The Philippine Statistician, 1995-1996
Vol. 44-45, Nos. 1-8, pp. 33-48

On the Application of Autoregressive Fractionally-Integrated Moving Average (ARFIMA) Model to Philippine Stock Exchange Oil Index

Antonina A. Binsol and **Daniel C. Bonzo**[1]

ABSTRACT

A systematic procedure of modeling a persistent time series is presented. The methodology which is known as ARFIMA (Autoregressive Fractionally-Integrated Moving Average) modeling is an extended version of Box-Jenkins or ARIMA iterative modeling. It involves three main stages, namely estimation (degree of differencing, autoregressive and moving-average parameters), diagnostic-checking and forecasting. The data set that was used in the application was the Philippine Stock Exchange (PSE) oil index daily series which was selected after an investigation of available economic variables. The use of fractional differencing in capturing long-range dependence of PSE in the years oil index 1994, 1993-94 and 1992-94 series was found to be more appropriate than integral differencing. Persistence in the series was evident in the autocorrelation plots which displayed a hyperbolic decay pattern. Furthermore, there is a decreasing persistence-degree pattern as the number of years decreases based on the fractional differencing degree estimates computed by using a modified version of the Hui-Li algorithm. In the final stage, the three periods were compared and the ARFIMA model fit for the one-year series provided the best set of forecasts.

KEY WORDS: Fractional differencing, Hyperbolic decay, Long-range dependence, Persistence

1. INTRODUCTION

Many time series analytical approaches focus on the traditional short-term memory models. The consideration herein is that distant observations are nearly independent. However, this is not always the case as proven in empirical studies done in the fields of hydrology and economics. Long-term dependent series display persistence. This is a characteristic showing correlations between distant observations to be slowly deteriorating which is uncommon in short-term dependent series. The univariate Box and Jenkins (UBJ) ARIMA approach could be generalized in order to model persistent series. The generalization calls for the estimation of the degree of differencing prior to ARMA fitting. The idea is called fractional differencing and the by-product model is known as an ARFIMA (Autoregressive Fractionally-Integrated Moving Average) model.

Modeling persistence through ARIMA approach poses three problems as discussed by Sowell (1992a). The first one is concerned with estimates of the parameter values. In an ARMA(p,q) model, the roots of the autoregressive and the moving-average polynomials are assumed to be outside the unit circle. However, for some applications the estimates are near the boundary of the unit circle. This implies that asymptotic distributions following the ARMA(p,q) representation may lead to poor approximations of the sampling distributions. To model positive dependence, a root of the autoregressive polynomial must approach the

[1] A. A. Binsol, Yokohama National University, Yokohama, Japan and D. C. Bonzo, UP Statistical Center, Diliman, Quezon City

unit circle. To model negative dependence, a root of the moving-average polynomial must approach the unit circle.

The second problem in using ARIMA models is that if an AR or MA parameter does capture the long-run behavior of a series then restrictions are imposed on the short-run behavior of the series. As a possibility, the long-run behavior is dominating in the series.

The last problem of ARIMA model application is concerned with model selection in small samples. There is no way to direct the fit of an AR or an MA parameter to the long-run characteristic of a series because it is being sacrificed just to obtain a better fit of the short-run behavior. F. Sowell mentioned two approaches to avoid these problems. He suggests reliance on nonparametric estimation and extensive parametric estimation techniques. However, the nonparametric estimates are too imprecise to give meaningful restrictions considering the sample sizes of econometric series being studied.

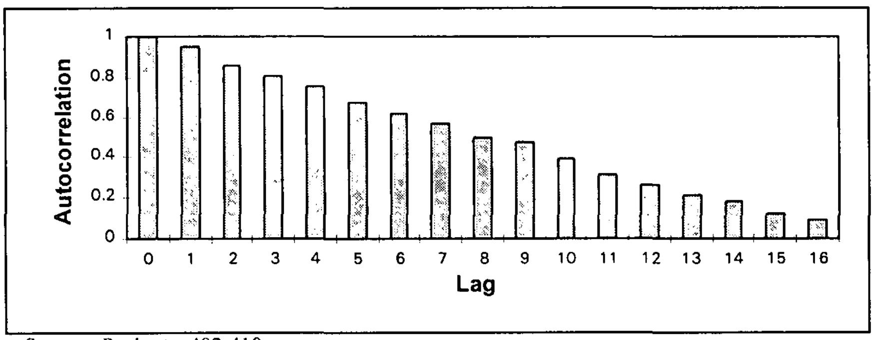
The consideration of a more general class of parametric models which will be less susceptible to these problems is needed. Such model is the Autoregressive Fractionally-Integrated Moving Average (ARFIMA) model represented by

$$(1+\phi_1B+\phi_2B^2+\phi_3B^3+\ldots+\phi_pB^p)\,(1-B)^d\,X_t = (1+\theta_1B+\theta_2B^2+\theta_3B^3+\ldots+\theta_qB^q)\,\varepsilon_t$$

where it will be assumed that $d < 0.5$ and that the roots of the autoregressive and the moving-average polynomials are outside the unit circle. In the model above, the AR and MA parameters seek to capture the short-run behavior of the series, while the differencing parameter seeks to explain long-range dependence.

Hosking (1981), Porter-Hudak (1990) and Beran (1992) characterized persistence or long-term dependence in a series by a hyperbolic decay-pattern of spikes representing the autocorrelation function values. This pattern displays a much slower decay of spikes than the known exponential-decay pattern. Figures 1(a) and 1(b) show the exponential decay pattern and hyperbolic decay pattern respectively. Beran (1992) pointed out that the two best known classes of stationary processes with this characteristic are increments of self-similar or fractional Gaussian Noise processes and fractional ARIMA processes.

