World edible oil prices prediction: Evidence from mix effect of ever difference on Box-Jenkins approach

Abdul Aziz Karia
Taufik Abd Hakim
Imbarine Bujang
Faculty of Business Management
Universiti Teknologi MARA (UiTM) Sabah, Malaysia

Keywords
First order difference; fractionally difference; overdifference; stationary; forecast

Abstract
The Box-Jenkins model assumed that the time series is stationary. Generally, researchers will conduct the first order difference as a necessary procedure of stationarity data. The first or second order difference seems to be a good solution towards nonstationarity counterparts, but this effort might lead into the possible over difference. Thus, alternative procedure of fractionally difference can be considered as a solution towards the over difference, since it permits the non-integer value of d *. However, the fractionally difference has been proved by several researchers to produced poor out-sample forecast as compared to its rival models. Therefore, we investigate the over difference's effect on five of the selected world edible oil prices that observed to have long memory behavior. Besides, we compare the performance of two difference models which are the autoregressive integrated moving average (ARIMA) and autoregressive fractionally integrated moving average (ARFIMA) models using the time series data that observed with the over difference and long memory behavior. The forecasting show mixed results and the addressed over difference seems not to give a significant effect neither ARIMA nor ARFIMA models. We also found that the ARFIMA model does not demonstrate poor out-sample forecasting.*

1. Introduction

Detecting the existence of long memory in time series data has been an issue for the attention of econometricians, statisticians and researchers. Likewise, if there exists of long memory, the will be the tendency of over difference. Suppose that $Y_t - \mu = \phi\left(Y_{t-1} - \mu\right) + e_t$, the value of $|\phi| < 1$ indicates the time series data $\left(Y_t\right)$ is stationary. At this point, researchers will implement the simple autoregressive moving average model (ARMA) of $\left(1-\phi_1 L - \phi_2 L^2 + \ldots + \phi_p L^p\right) Y_t = \mu + \left(1 - \theta_1 L - \theta_2 L^2 - \ldots - \theta_q L^q\right)\varepsilon_t$. If $\phi = 1$, the time series of Y_t is considered nonstationary. The issue arises when Box-Jenkins model assumed that the Y_t must be stationary. In order to meet this assumption, the necessary procedure of differencing $\left(\Delta Y_t = Y_t - Y_{t-1}\right)$ is performed, generally the first order difference towards achieving the stationary of Y_t .In this case, researchers will adopt the autoregressive integrated moving average model (ARIMA).

The necessary procedure of differencing seems to be very good solutions toward the nonstationary counterpart. However, it might lead to a tendency to over difference (Erfani & Samimi, 2009). Based from the study conducted byKaria, Bujang, and Ahmad (2013), the time series suffers from over difference if the reported result from the analysis of unit root test indicate large value statistics of Augmented Dickey Fuller (ADF), Phillips Perrons (PP) and Dickey Fuller using Generalized Least Squares (DF GLS), while small value statistic for

Kwiatkowski, Phillips Schmidt and Shin (KPSS). Besides, the previous studies also indicate the tendency of over difference toward Y_t if the plots of autocorrelation function (ACF) decays at a very hyperbolic rate or sluggish, (Arouri, Hammoudeh, Lahiani, & Nguyen, 2012; Diebold &Inoue, 2001; Karia et al., 2013; Kwan, Li, & Li, 2012; Perron & Qu, 2007; Tan, Galagedera, & Maharaj, 2012; Xiu & Jin, 2007).Hurvich and Ray (1995) found that the time series that suffer from over difference could be biased for long memory prediction since it ineffective in parameters estimation.

Meanwhile, an alternative necessary procedure of fractional difference is one of the most popular approaches in dealing with long memory time series analysis. The ARFIMA is outperformed compared to its rival models in predicting varieties of time series area (Baillie & Chung, 2002; Reisen & Lopes, 1999). Moreover the ARFIMA is good in predicting out-sample time series prediction (Bhardwaj & Swanson, 2006; Chortareas, Jiang, & Nankervis, 2011; Chu, 2008; Koopman, Jungbacker, & Hol, 2005). On the other hand, Xiu and Jin (2007) and Ellis and Wilson (2004) found that the ARFIMA produced poor out-sample forecast for their time series data.

Considering the analysis from the literature, we have several concerns about the subsequent vital issues, which are: (1) whether the possible over difference could degrades the performance of ARIMA model? (2) Is the necessary procedure of fractionally integrated produces poor in-sample and out-samples forecasting? In order to examine the existence of the possible over difference, this study decided to select five world edible oils which are crude palm oil (CPO), soybean, rapeseed, sunflower and linseed.

2. Data and Methodology

In regard to the addressed issues such the existence of the over difference in relation to the performance of ARIMA model and the poor performance of ARFIMA model, this study utilized the world edible oil pricesthat we observe shows the long memory behaviour. This study obtained CPO, soybean, rapeseed, sunflower and linseed prices from Datasteam. The data was in daily basis from first of January 2008 to end of December 2013 at Free-on-Board (FOB) Malaysian Ringgit (RM) to US dollar ($) per tonne. Every five of these data consisting of1566 observations.

