>
<th>Kurtosis</th>
<th>JB</th>
<th>Prob.</th>
</tr>
</thead>
<tbody>
<tr>
<td>GDP</td>
<td>667758.1</td>
<td>15006.6</td>
<td>6376225</td>
<td>497.7</td>
<td>1407810</td>
<td>2.5249</td>
<td>8.5559</td>
<td>460.351</td>
<td>0.0000</td>
</tr>
</tbody>
</table>

The time plot presented in Figure 1 shows that there has been astronomical increase in the Nigeria GDP series.
The data will be modelled in two scenarios. First the data will be modelled as ARIMA(p, d, q) and secondly as ARFUIMA(p, d0, q) model. Then their forecast performance will be examined with a view to determining the better model on the basis of the RMS forecast error.

From Table 2, the hypothesis of unit root of the augmented Dickey Fuller (ADF) test is not significant at level and first differenced series of the Nigerian GDP at 5% level but after second difference, the series is stationary which implies that Nigerian GDP is \( I(2) \) series. The second difference series is further subjected to KPSS test of long memory and at 5% level of significance, null hypothesis of series stationarity is accepted again the alternative of long memory.
The GDP series attains stationarity after second difference, \( I(u = 2) \). Using the model identification tools of ACF and PACF and other diagnostic tools, the most appropriate model for the data ARIMA (4,2,0) model with minimum AIC of 25.53 as shown below.

**Table 3: Estimation of Parameters of Subset of ARIMA (4, 2, 0) Model**

<table>
<thead>
<tr>
<th>Estimators</th>
<th>Coefficient</th>
<th>Standard Error</th>
<th>t-probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \hat{\mu} )</td>
<td>2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \hat{\phi}_1 )</td>
<td>-0.986847</td>
<td>0.0709</td>
<td>0.0000</td>
</tr>
<tr>
<td>( \hat{\phi}_2 )</td>
<td>-1.109800</td>
<td>0.0813</td>
<td>0.0000</td>
</tr>
<tr>
<td>( \hat{\phi}_3 )</td>
<td>-0.863167</td>
<td>0.0836</td>
<td>0.0000</td>
</tr>
<tr>
<td>( \hat{\phi}_4 )</td>
<td>-0.166294</td>
<td>0.0743</td>
<td>0.0260</td>
</tr>
</tbody>
</table>

\[
(1-L)^2 X_t = -0.986847(1-L)^2 X_{t-1} -1.109800(1-L)^2 X_{t-2} -0.863167(1-L)^2 X_{t-3} -0.166294(1-L)^2 X_{t-4} + \epsilon_t
\]

Log-likelihood -2471.46903  Skewness 0.48381
AIC 25.5306085  Excess Kurtosis 12.111
Variance of Residuals 6.68222E+09  Normality test 278.06 [0.0000]**
Portmanteau test 32.597 [0.0002]**  ARCH test 7.2796 [0.0076]**

The nonlinear estimation with PcGive version 10.0 following the GPH approach estimated \( d = 0.7539 \) with standard error of 0.06120. The value of \( d \) is within the confidence interval of [0.633987, 0.873891]. The estimated fractional unit integrated parameter is obtained using (2) as \( d_0 = 1.7539 \) and the resulting series is assumed to be in the stationary fractionally integrated range according to proposition 1. Rescaled Statistics (RS) approach employed on the second unit difference \( I(u = 2) \) series of the GDP data computed \( d \) as -0.1350, which implies \( d_0 = 1.8650 \).
The estimated ARFUIMA (3, 1.7539, 0) model is presented below:

\[
(1-L)^{1.7539} X_t = 25218.1 - 0.789660 (1-L)^{1.7539} X_{t-1} - 0.916117 (1-L)^{1.7539} X_{t-2} - 0.662994 (1-L)^{1.7539} X_{t-3} + \epsilon_t
\]

Log-likelihood: -2540.0045
Skewness: 1.1124
Excess Kurtosis: 11.661
Variance of Residuals: 6.32124E+09
Normality test: 183.86 [0.0000]**
ARCH test: 8.9772 [0.0000]**

Table 5: Forecasts of ARFUIMA model based on Modified GPH estimation Approach of Fractional Difference

<table>
<thead>
<tr>
<th>Horizon</th>
<th>ARIMA</th>
<th>ARFUIMA</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Forecasts (in Millions Naira)</td>
<td>Std. Error (in Millions Naira)</td>
</tr>
<tr>
<td>2009Q1</td>
<td>6289141.0</td>
<td>81745</td>
</tr>
<tr>
<td>2009Q2</td>
<td>6618196.8</td>
<td>114850</td>
</tr>
<tr>
<td>2009Q3</td>
<td>7132262.6</td>
<td>115380</td>
</tr>
<tr>
<td>2009Q4</td>
<td>7207918.4</td>
<td>119200</td>
</tr>
<tr>
<td>2010Q1</td>
<td>7172164.2</td>
<td>125370</td>
</tr>
<tr>
<td>2010Q2</td>
<td>7504010.0</td>
<td>134430</td>
</tr>
<tr>
<td>2010Q3</td>
<td>7944395.8</td>
<td>138500</td>
</tr>
<tr>
<td>2010Q4</td>
<td>8038781.6</td>
<td>140060</td>
</tr>
<tr>
<td>2011Q1</td>
<td>8055389.4</td>
<td>141920</td>
</tr>
</tbody>
</table>

Using the Meese and Rogoff (MR) (1988), who developed the MR-statistic as reviewed in Section 4:

\[ H_0: \text{Forecast Errors are the same (Models are identical)} \]
\[ H_1: \text{Forecast Errors are not the same (Models are not identical)} \]
Since the computed $MR=2.186$ is greater than $Z_{0.025}=1.96$, we reject the null hypothesis of equality of forecast errors for the two models. Thus, the estimated models are not identical. The ARFIMA $(10,2.27,0)$ is considered considering the smaller MAD of $571110.1$ against $677692.7$ of ARIMA $(1,2,1)$ model.

**CONCLUSION**

We have considered in the paper autoregressive fractionally unit integrated moving average and (ARFIMA) model. Its performance was compared with the ARIMA($p,d,q$) model in terms of model adequacy and forecast performance. The ARFIMA model performed better when applied to GDP data than the ARIMA model as indicated by smaller mean squared error of forecast.

**REFERENCES**


**ABOUT THE AUTHORS:**

Olanrewaju I. Shittu and OlaOluwa S. Yaya: Department of Statistics, University of Ibadan, Nigeria