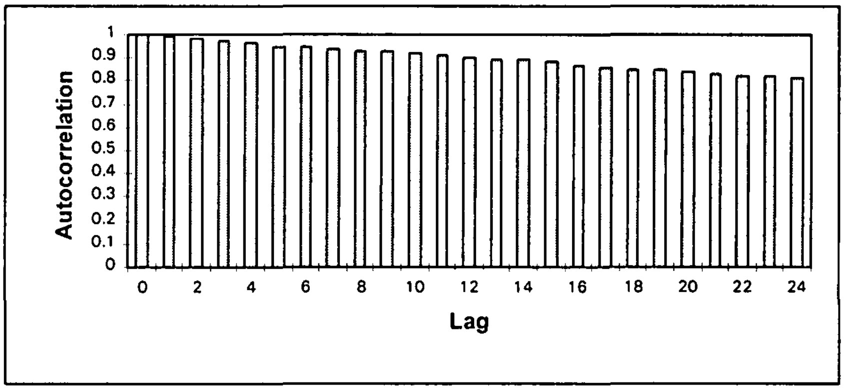
Figure 1a
ACF of AT&T Stock Price (Undifferenced)



Source: Pankratz, 402-410

ARFIMA modeling applications were vastly done in hydrology. The long-term persistence phenomenon was discussed by Lawrance and Kottegoda (1977) in the riverflow time series analysis. Self-similarity is a characteristic inherent in this type of series which is a form of statistical invariance with respect to the change in time scale. Long-term series were preferred over short-term series in order to reproduce distributions of deficits and durations of extreme overflows. Most hydrologists believe that long-range dependence is a rule rather than an exception. Hydrologic time series like amount of rainfall, lake levels and riverflows are known to be persistent series. Such series contain more likely extreme events which must be treated with caution.

Figure 1b
ACF of PNB-Mkt 1990 Stock Prices (Undifferenced)



The discussions provided by Davies and Harte (1987) show that if $X_1, ..., X_n$ denote a series of observations that are normally distributed and considered to test for long-term dependence then $\{X_n\}$ is known as fractional Gaussian noise if $cov(X_i,X_j) = \theta \; \rho(i - j)$ where $\rho(i) = 0.5\,|i + 1|^{2H} + 0.5\,|i - 1|^{2H} - |\,i\,|^{2H}$. If i is large then $\rho(i) \sim H\,(2H\text{-}1)|\,i\,|^{2H-2}$. The parameter H, known as the *Hurst coefficient*, is the measure of the extent of dependence. If H has a value of 1/2 then it implies independence of observations while $0 < H < 1/2$ corresponds to negative dependence and $1/2 < H < 1$ corresponds to long-term dependence.

In economics, however, the usefulness of ARFIMA modeling was applied only in the analysis of monetary aggregates and asset returns which are known to be producing persistent series. A monetary aggregate is a mass of similarly classified economic indicators. The modeling technique was particularly useful to predict future movements in aggregate economic activity and to shape monetary policies. The presence of long-memory components in such series has important implications to the models and paradigms that had been used to understand trend and behavior. Long-run swings discovered in asset returns may prove wrong the existing speculations about diversity. Geweke and Porter-Hudak (1983) found out that fractionally-integrated models provided more reliable out-of-sample forecasts, although not the best, for the selected U.S. post-war economic time series such as CPI, Food CPI and WPI. Persistence in U.S. aggregate output (GNP and NNP) was examined by Diebold and Rudebusch (1989) through estimation of fractionally-integrated ARMA models.

There is a number of other ARFIMA modeling applications. Hui and Li (1988) illustrated the use of a fractionally-differenced periodic process in the analysis of Hongkong United Christian Hospital attendance series. A spatial model using fractional differencing in the time domain was formulated by Haslett and Raftery (1989) in their study of Ireland's wind power resource. The focus of the study was on the evaluation of the average wind power to be expected in the long term from a wind turbine at a given new site.

The recent modeling technique will be applied to the Philippine Stock Exchange oil index (daily figures). The time periods that will be considered are 1992-94, 1993-94 and 1994. The actual data in the first two weeks of January of 1995 shall be used for forecast comparisons. The study, however, will not venture into explaining volatile conditions and activities bringing forth accession or recession in the money market or investment world. It simply encompasses series trend- and behavior-description and model- formulation with test for optimal forecasting. Existing measures that will prove existence of persistence will not be used. Instead, the correlation plot will be examined. Moreover, the endeavor is to exhibit a methodology of the model-formulation of persistent time series. This methodology, however, does not totally depart from the usual Box-Jenkins iterative model-building approach. The only difference is that prior to the identification stage, the differencing parameter estimate is a fraction.

The organization of this paper is as follows. We present the working data in section 2. The rudiments of model identification are reviewed in section 3. Estimation is discussed in section 4 and model selection in section 5. The results and discussions are given in section 6. Some concluding remarks and recommendations are given in sections 7 and 8, respectively.

2. THE DATA

This study features the modeling of the Philippine Stock Exchange (PSE) oil index. This variable was selected after investigating about 18 blue-chip stocks, commercial and industrial, mining and composite indices, foreign exchange (dollar to peso) rates and T-bill rates. Results of initial processing of the data by using the prepared SAS program through its PROC MODEL had shown that the preliminary d-estimate of the oil index seems the lowest.

A stock index represents a central measure of the prices and serves as a reflection of the price movement of its component stocks. At present, there are five (5) stock indices identified in the PSE. These are CI or commercial-industrial index, mining index, property index (the most recent addition), oil index and PHISIX or composite index. Daily figures of each of the stock indices are listed in newspapers. The PSE has the following formula to compute for each stock index which is based on traded prices of stocks:

$$\text{TCLIX} = \text{YCLIX} \times \frac{\text{TMCAP}}{\text{YRMCAP}} \qquad (1)$$

where TCLIX is today's closing index, YCLIX is yesterday's closing index, TMCAP is today's market capitalization and YRMCAP is yesterday's revised market capitalization adjusted for any capital change in the constituent stocks.

As of January 20, 1995 the following stocks are listed under Oil:

Alcorn Petroleum - A	Alcorn Petroleum - B
Basic Petroleum - A	Basic Petroleum - B
Oriental Petroleum - A	Oriental Petroleum - B
Palawan Oil	Petrofields - A
Petrofields - B	Seafront Resources Corporation - A
Seafront Resources Corporation - B	South China
The Philodrill Corporation - A	The Philodrill Corporation - B
Trans-Asia	Vulcan Industrial

3. MODEL IDENTIFICATION

The time series plot is first examined to check for nonstationary tendencies. If the need arises, proper transformation of the original series is done in order to induce a constant variance. The produced variance-stabilized series, then, will be used to compute and plot the estimated autocorrelation coefficients and partial autocorrelation coefficients so as to determine the appropriate degree of differencing. Persistence or long-term dependence in a series is characterized by a hyperbolic decay-pattern in the autocorrelation plot. This pattern exhibits a slower dying-out to zero trend of the autocorrelation coefficients than the usual exponential decay-pattern.