2.1 Autoregressive Integrated Moving Average Model

The autoregressive moving average model is the combination between the autoregressive (AR) and moving average (MA) model. If the $Y_t - \mu = \phi\left(Y_{t-1} - \mu\right) + e_t$ shows the value of $|\phi| < 1$, it gives an impression that the intended time series data is stationary. The model will be implemented if the intended time series is said to be stationary around the mean. The basic ARMA (p,q) model can be derived as:

$$\Phi(L)\left(Y_t - \mu\right) = \Theta(L)\varepsilon_t \quad (1)$$

However if $Y_t - \mu = \phi\left(Y_{t-1} - \mu\right) + e_t$ and $\phi = 1$,the time series of Y_t is considered nonstationary. In order to meet the stationarity assumption of the Box and Jenkins (1976) model, the necessary procedure of differencing $\left(Y_t^* = Y_t - Y_{t-1}\right)$ need to be done, generally integer value of $d = 1$ and $d = 2$. With the implementation of the necessary procedure of differencing, the time series will be at detrended value of Y_t^*. It is also known as autoregressive integrated moving average model, ARIMA(p,d,q).

2.2 Autoregressive Fractionally Integrated Moving Average Model

The ARFIMA model can be considered as a very useful for time series data that has a strong persistency level towards nonstationary (Mostafaei & Sakhabakhsh, 2011). It is important to give a special attention toward the ARFIMA package introduced by Doornik and Ooms (2004) which has the capability to adopt the Maximum Likelihood Estimation (MLE) to the long memory time series data. Various literature has noted the main weakness of adopting the MLE towards the ARFIMA estimation procedure and the problem has essentially been solved by Hosking (1981) and Sowell (1987). However, Ooms and Doornik (1999) list the reasons why some problems remained unsolved. There will be problems in variance matrix into account which is totally inappropriate for extensions with regression parameters. Therefore, the study conducted by Doornik and Ooms (2004)was proposed in order to tackle the problems in variance matrix that can be expose as follow:

Assuming either $\varepsilon_t \sim NID\left[0,\sigma_\varepsilon^2\right]$, or $E\left[\varepsilon_t\right]=0$ and $E\left[\varepsilon_t^2\right]=\sigma_\varepsilon^2$

Therefore, the basic ARMA (p,q) model can be derived as:

$$\Phi(L)(Y_t-\mu)=\Theta(L)\varepsilon_t \quad (2)$$

Whereby, L and ε_t are the lag operator and a white noise of a series respectively. For the nonstationary solution, the fractionally difference d or the ARFIMA (p,d,q) can be derived as:

$$\Phi(L)(1-L)^d(Y_t-\mu)=\Theta(L)\varepsilon_t \quad (3)$$

Where p and q are integers while the d is real. The main player in the ARFIMA model is $(1-L)^d$ which is the fractionally difference operator and defined as the binomial equation as follows:

$$(1-L)^d=\sum_{j=0}^{\infty}\delta_j L^j=\sum_{j=0}^{\infty}\binom{d}{j}(-L)^j \quad (4)$$

With this, the stationary auto covariance function with μ is written as follow:

$$\gamma_i=E\left[(Y_t-\mu)(Y_{t-i}-\mu)\right] \quad (5)$$

Therefore, Doornik and Ooms (2004) provide the solutions towards the variance matrix of the joint distribution of $Y=(Y_1,\ldots,Y_t)'$ which is presented as follows:

$$V[y]=\begin{pmatrix}\gamma_0 & \gamma_1 & \cdots & \gamma_{T-1}\\ \gamma_1 & \gamma_0 & \ddots & \vdots\\ \vdots & \ddots & \ddots & \gamma_1\\ \gamma_{T-1} & \cdots & \gamma_1 & \gamma_0\end{pmatrix}=\Sigma \quad (6)$$

The Toeplitz matrix presented by $\mathrm{T}\left[\gamma_0,\ldots,\gamma_{T-1}\right]$ under the normality assumption of:

$$Y \sim N_T(\mu,\Sigma) \quad (7)$$

The variance matrix of joint distribution as shown in 4.46 combined with Toeplitz matrix, shows as the log-likelihood equation as follows:

$$\log L\left(d,\phi,\theta,\beta,\sigma_\varepsilon^2\right)=-\frac{T}{2}\log(2\pi)-\frac{1}{2}\log|\Sigma|-\frac{1}{2}z'\Sigma^{-1}z \quad (8)$$

Therefore, the ARFIMA model proposed by Doornik and Ooms (2004) is a very powerful model to predict the time series data that has a strong persistency towards the nonstationary. Besides, the proposed ARFIMA model also solved two issues in implementing the MLE compared to the existing ARFIMA model. Moreover, the ARFIMA model which consists of the elements of d for the ranging between $(0.0 < d < 0.5)$ is good in capturing the time series data that are persistence towards the nonstationary and has been considered by a number of literatures in many fields of time series study[1].