Hosking (1981) pointed out assumptions about real values of differencing parameter (d). For $0 < d < 1/2$, the process is stationary with long-memory and may be expected to be useful for modelling long-term persistence. When $d = 0$, the process becomes a white noise with zero correlations. If $-1/2<d<0$, the process has a short-memory and is antipersistent characterized by negative autocorrelations and partial autocorrelations which decay monotonically and hyperbolically to zero.

The partial autocorrelation function is of little use in the identification stage when the series considered is persistent. This is so because of the complicated form of its plot pattern. However, Hosking (1981) showed through experimental evidence that for the estimated partial autocorrelation coefficient (ϕ_{kk}) and k number of lags considered,

$$\phi_{kk} \sim \frac{d}{k} \quad \text{as} \quad k \rightarrow \infty$$

Thus, $d \sim \phi_{kk} \bullet k$ as $k \rightarrow \infty$.

If evidence of persistence is detected, a new technique called fractional differencing will be applied. The evidence that is considered in this study is the hyperbolic decay-pattern in the estimated ACF. It should be noted, however, that there are existing measures or tests of long-range dependence. Some of these include the rescaled range test by Hurst (see Kottegoda, 1977), optimal tests other than rescaled-range test by Davies and Harte (1987) and the cumulative impulse-response function included in the discussion of Diebold and Rudebusch (1989).

The estimation procedure that will be applied in this paper is a modified version of the Hui-Li (1994) algorithm. The algorithm was a modified one since the estimation starts with an initial d-estimate and the formulation is for one period only. It should be noted that the original algorithm was formulated for two periods. The fractionally-differenced series becomes the basis in the computation of the new set of autocorrelation and partial autocorrelation coefficients which are considered in the determination of the closest fitting ARMA (Autoregressive Moving Average) model. The final step in the identification stage is the test for model inclusion of the deterministic trend element.

4. FRACTIONAL DIFFERENCING PARAMETER ESTIMATION

The stationary and variance stabilized series $\{X_t\}$ will be used to determine the fractional degree of differencing. Let $Z_t = (1 - B)^d X_t$ be the fractionally-differenced outcomes for $t = 1, 2, \ldots, n$, $0 < d < 1$ and with $(1 - B)^d$ as the fractional differencing operator. In the model building stage, computation of the preliminary fractional differencing parameter will be carried out. We modify Hui and Li's algorithm (see Hui and Li, 1994) to carry out the estimation of d. The modification is necessary since only one period is being considered in this study.

The fractional differencing operator $(1 - B)^d$ can be expressed as:

$$(1-B)^d = \nabla^d(B) = \sum_{k=0}^{\infty} \binom{d}{k} (-1)^k B^k. \quad (2)$$

Persistence is evident if $0 < d < 1/2$. Li and McLeod (1986) pointed out that the fractional differencing operator, $\nabla^d(B)$, can be approximated through the binomial expansion:

$$\nabla_r^d(B) = \sum_{k=0}^{r} \binom{d}{k} (-1)^k B^k. \quad (3)$$

where, according to Hui and Li (1994), the value of r is such that $r \to \infty$ as $n \to \infty$ but r/n is o(1). With the fixed value of r, the ARIMA (0, d, 0) can be estimated.

The following equations featuring the forward-shift operator (F) and the backward-shift operator (B) where $F = B^{-1}$ are used in the algorithm:

$$\nabla_r^d(B)\ X_t = a_t \qquad t = 1, 2, ..., n \quad (4)$$

where $\{a_t\}$ are identical, independent and normally-distributed errors with mean 0 and variance σ^2, $0 < d < 1$ and $B(X_t) = X_{t-1}$;

$$\nabla_r^d(F)\ X_t = c_t \qquad t = 1, 2, ..., n \quad (5)$$

where $\{c_t\}$ are identical, independent and normally-distributed errors with mean 0 and variance σ^2, $0 < d < 1$ and $F(X_t) = X_{t+1}$.

The algorithm has the following steps:

1. Get a preliminary estimate d by minimizing

   $\sum_{t=1}^{n} a_t^2$ with respect to d.
2. Using the d in (1), calculate $\{c_t\}$ for $t = n\text{-}r, n\text{-}r\text{-}1, ..., 1$ using (5).
3. Set $\{c_t\} = 0$ for $t = 0, -1, -2, ..., 1-r$.
4. Backcast $\{X_t\}$ for $t = 0, -1, -2, ..., 1-r$.
5. Calculate $\{a_t\}$ for $t = 1, 2, ..., n$ using (4).
6. Minimize $\sum_{t=1}^{n} a_t^2$ with respect to d.

An interactive macro program was prepared to determine the fractional differencing parameter estimate (d). Basically, it has three parts. The first part computes for an initial d-value. This is followed by backcasting. The number of backcasts produced will depend on the r-value determined using PROC IML of SAS. In the program, the value of r is known by setting a minimum for the (dCk) value like 0.001. The last part is d-final estimation. Here, the backcasts that were computed are included. The estimate that will be computed through this program is the d-value that will be used to difference the variance-stabilized series.

Procedure Model or PROC MODEL is the main part of the prepared SAS program. This procedure included in SAS/ETS is capable of solving nonlinear systems of equations. For the designed program, the solution mode used is NEWTON which computes a simultaneous solution and the iterative minimization method is GAUSS. Newton's method computes analytic derivatives of the equation errors with respect to the solution variables. Analytic derivatives are efficiently calculated using exact formulas for computing derivatives instead of numerical approximations (SAS/ETS Guide Version 6). For more details see Binsol (1995).

5. ARMA Model Selection

The estimated ACF and PACF of the fractionally-differenced series are compared with theoretical ACF's and PACF's in order to determine the initial model to be fitted. The PROC ARIMA in SAS/ETS is used to produce the correlation plots. Three major types of processes are considered. These are the autoregressive (AR), moving-average (MA) and autoregressive-moving average (ARMA) processes. In practice, the tentative model chosen should be parsimonious. That is, it must contain the smallest number of parameters to be estimated. This case is true for short-term dependent or anti-persistent series. There may be a difference when long-term dependent series is considered. Geweke and Porter-Hudak (1983) pointed out that a non-parsimonious model may better represent long-memory persistence.

The non-seasonal ARFIMA (p, d, q) model has the following form:

$$\phi_p(B) \ \nabla^d(B) \ X_t = \theta_q(B) \ a_t \tag{6}$$

where $\{a_t\}$ are identical, independent and normally-distributed errors with mean 0 and variance σ^2, $\phi_p(B)$ is the autoregressive polynomial operator and $\theta_q(B)$ is the moving-average polynomial operator.

After fractionally-differencing the variance-stabilized series, the mean of the new series is either zero or nonzero. A model set with the differenced series having a nonzero mean has a constant term which is called the deterministic trend element. This value when included in the model brings about a certain shift in the forecast. The nonseasonal ARFIMA (p,d,q) model with the deterministic trend element (θ_0) has this form:

$$\phi_p(B) \ \nabla^d(B) \ x_t = \theta_0 + \theta_q(B) \ a_t \tag{7}$$

A preliminary statistical test using t-test is performed to find out if θ_0 is zero based on the mean of $\{Z_t\}$ and its approximate standard error. Box and Jenkins exhibited the approximate standard error for the mean of $\{Z_t\}$ for some ARMA processes. If θ_0 is significant then it must be included in the model.

The parameter estimates are then checked for possible inclusion in the model. Specifically, the coefficients will be examined if they meet the stationarity or invertibility conditions, high-quality criteria, being uncorrelated from each other and being off near-redundancy. Finally, the adequacy of the model set is tested in terms of forecast errors using the root mean square error (RMSE), the mean absolute percent error (MAPE) and the mean square prediction error (MSPE).

After the initial model selection and parameter estimation, the next and final stage in the model formulation is the check whether the residuals follow a white noise behavior. The residuals are expected to be uncorrelated with zero mean and a constant variance. If these conditions fail then model-reformulation is necessary which means going back to the identification stage. A Pormanteau test for lack of fit as featured in SAS and explained by Wei is used to determine model adequacy. This statistic approximately follows a chi-square distribution. It will be used to test the null hypothesis that all the errors are significantly different from zero.

When a set of apt models is at hand, other model selection criteria based on the residuals are used to select the best model. These are the Akaike-Information Criterion (AIC) and the

Schwartz-Bayesian Criterion (SBC). The model with the minimum value for these criteria is considered as the best fitting model for the series.

6. RESULTS AND DISCUSSION

6.1 Fractional Differencing Parameter Estimation

In order to determine if there is a change in the persistence pattern of the Philippine Stock Exchange oil index, three periods were considered. Daily data were gathered from leading newspapers (Manila Standard, The Philippine Daily Inquirer and The Philippine Star) for years 1992-94. The analysis periods used were 1994, 1993-94 and 1992-94.

Figure 2
Oil Index Series 1992-1994

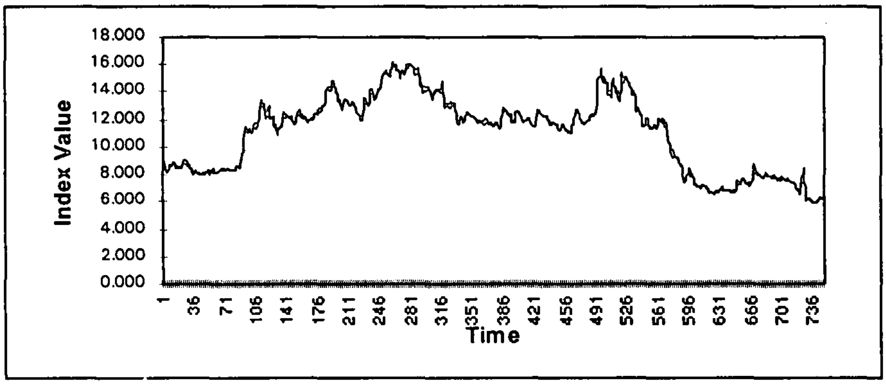
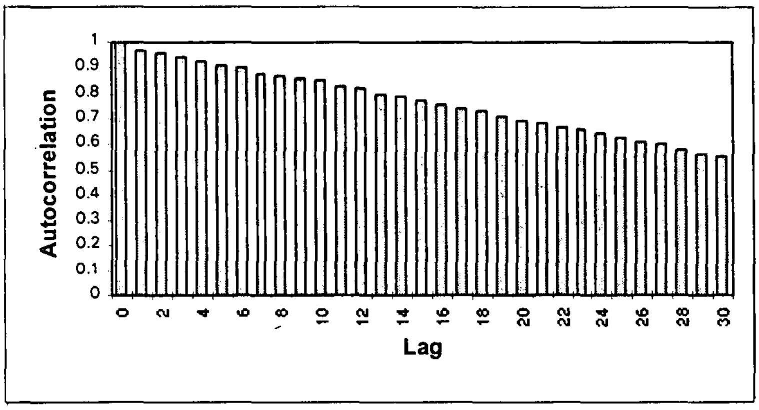


Figure 3a
ACF of Oil Index (Log) 1994 Series



As can be observed from Figure 2, the trend of the daily figures of oil index 1994, oil index 1993-94 and oil index 1992-94 generally shows a downward movement. Notice that fluctuations, variations or periodicities that are contained in the data are not regular. Furthermore, the plots indicate that the series mean changes over time and that the variance is unstable. The natural log transformation was used to stabilize the variance and nonseasonal differencing is required to have a constant mean. The sample ACF's (autocorrelation functions) exhibit a slow decay of positive values which is more reminiscent of a fractional or hyperbolic decay (refer to Figures 3(a), 3(b) and 3(c)).