2.3 Forecasting evaluation criterions

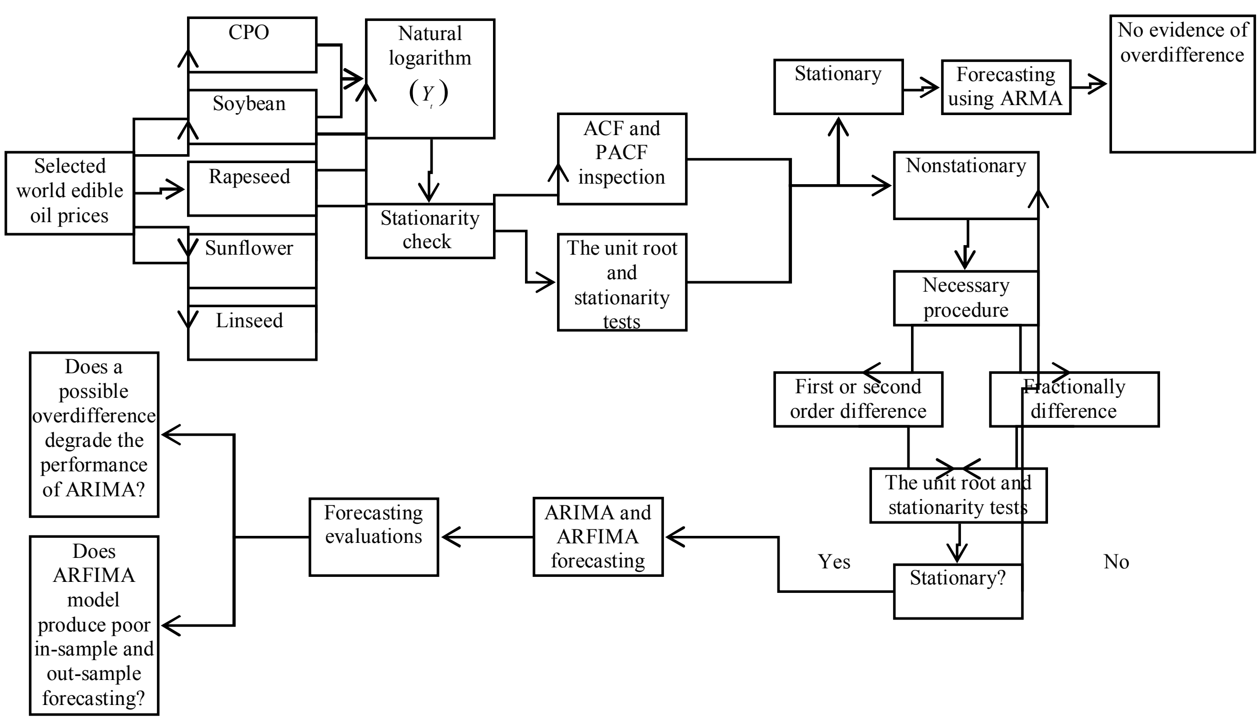
This study utilized the root mean squared error (RMSE) to evaluate the performance of the ARIMA and ARFIMA models in predicting in-sample and out-sample for five of the selected edible oil prices. This statistical evaluation criterion can be derived as follows:

$$RMSE = \sqrt{\frac{\sum_{t}^{n} e_t^2}{n}} \tag{9}$$

However this study also improvised the use of RMSE to $\%\Delta RMSE$ by means to identify whether the performance of the ARIMA and ARFIMA models degrades as it move from in-sample to out-sample forecasting. The positive sign of $\%\Delta RMSE$ indicates that the model perform poor out-sample forecasting. Meanwhile, negative sign of $\%\Delta RMSE$ reveals that the out-sample forecasting is outperform. Their expressions are given by

$$\%\Delta RMSE = \frac{\sqrt{\frac{\sum_{t}^{n} e_t^2}{n}}_1 - \sqrt{\frac{\sum_{t}^{n} e_t^2}{n}}_0}{\sqrt{\frac{\sum_{t}^{n} e_t^2}{n}}_0} \times 100\% \tag{10}$$

Figure 1: The overall research flow



[1] Refer Doornik and Ooms (2004) for complete explanation on improvise model of ARFIMA.

3. Results and discussions

Table 1showsthe descriptive statistics of the original series (X_t) and transformed into natural logarithm (Y_t) of five world edible oil prices. Figure 2 shows the plots of original daily prices of selected edible oils (X_t) in Malaysia. This figure indicates that five of the selected oils prices show a decreasing trend from half year 2008 to early of 2009. This figure also proved that five of the selected oil prices show similar movement as increase in one price will lead to increase in another prices and vice versa.

Table 1: Descriptive statistics of the original series X_t and $Y_t = \log(X_t)$ for five of the selected world edible oil prices (from 1 January 2008 to 31 December 2013)

Descriptive Statistics	CPO		Soybean		Rapeseed		Sunflower		Linseed	
	Series (X_t)	Series (Y_t)	Series (X_t)	Series (Y_t)	Series (X_t)	Series (Y_t)	Series (X_t)	Series (Y_t)	Series (X_t)	Series (Y_t)
Mean	867.370	6.736	479.936	6.160	1145.245	7.023	1090.024	6.965	1000.270	6.886
Median	805.000	6.691	490.000	6.194	1178.665	7.072	1100.000	7.003	1040.000	6.947
Max	1350.000	7.208	683.000	6.526	1639.440	7.402	1915.000	7.557	1420.000	7.258
Min	390.000	5.966	196.000	5.278	704.760	6.558	612.000	6.417	570.000	6.346
Std. Dev.	203.5712	0.249	76.554	0.166	225.164	0.201	267.204	0.243	203.686	0.213
N	1566	1566	1566	1566	1566	1566	1566	1566	1566	1566