Figure 3b
ACF of Oil Index (Log) 1993-94 Series

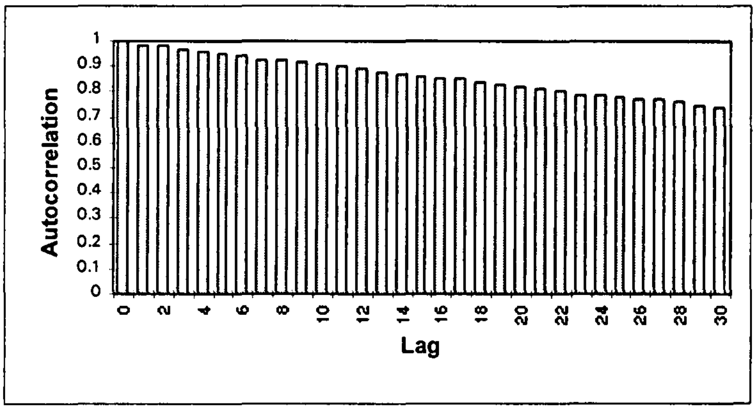
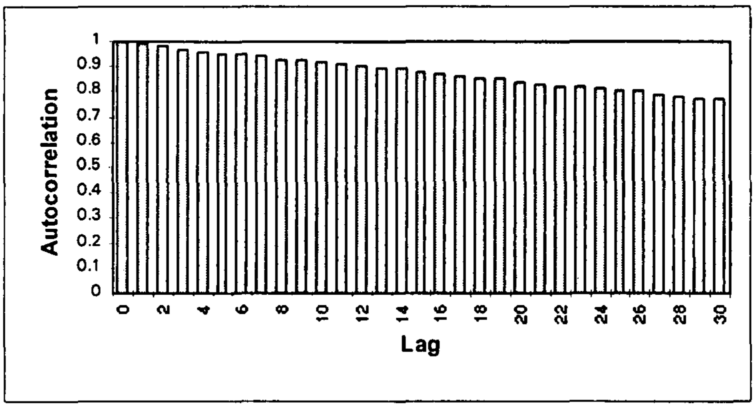


Figure 3c
ACF of Oil Index (Log) 1992-94 Series



Crucial to the estimation of the fractional differencing parameter (d) is the determination of an appropriate r-value in the binomial expansion

$$\nabla_r^d(B) = \sum_{k=0}^{r} \binom{d}{k} (-1)^k B^k.$$

The weights (dCk), k =0,...,r are almost 0 for large k. The output in Table 1 shows that for specific values of d (0.01 and 0.99 and 0.1 to 0.9 with increment of 0.1) and setting |(dCk)| < 0.001, r is highest at 73 when d = 0.2. It could have been larger if the d-increment was made smaller. The result, r = 73, implies that 73 observations are needed just to estimate d. Denoting by n_O the remaining number of observations (n - r), for oil index 1994, n_O = 179, for oil index 1993-94, n_O = 428 and for oil index 1992-94, n_O = 673. Further analysis of results in Table 1 suggests that the highest value of r, at any d-increment, occurs between d = 0.1 and d = 0.3.

Although the optimization procedure is not the best, it serves as a beginning of an exploration. Through the application of the interactive macro program the values of d are determined. The estimation procedure is maximum likelihood estimation as patterned after Hui-Li using the default set in PROC MODEL of SAS like Newton's method as the solution mode, the Gaussian iterative procedure and convergence criterion of 0.001. After processing, d = 0.86 was obtained for oil index 1994, d = 0.92 for oil index 1993-94 and d = 0.95 for oil index 1992-94. The three d-values show a decreasing trend as the number of observations decreases. This could be an indication that the degree of persistence decreases as the number of years decreases. Furthermore, the values are greater than 0.5 which, as discussed in Sowell (1992a), suggest less persistence.

Table 1

Values of dCk for Various Combinations of d and r

					d - Values					
r	0.01	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	0.99
1	0.01000	0.10000	0.20000	0.30000	0.40000	0.50000	0.60000	0.70000	0.80000	0.90000	0.99000
2	-0.00495	-0.04500	-0.08000	-0.01500	-0.12000	-0.12500	-0.12000	-0.10500	-0.08000	-0.04500	-0.00495
3	0.00328	0.02850	0.04800	0.05950	0.06400	0.06250	0.05600	0.04550	0.03200	0.01650	0.00167
4	-0.00245	-0.02066	-0.03360	-0.01016	-0.04160	-0.03906	-0.03360	-0.02616	-0.01760	-0.00866	-0.00084
5	0.00196	0.01612	0.02554	0.02995	0.02995	0.02734	0.02285	0.01727	0.01126	0.00537	
6	-0.00163	-0.01316	-0.02043	-0.02328	-0.02296	-0.02051	-0.01676	-0.01237	-0.00788	-0.00367	
7	0.00139	0.00109	0.01693	0.01896	0.01837	0.01611	0.01293	0.00937	0.00586	0.00267	
8	-0.00122	-0.00957	-0.01439	-0.01588	-0.01516	-0.01309	-0.01034	-0.00738	-0.00454	-0.00204	
9	0.00108	0.00840	0.01247	0.01358	0.01280	0.01091	0.00850	0.00598	0.00363	0.00161	
10	-0.00097	-0.00748	-0.01097	-0.01182	-0.01101	-0.00927	-0.00714	-0.00497	-0.00298	-0.00130	
20		0.00348	-0.00475	0.00475	0.00411	0.00322	-0.00230	0.00148			
30		-0.00222	-0.00291	-0.00280	-0.00232	-0.00174	-0.00119				
40		-0.00162	-0.00206	-0.00192	-0.00155	-0.00113					
50		-0.00127	-0.00158	-0.00144	-0.00113						
60		-0.00104	-0.00127	-0.00113							
65		0.001149	0.00101941								
70		-0.001051									
73		0.0009994									

6.2 ARMA Model for the Fractionally-Differenced Series

In order to determine the ARMA model to be fitted in each of the three series, reference to theoretical ACFs, IACFs and PACFs was considered. However, the patterns displayed in the sample ACFs, IACFs and PACFs are atypical. Hence, AR, MA and ARMA models of order 1 were all tried to determine the best representation. Although it was discovered that the ARFIMA(0,d,0) model or the fractional white noise model of each of the three series has statistically independent residuals, the consideration of ARMA fitting proved to be important due to the findings in Table 2.

For all the three series, the AR(1) model appears to be appropriate because of higher absolute t-values and fairly good coefficients. It should be noted that all the coefficients are of high quality since their absolute values are less than 1 and the corresponding t-values are significant at most at 0.10 level. In Table 3 the tentative ARFIMA(p,d,q) models selected are presented. It is noticeable that the AR coefficient in the third series is small. One might suggest the appropriateness of the fractional white noise model. For this case, what accounts for is the fact that the residuals are white noise at ARFIMA(0,d,0) fit. However, this small AR coefficient is significant as displayed in Table 5.

Table 2
Parameter Estimates for ARFIMA and ARIMA Models
Oil Index 1994, 1993-1994, 1992-94 (Log Series)

Model	Oil Log 1994 Series d = 0.86 AR1	MA1	Oil Log 1993-94 Series d = 0.92 AR1	MA1	Oil Log 1992-94 Series d = 0.95 AR1	MA1
ARFIMA(1,d,0)	-0.13481 (-1.81)		-0.1058 (-2.09)		-0.08015 (-2.08)	
ARFIMA(0,d,1)		0.11824 (1.59)		0.09037 (1.87)		0.07448 (1.94)
ARFIMA(1,d,1)	-0.39109 (-0.94)	-0.52217 (-1.35)	-0.38002 (-0.96)	-0.27921 (-0.68)	-0.33695 (-0.86)	-0.4162 (-1.11)
ARIMA(1,1,0)	-0.24673 (-3.39)		-0.17339 (-3.64)		-0.13098 (-3.42)	
ARIMA(0,1,1)		0.24347 (3.35)		0.16768 (3.51)		0.12901 (3.37)
ARIMA(1,1,1)	-0.3092 (-1.07)	-0.06676 (-0.22)	-0.22427 (-0.83)	-0.05246 (-0.19)	-0.05788 (-0.20)	-0.18869 (-0.66)

* Values in parenthesis correspond to t-values.

Measures such as AIC (Akaike Information Criterion) and SBC (Schwartz-Bayesian Criterion) were used to determine the model that best represents the fractionally-differenced series. The AIC and SBC results for the oil index series are in Table 4. The findings reveal that the selected model for each series is not the best if the criteria are used though these criteria do not agree. AIC selects the corresponding AR(1) representation of the fractionally-differenced series and SBC selects the corresponding fractional white noise model. Both models have no constant trend. This is a crucial situation because both models show promise.

Table 3
ARFIMA and ARIMA Models for Oil Index Series

Series	ARFIMA(p,d,q)	ARIMA(p,d,q)
Oil Index 1994	$(1+0.13841B)^{0.86}X_t=a_t$	$(1+0.24673B)(1-B)X_t=a_t$
Oil Index 1993-94	$(1+0.10058B)^{0.92}X_t=a_t$	$(1+0.17339B)(1-B)X_t=a_t$
Oil Index 1992-94	$(1+0.08015B)^{0.86}X_t=a_t$	$(1+0.13098B)(1-B)X_t=a_t$

In order to determine the better model, each must be tested in terms of forecasting ability. However, if consideration is modeling the short-run behavior-part of the persistent series then the fractional white noise model may not very well fit this possibility. With the available measures, the better model for the oil index series could be found. Such measures include the adjusted root mean-squared error (RMSE) and mean absolute percent error (MAPE).

From the results in Table 5 it is noticeable that there is no clear boundary between the two models. MAPE values are lower for the fractional white noise model while RMSE's are lower for the AR(1) model. Since both models seem appropriate if based on the measures, a choice of one of them maybe suggestive. The choice is still the AR(1) representation. The selection is but proper for succeeding comparisons with their corresponding ARIMA(p,1,q) counterpart which at an initial stage of the investigation is AR(1) also. This will permit a clearer comparison and easier interpretation of ARFIMA and ARIMA models since for both sets, the AR parameter is the only one estimated after considering the differencing degree.

For further diagnostic checks, the residual autocorrelations were examined. Based on the Ljung-BOX statistic computed using PROC ARIMA of SAS, the residuals are white noise and therefore, the oil index series models are apt.

6.3 Comparison of ARIMA and ARFIMA Models

For comparison, in both ARFIMA and ARIMA modeling n_O (number of actual figures less r) observations for each oil index series are considered. It is noteworthy to mention that the fractionally-differenced series has a more stationary mean and variance than the corresponding first-differenced series.

For the fractionally-differenced series, AR(1) model was selected for the three oil index series. In order to test the ability of these fractionally-differenced models as compared with conventional models, it is but wise to consider applications of ARIMA modeling.

The iterative approach starts with model identification then followed by estimation of ARMA model parameters, evaluation of parameter estimates, diagnostic-checking and lastly, forecasting. The nonstationary tendencies which are evident in the natural log series plot suggest nonseasonal differencing. First-differencing was applied on each of the oil index series. For the ACF, the same pattern of having a spike in lag 1 then cutting off to zero was found. Again, AR, MA and ARMA models of order 1 were all fitted on the first-differenced series. Table 3 displays each model coefficients with the corresponding t-value. Based on these, it turns out that the AR(1) model is the best representation for the three oil index series.

Table 4

AIC, SBC of ARFIMA and ARIMA Models for Oil Index Series

Model	Oil Log 1994 Series d = 0.86		Oil Log 1993-94 Series d = 0.92		Oil Log 1992-94 Series d = 0.95	
	AIC	SBC	AIC	SBC	AIC	SBC
ARFIMA(0,d,0)	-613.084	-613.084	-1677.278	-1677.278	-2731.274	-2731.274
ARFIMA(1,d,0)	-614.358	-611.170	-1679.624	-1675.565	-2733.608	-2729.096
ARFIMA(0,d,1)	-613.934	-610.746	-1679.180	-1675.121	-2733.297	-2728.786
ARFIMA(1,d,1)	-613.254	-606.879	-1678.326	-1670.208	-2732.364	-2723.340
ARIMA(0,1,0)	-602.657	-602.657	-1668.115	-1668.115	-2726.778	-2726.778
ARIMA(1,1,0)	-611.884	-608.696	-1679.168	-1675.109	-2736.415	-2731.903
ARIMA(0,1,1)	-611.456	-608.268	-1678.650	-1674.591	-2736.186	-2731.674
ARIMA(1,1,1)	-609.913	-603.538	-1677.204	-1669.444	-2734.444	-2725.420

The selection of AR(1) representation for the first-differenced oil index series proved to be promising because AIC and SBC are lowest for these models. The findings are in Table 4. These are supported by a white noise pattern in the corresponding set of residual autocorrelations. It should be noted that after AR(1) fitting of the first-differenced observations, the residuals became white noise.