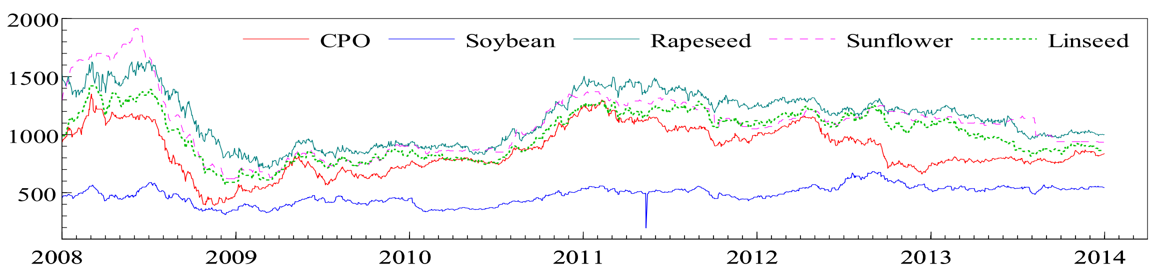


Figure 2: Daily Prices of selected world edible oil in free-on-board MYR/US$ per metric ton from 31 January 2008 to 31 December 2013

Examining the time series using the autocorrelation function (ACF) and partial autocorrelation function (PACF) are important since, (1) it helps to identify the order of p and q for ARMA model (Ibrahim et al., 2015), (2) identifying the stationarity of time series data (Zhang, Pang, Cui, Stallones, & Xiang, 2015) and indicating the tendency of over difference (Karia et al., 2013).Figure 2 shows the ACF and PACF inspection on five of the selected edible oils prices that have been transformed into natural logarithm (Y_t). This figure reveals that all of the data show a covariance stationary of the CPO, soybean, rapeseed, sunflower and linseed exhibits statistically significant dependence between the observations. We also found that five of the selected edible oils prices ACF's demonstrated decays at a hyperbolic rate than short memory time series data. Besides we detect several spikes on PACF that are ranging of $0 \le p \le 2$. The illustrations from the ACF and PACF proved that five of the selected edible oils are nonstationary and need for the necessary procedure of differencing. Other than that, an illustration from the ACF also showed that it decays at a hyperbolic rate whereby it gives an indication of possible of over difference(Arouri et al., 2012; Diebold & Inoue, 2001; Kwan et al., 2012; Maqsood & Burney, 2014; Perron & Qu, 2007; Tan et al., 2012; Xiu & Jin, 2007).

In order to meet stationarity assumption, this study utilized the necessary procedure as conducted by Karia et al. (2013)that uses first or second order difference and fractionally difference. The first or second order difference will be estimated using integer of d that are $d=1$ or $d=2$, generally. The fractionally difference will be estimated using non-integer of d with $0<d<0.5$ will be considered as stationarity series (Doornik & Ooms, 2004). Table 2 illustrates the perspective of fractionally difference parameter values based from the study of Coleman and Sirichand (2012) and Tkacz (2001).

Table 2: Perspective in determining the fractionally difference parameter values

d	Variance	Shock duration	Stationarity
$d=0$	Finite	Short-lived	Stationary
$0<d<0.5$	Finite	Long-lived	Stationary
$0.5\le d<1$	Infinite	Long-lived	Nonstationary
$d=1$	Infinite	Infinite	Nonstationary
$d>1$	Infinite	Infinite	Nonstationary

Source: Coleman and Sirichand (2012) and Tkacz (2001)

Detecting the stationarity is important since the autoregressive moving average model (ARIMA) and autoregressive fractionally integrated model (ARFIMA) assume stationarity time series data. Relying to the Figure 3, the ACF and PACF inspection reveals that five of the selected edible oil prices show tendency of over difference and nonstationarity. Therefore, this study utilized the unit root and stationarity tests for five of the time series data. Since there is no predetermined set of rules on which of the particular unit root and stationarity tests to be adopted for five of the selected edible oil prices, this study consider the augmented Dickey and Fuller(1981) [ADF] test in detecting the existence of unit root. While, this study decided to implement the Kwiatkowski, Phillips, Schmidt and Shin (1992) [KPSS] for stationarity test.

Table 3 shows the results obtained from the ADF and KPSS tests on five of the selected edible oils prices for original series, first order difference and fractionally difference at Y_t. The ADF test results toward the original series of five of the selected edible oils prices shows that there is no evidence of significant difference between the computed values of statistics with the critical value at 1%, 5% and 10% level. In addition, the P-value are also insignificant for five of the time series data. Based from the ADF test, five of the original series at Y_t, that are CPO, soybean, rapeseed, sunflower and linseed are insignificant and nonstationary as it fails to reject the H_0 of time series has unit root. Observing the KPSS test for original series at Y_t, five of the time series data indicates significant at 1% level. Therefore, we reject the H_0 of time series is stationary at 99% confidence interval. Thus, KPSS test has confirmed that the original series of five selected edible oil prices are nonstationary. The results from the unit root and stationarity tests are consistent with the ACF and PACF inspection. As a result it needs the necessary procedure of first order difference and fractionally difference as fulfilling the assumption of ARIMA and ARFIMA model.