After choosing the best set of ARFIMA and ARIMA model representations for the oil index series, it is proper to decide the better set between them. The first check for model adequacy is through AIC and SBC measures. The criteria are computed for the two sets and are displayed in Table 4. Both AIC and SBC are lower for the ARFIMA(1,d,0) model representations of oil index 1994 and 1993-94 series. While for oil index 1992-94 series, the better choice is ARIMA(1,1,0).

Table 5
RMSE, MAPE of ARFIMA(0,d,0), ARFIMA(1,d,0) and ARIMA(1,1,0) Models

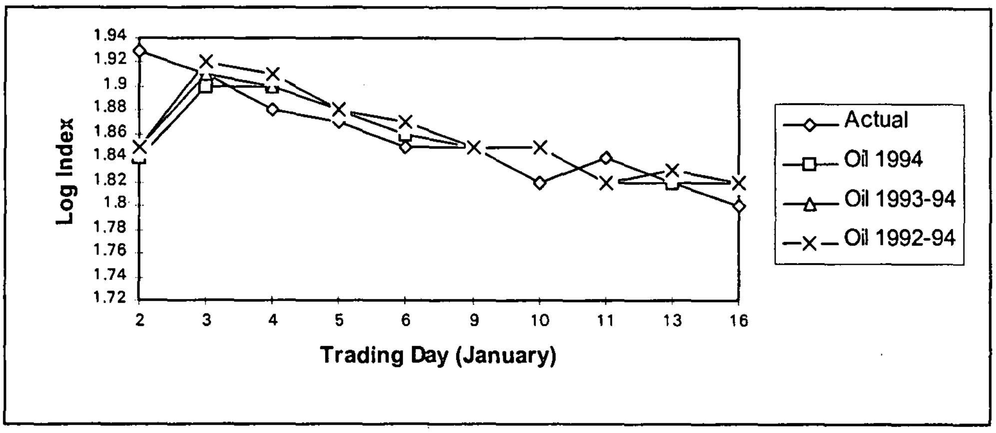
Model	Oil Log 1994 Series d = 0.86		Oil Log 1993-94 Series d = 0.92		Oil Log 1992-94 Series d = 0.95	
	RMSE	MAPE	RMSE	MAPE	RMSE	MAPE
ARFIMA(0,d,0)	0.001890	1.175468	0.001151	0.872033	0.001008	0.826097
ARFIMA(1,d,0)	0.001916	1.165018	0.001165	0.869708	0.001013	0.822033
ARIMA(1,1,0)	0.001918	1.203441	0.001152	0.884250	0.01003	0.830849

The next criteria for the selection of the better model is through forecasting ability. The choice is made through comparison of forecast errors. This is accomplished through computation of the respective RMSE, and MAPE for each model. In Table 5, comparative results of the measures are given. Based on the RMSE and MAPE, the ARFIMA(1,d,0) model representations of the oil index series outperform the ARIMA(1,1,0) model representations. The findings are very well-supported through comparison of the actual and fitted values.

The final measure to be considered in determining the model with a better quality of fit is the mean-squared prediction error (MSPE). Its computation depends on the out-of-sample forecasts compared with the additional ten actual data from January 2-16, 1995 trading days (see Fig. 4). Prior to its computation is the presentation of the plots featuring the actual values and the one-step-ahead forecasts. For oil index 1994 (log) series, ARFIMA(1,0.86,0) model has six out of ten better forecasts. For oil index 1993-94 (log) series, ARFIMA(1,0.92,0) model outperformed ARIMA(1,1,0) model with seven out of ten better forecasts. While for oil index 1992-94 (log) series, ARFIMA(1,0.95,0) model forecasted better seven out of ten indices.

While it is discovered that ARFIMA models perform better than ARIMA models in modelling persistence of the three oil index series, there is one best ARFIMA representation. Figure 4 provides the information. Considering the three comparative time-periods and if time factor is included in the analysis, the model fit for the one-year oil index data, that is, ARFIMA(1,0.86,0) provides the best set of forecasts. This suggests that the more recent available data of oil index will explain better its present or future behavior. At this stage, it remains to be discovered as a generalization if longer time periods are needed. As another point of discussion, the selected series may not have represented adequately a persistent series based on the estimates that were computed but the use of fractional differencing proved to be more advantageous than integral differencing based on the findings.

Figure 4
Actual Values and One-Step Ahead Forecasts for the ARFIMA Models



7. CONCLUDING REMARKS

This study serves as an evidence of the usefulness of autoregressive, fractionally-integrated and moving-average (ARFIMA) models in capturing the necessary properties of the Philippine Stock Exchange oil index series. The usefulness holds true for the 1994, 1993-94 and 1992-94 daily figures that were covered. More illustrative findings and generalizations regarding the behavior pattern of the oil index could have been made possible if longer periods of time were considered. Nevertheless, this undertaking had attained its purpose of showing that fractional differencing is a better technique to apply than integral differencing in removing long-term persistence. The use of a statistic like R/S analysis in order to determine the existence of persistence phenomenon was not considered. Instead, visual means were used. This is through the inspection of the autocorrelation plots. The hyperbolic pattern in the ACFs of the oil index series, just like in most empirical studies, is an indication of its existence. Due to this, fractional differencing is the more appropriate technique to achieve stationarity.

As regards the estimation procedure which was designed in the context of maximum likelihood estimation, the values that were determined are promising. The differencing parameter estimates obtained were 0.86, 0.92 and 0.95 for oil index 1994, 1993-94 and 1992-94 series, respectively. As a realization, these values do not agree with the assumed values of 0 to 0.5 of proponents like Beran, Diebold, Porter-Hudak and Sowell for long-term persistence to exist. In spite of that, the conjecture of persistence existence in the oil index series remains. The optimization procedure of estimating the nonseasonal fractional differencing parameter is only one of the many possible means that could be explored. It may not be considered as the best, but for the time being, it is suggestive. Certainly, the methodology which was patterned after Hui and Li (1994) has provided a system which is not very complicated.

A systematic iterative procedure of modeling a persistent time series had been formulated. This was strictly followed in modeling the three oil index series. As mentioned in the data methodology and described in Hosking(1981), it is a generalized version of ARIMA iterative modeling. The first step in the procedure is the estimation of the differencing parameter of the variance-stabilized series. This is followed by the selection of the tentative autoregressive and moving-average (ARMA) representation and evaluation of the parameter estimates. The next step is diagnostic checking to determine the adequacy of the fitted model. If there is lack of fit then another ARMA model is selected. Finally, the selected model is tested for efficient forecasting.

Following the above iterative procedure, it was found out that ARFIMA1,d,0) with no deterministic trend best fitted the behavior pattern of the oil index series. Specifically, oil index 1994 (log) series admitted ARFIMA(1,0.86,0) representation. Oil index 1993-94 (log) series admitted ARFIMA(1,0.92,0) and oil index 1992-94 (log) series was represented by ARFIMA(1,0.95,0). These results were arrived at after comparisons of one-step-ahead forecasts with actual values, model selection criteria such as AIC and SBC and realization of the presence of reasonably clear white noise pattern in the residual ACF's.

8. RECOMMENDATIONS

Long memory models with two classifications - fractional Gaussian noise and fractionally-differenced models - are areas in time series analysis where research is in its infancy. These classes of models are needed to refine the conventional modeling approaches so as to increase precision in terms of modeling or capturing the trend or behavior in a series. This study serves only as an introduction of the usefulness of fractionally-differenced models.

The first possible area of consideration is the measure of persistence. Persistence or long-run behavior is a qualifying characteristic before possible application of a long-memory model. Qualified as robust techniques - R/S analysis and cumulative response function - were not yet generalized as applicable to series that have the tendency to exhibit short-run behavior.

The second possible exploration is on fractional differencing estimation. Areas that might be covered are estimations considering conditional least squares and unconditional least squares and still, maximum likelihood estimation. The study just presented a version of maximum likelihood estimation which could be improved.

The Hui-Li algorithm had been suggestive in terms of determining a d-estimate but it could be further improved if the system of fixing r depending on the sample size is incorporated. A so-designed procedure when formulated will increase the level of its precision in terms of estimating the fractional differencing parameter. As a realization based on the trials conducted for this study,. the identification of an appropriate r-value may be done by trial and error assignment.

In addition to fractional differencing parameter estimation, a unified program consisting of ARMA model parameter estimations, diagnostic-checks and forecasting as in PROC ARIMA must be constructed as if the SAS procedure is extended. In this study, a program was designed for each step.

Most of the studies done in this area made use of the spectrum density approach. That is why the available estimation procedures are in line with this. The recommendation is to further strengthen studies on time series domain. Furthermore, seasonal variations are common to monthly, quarterly and annual data. The application of a fractionally-differenced model that is fitting to this kind of data may be explored.

REFERENCES

BERAN, J. (1989), A Test of Location for Data with Slowly Decaying Serial Correlations, Biometrika 76, 261-269.

BINSOL, A. (1995), On the Application of Autoregressive Fractionally-Integrated, Moving Average (ARFIMA) Model to Philippine Stock Exchange Oil Index, Masteral Report, UP Statistical Center.

BOX, G. E. P. and JENKINS, G. M. (1970), Time Series Analysis: Forecasting and Control, San Francisco: Holden-Day.

DAVIES, R. B. and HARTE, D. S. (1987), Tests for Hurst Effect, Biometrika 74, 95-102.

DIEBOLD, F.X. and RUDEBUSCH, G. D. (1989), Long Memory and Persistence in Aggregate Output, Journal of Monetary Economics 24, 189-209.

GEWEKE, J. and PORTER-HUDAK, S. (1983), The Estimation and Application of Long Memory Time Series Models, Journal of Time Series Analysis 4, 221-237.

HASLETT, J. and RAFTERY, A. E. (1989), Space-time Modeling with Long-Memory Dependence: Assessing Ireland's Wind Power Resource (With Discussion), Journal of Royal Statistical Society Series (C) 38, 1-21.

HOSKING, J. R. M. (1981), Fractional Differencing, Biometrika 68, 165-176.

HUI, Y.V. and LI, W.K. (1994), On Fractionally Differenced Periodic Processes, Unpublished Manuscript, City Polytechnic of Hongkong and University of Hongkong.

LAWRENCE, A. J. and KOTTEGODA, N. T. (1977), Stochastic Modeling of Riverflow Time Series (With Discussion), Journal of Royal Statistical Society Series (A) 140, 1-47.

LI, W.K. and MCLEOD, A. I. (1986), Fractional Time Series Modeling, Biometrika 73, 217-221.

PANKRATZ, A. (1983), Forecasting with Univariate Box-Jenkins Models (Concepts and Cases), New York: John Wiley and Sons.

PORTER-HUDAK, S. (1990), An Application of Seasonal Fractionally-Differenced Model to the Monetary Aggregates, Journal of American Statistical Association 85, 338-344.

SOWELL, F. (1992a), Modeling Long-run Behavior with the Fractional ARIMA Model, Journal of Monetary Economics 29, 277-302.

SOWELL, F. (1992b), Maximum Likelihood Estimation of Stationary Univariate Fractionally Integrated Time Series Models, Journal of Econometrics 53, 165-188.

WEI, W. (1990), Time Series Analysis (Univariate and Multivariate Methods), California: Addison-Wesley.

SAS/ETS Guide (Version 6.0).

SAS/IML Guide (Version 6.0).

SAS Macro Processing Guide (Version 6.0).