As mentioned previously, ensuring the stationarity of the time series data is vital in fulfilling the assumption of the ARIMA and ARFIMA model. Therefore we start with the necessary procedure of first order difference towards five of the selected edible oil prices. The result demonstrated in Table 3. The ADF test shows that five of the selected edible oil prices are significant at 1% level. As a result, ADF test reject the H_0 of time series has unit root at 99% confidence interval. With this, the ADF test has confirmed that the effort of first order difference is stationary for five of the time series data. The KPSS test for five of the time series data shows insignificant either at 1%, 5% and 10% level. Therefore we do not reject H_0 of time series is stationary. In regard to this matter, the ADF and KPSS test have confirmed that five of the time series data are stationary at first order difference. However, we found that there is a tendency of possible over difference as reported in Figure 3. Since the ADF test show large values of their statistics and KPSS test show small value of statistic. This results tend to be consistent with the previous study byKaria et al. (2013).

Next we proceed with the analysis of necessary procedure of fractionally difference. From it, we found that the value of noninteger of d for CPO, soybean, rapeseed, sunflower and linseed are $d = 0.1179$, $d = 0.0359$, $d = 0.2274$, $d = 0.2491$ and $d = 0.1444$ respectively. All of the five time series data show the noninteger value of d that are still within the range of $0 < d < 0.5$. Relying to the perspective of Coleman and Sirichand (2012) and Tkacz (2001), five of the time series data are long-lived and stationary. Considering the results of ADF test, all of the fractionally difference are significant at 5% level for five of the time series data. Thus, we reject the H_0 of time series has unit root at 95% confidence interval. The ADF test confirm that five of the time series data are found stationary. Analyzing the KPSS test found that there are significant at 10% level for five of the time series data. Therefore the KPSS test reject the H_0 of stationary at 90% confidence interval. Considering the results of ADF and KPSS tests, we conclude that the fractionally difference towards five of the time series data are found stationary.

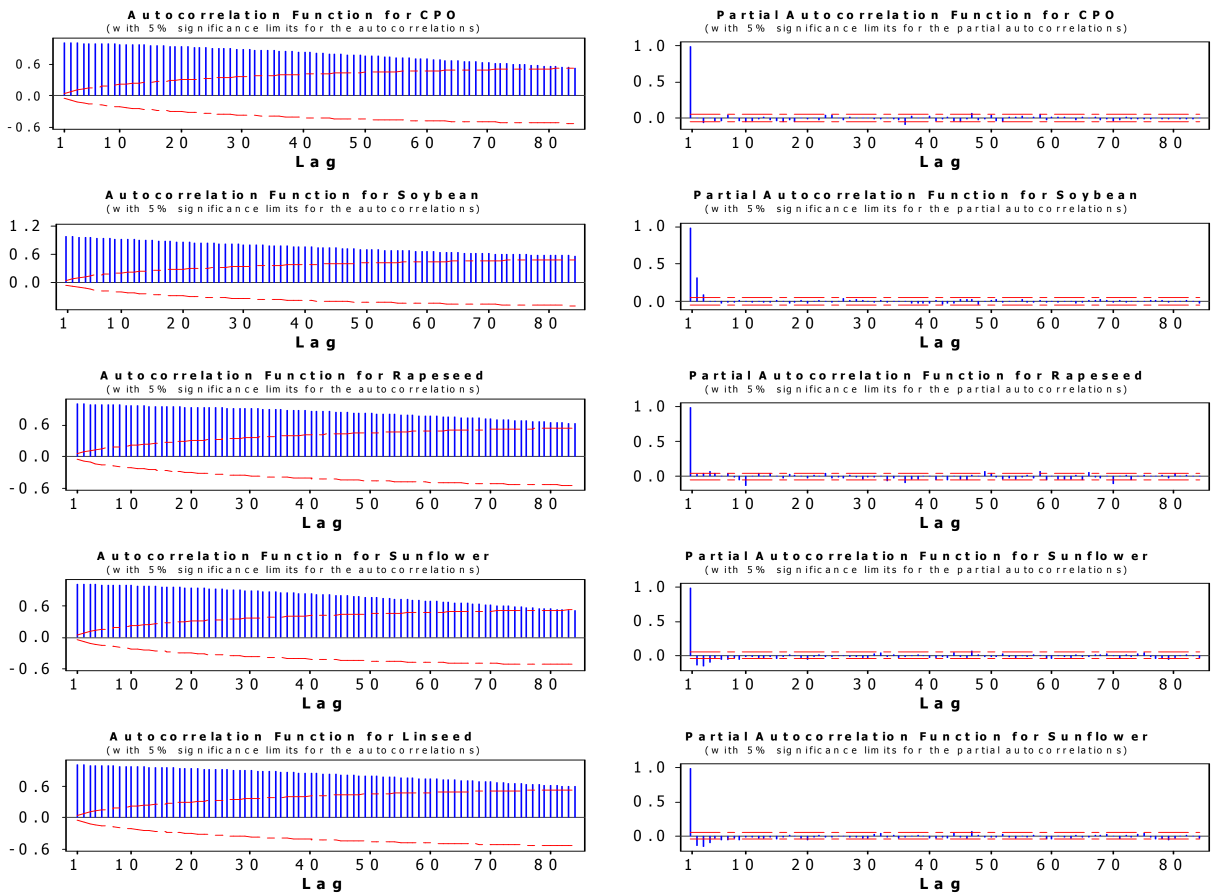


Figure 3: Autocorrelation function (ACF) and partial autocorrelation function (PACF) of original series (Y_t) for five of the selected world edible oil prices

Table 3: The unit root and stationarity tests for selected world edible oil prices of original series, first order difference and fractionally difference at Y_t

Time series	Test	Value of statistic	1% Critical value	5% Critical value	10% Critical value	P-value[a]
CPO prices						
Original series	ADF	-1.453	-3.964	-3.413	-3.128	0.845
(Y_t)	KPSS	0.410***	0.216	0.146	0.119	-
First order difference	ADF	-17.676***	-3.964	-3.413	-3.128	0.001
$(d=1)$	KPSS	0.115	0.216	0.146	0.119	-
Fractionally difference	ADF	-3.435**	-3.964	-3.413	-3.128	0.047
$(d=0.1179)$	KPSS	0.146*	0.216	0.146	0.119	-
Soybean prices						
Original series	ADF	-3.077	-3.964	-3.413	-3.128	0.112
(Y_t)	KPSS	0.333***	0.216	0.146	0.119	-
First order difference	ADF	-39.669***	-3.964	-3.413	-3.128	0.001
$(d=1)$	KPSS	0.033	0.216	0.146	0.119	-
Fractionally difference	ADF	-3.489**	-3.964	-3.413	-3.128	0.041
$(d=0.0359)$	KPSS	0.144*	0.216	0.146	0.119	-
Rapeseed prices						
Original series	ADF	-1.992	-3.964	-3.413	-3.128	0.605
(Y_t)	KPSS	0.468***	0.216	0.146	0.119	-
First order difference	ADF	-42.101***	-3.964	-3.413	-3.128	0.001
$(d=1)$	KPSS	0.118	0.216	0.146	0.119	-
Fractionally difference	ADF	-3.459**	-3.964	-3.413	-3.128	0.044
$(d=0.2274)$	KPSS	0.143*	0.216	0.146	0.119	-
Sunflower prices						
Original series	ADF	-1.515	-3.964	-3.413	-3.128	0.824
(Y_t)	KPSS	0.396***	0.216	0.146	0.119	-
First order difference	ADF	-17.430***	-3.964	-3.413	-3.128	0.001
$(d=1)$	KPSS	0.118	0.216	0.146	0.119	-
Fractionally difference	ADF	-3.462**	-3.964	-3.413	-3.128	0.044
$(d=0.2491)$	KPSS	0.145*	0.216	0.146	0.119	-
Linseed prices						
Original series	ADF	-1.379	-3.964	-3.413	-3.128	0.867
(Y_t)	KPSS	0.429***	0.216	0.146	0.119	-
First order difference	ADF	-36.111***	-3.964	-3.413	-3.128	0.001
$(d=1)$	KPSS	0.118	0.216	0.146	0.119	-
Fractionally difference	ADF	-3.599**	-3.964	-3.413	-3.128	0.030
$(d=0.1444)$	KPSS	0.145*	0.216	0.146	0.119	-

Note:[a]Based from MacKinnon (1996) one-sided p-values. The critical values are based on percentage levels of 1%, 5% and 10%, which correspond to 99%, 95% and 90% of confidence level.
* Significant at levels of 10%

**Significant at levels of 5%
*** Significant at levels of 1%.

The results from the ACF and PACF inspections together with analysis of unit root and stationarity tests suggest for the necessary procedure of first order difference $(d = 1)$ and fractionally difference of order d stat ranging $0 < d < 0.5$. Now for the first order difference, we found that all of the selected five edible oil prices displaying a stationary pattern. However, from Figure 4 it clearly shows that it reducing the original trend characteristics for five of the selected edible oil prices. Besides, the efforts of first order differencing were not only attenuated but nearly annihilated the characteristics like a trend for five of the time series data. We believed that the first order difference seems to eliminate too much of the important information from the original series data. Moreover, we found that the result shows in Figure 4 is consistent with the results of ACF and PACF and unit root and stationarity tests. The effort of first order difference seems to demonstrate the tendency of over difference as the time series is in long memory or long-lived duration. In this study we also intend to compare the ARIMA and ARFIMA performance which covered in-sample and out-sample forecasting.

For the ARFIMA model, we obtain the non-integer d from package developed by Doornik and Ooms (2004). The value of non-integer d for CPO, soybean, rapeseed, sunflower and linseed can be derived by the following models respectively.

$$\Phi(L)(1-L)^{0.1179}(Y_t-\mu_t)=\Theta(L)\varepsilon_t \tag{11}$$

$$\Phi(L)(1-L)^{0.0359}(Y_t-\mu_t)=\Theta(L)\varepsilon_t \tag{12}$$

$$\Phi(L)(1-L)^{0.2274}(Y_t-\mu_t)=\Theta(L)\varepsilon_t \tag{13}$$

$$\Phi(L)(1-L)^{0.2491}(Y_t-\mu_t)=\Theta(L)\varepsilon_t \tag{14}$$

$$\Phi(L)(1-L)^{0.1444}(Y_t-\mu_t)=\Theta(L)\varepsilon_t \tag{15}$$

The resulting series from fractionally differencing towards five of the selected edible oil prices are shows in Figure 4. The results indicate that there is not much loss in important data if we compared it with the first order difference. This is because the necessary procedure of first order difference is still displaying the characteristic like the trend for five of time series data.

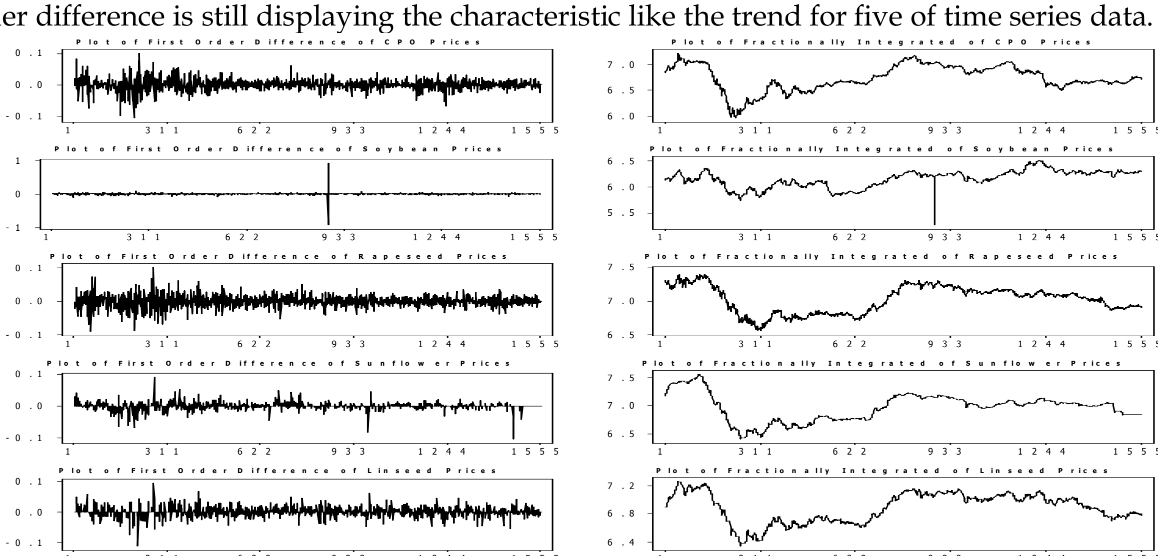


Figure 4: Plots of the first order difference and fractionally difference for five of the selected world edible oil prices at Y_t

Table 4 reported the in-sample and out-sample forecasting performances from ARIMA and ARFIMA models in predicting five of the selected edible oil prices at Y_t.The evidence demonstrated from this table show mixed results. We found that the ARIMA model is outperformed for the in-sample and out-sample predictions of CPO and linseed prices. Meanwhile, the ARFIMA is better fit for the rapeseed prices predictions. Besides, we found the inconsistent results for the soybean prices prediction since its demonstrated that in-sample and out-sample are associated for the ARIMA and ARFIMA respectively. This inconsistency also revealed for the sunflower prices prediction whereby the ARFIMA and ARIMA are associated for the in-sample and out-sample prediction. These inconsistencies shows a similarity with the case addressed by the study of Kang and Yoon (2013). We do not find any clear model for the edible oils prediction.

The evidence from the analysis of percentage change in root mean squared error $(\%\Delta RMSE)$ also revealed mixed results. From it, we found that there are negative and positive signs in ARIMA and ARFIMA models respectively. It is found that CPO, soybean, rapeseed and sunflower prices prediction demonstrated $-\%\Delta RMSE$. It gives the impression that the ARIMA model is performing better in out-sample forecasting. Whereby the linseed prices suggesting an opposing result of $+\%\Delta RMSE$ that indicates the out-sample prediction using ARIMA model is degraded. In one hand, the ARFIMA model also suggesting similar result. This model depict $+\%\Delta RMSE$ for the CPO and linseed prices. Meanwhile, there are $-\%\Delta RMSE$ for the soybean and sunflower prices. However, there is a mixed result of $\%\Delta RMSE$ for the rapeseed prices.

From the analyses of the Table 4, we found that:

(1) the ARFIMA model do not show poor out-sample prediction from that have been found by the study of Xiu and Jin (2007) and Ellis and Wilson (2004) and the reference therein. The ARFIMA model shows decent result and its performance is slightly different with the ARIMA model. Similar with the ARIMA model, it showed $+\%\Delta RMSE$ and $-\%\Delta RMSE$ that gives impression of degrades and performing better in out-sample forecasting, respectively.

(2) The tendency of over difference seems not to give a significant impact toward neither ARIMA nor ARFIMA models. This proven with the results of ARIMA and ARFIMA that displayed mixed results although the analyses of ACF and PACF, and unit root and stationarity tests indicated the tendency of over difference.

(3) Consistent with the study byMaqsood and Burney (2014), we found that the ARIMA model is healthier model in forecasting world edible oils prices due to its simplicity rather than complex ARFIMA model.

Table 4: The ARIMA and ARFIMA forecasting performances in predicting five of the selected world edible oil prices from 1 January 2008 to 31 December 2013

World edible oils	Model	In-sample	Out-sample*	$\%\Delta RMSE$	In-sample		Out-sample	
		$RMSE_0$	$RMSE_1$		ARIMA	ARFIMA	ARIMA	ARFIMA
CPO prices	ARIMA(1,1,0)	0.017128	0.009000	-47.45				
	ARIMA(1,1,1)	0.017043	0.010706	-37.18				
	ARIMA(2,1,0)	0.017090	0.008840	-48.27				
	ARIMA(2,1,1)	0.017067	**0.008695**	-49.05			✓	
	ARIMA(2,1,2)	**0.017010**	0.011328	-33.40	✓			
	ARFIMA(1,0.1179,0)	0.017073	0.030308	77.52				
	ARFIMA(1,0.1179,1)	0.017074	0.030323	77.60				
	ARFIMA(2,0.1179,0)	0.017073	0.030300	77.47				

	ARFIMA(2,0.1179,1)	0.017056	**0.030248**	77.35				
	ARFIMA(2,0.1179,2)	**0.017039**	0.030373	78.26				
Soybean prices	ARIMA(1,1,0)	0.033442	**0.006761**	-79.78				
	ARIMA(1,1,1)	0.032796	0.006952	-78.80				
	ARIMA(2,1,0)	0.032877	0.006762	-79.43				
	ARIMA(2,1,1)	**0.032790**	0.006910	-78.93	✓			
	ARIMA(2,1,2)	**0.032790**	0.006899	-78.96				
	ARFIMA(1,0.0359,0)	0.033630	**0.005708**	-83.03				✓
	ARFIMA(1,0.0359,1)	0.032853	0.006889	-79.03				
	ARFIMA(2,0.0359,0)	0.032976	0.006120	-81.44				
	ARFIMA(2,0.0359,1)	**0.032848**	0.006821	-79.23				
	ARFIMA(2,0.0359,2)	**0.032848**	0.006822	-79.23				
Rapeseed prices	ARIMA(1,1,0)	0.016567	**0.003400**	-79.48				
	ARIMA(1,1,1)	**0.016516**	0.004310	-73.90				
	ARIMA(2,1,0)	0.016552	0.004004	-75.81				
	ARIMA(2,1,1)	0.016552	0.004004	-75.81				
	ARIMA(2,1,2)	-	-	-				
	ARFIMA(1,0.2274,0)	0.000287	0.018036	6184.32				
	ARFIMA(1,0.2274,1)	0.016518	0.018072	9.41				
	ARFIMA(2,0.2274,0)	0.016820	**0.001448**	-91.39				✓
	ARFIMA(2,0.2274,1)	0.016514	0.021065	27.56				
	ARFIMA(2,0.2274,2)	**0.016512**	0.018043	9.27		✓		
Sunflower prices	ARIMA(1,1,0)	0.010275	0.003945	-61.61				
	ARIMA(1,1,1)	0.010069	0.004479	-55.52				
	ARIMA(2,1,0)	0.010140	**0.003902**	-61.52			✓	
	ARIMA(2,1,1)	0.010068	0.004418	-56.12				
	ARIMA(2,1,2)	**0.010066**	0.004421	-56.08				
	ARFIMA(1,0.2491,0)	0.010073	**0.005202**	-48.36				
	ARFIMA(1,0.2491,1)	0.010072	0.005207	-48.30				
	ARFIMA(2,0.2491,0)	0.010073	0.005207	-48.31				
	ARFIMA(2,0.2491,1)	0.010073	0.005210	-48.28				
	ARFIMA(2,0.2491,2)	**0.010061**	0.005243	-47.89		✓		
Linseed prices	ARIMA(1,1,0)	0.014267	0.024042	68.51				
	ARIMA(1,1,1)	0.014245	0.023205	62.90				
	ARIMA(2,1,0)	0.014255	0.024047	68.69				
	ARIMA(2,1,1)	0.014245	0.023208	62.92				
	ARIMA(2,1,2)	**0.014244**	**0.023140**	62.45	✓		✓	
	ARFIMA(1,0.1444,0)	0.014268	0.038263	168.17				
	ARFIMA(1,0.1444,1)	0.014256	**0.038163**	167.70				
	ARFIMA(2,0.1444,0)	0.014263	0.038206	167.87				
	ARFIMA(2,0.1444,1)	0.014263	0.038207	167.87				
	ARFIMA(2,0.1444,2)	**0.014255**	0.038176	167.81				
Total outperformed model					3	2	3	2

Note: * Indicated 10-step ahead forecasting. Whereby researchers will keep 10 last observations from original series of Y_t and will be compared with the output from ARIMA and ARFIMA models.

4. Conclusion

In this study we sought to identify a good model in predicting the time series data that observed with the tendency of over difference and long memory behavior. Whereby the previous sections demonstrated that five of the selected world edible oil prices,that are CPO, soybean, rapeseed, sunflower and linseed prices are predicted using ARIMA and ARFIMA models. We also found that five of these time series data demonstrated highly persistence towards the nonstationarity. Moreover, the analysis of ACF indicated decays at a very hyperbolic rate, in which giving the impression of long memory behaviour and persistence towards nonstationary. Therefore it needs a necessary procedure of differencing as means in fulfilling the assumption of the Box and Jenkins (1976) model. Whereby, this study considered first order difference and fractionally difference. Consistent with the evidence from ACF and PACF inspection, the necessary procedure of first order difference seems to be a good solution in nonstationary behaviour, but the unit root and stationarity tests proved there is a presence of overdifference.

Whilstmethodology seems to be sound, we found mixed results. The addressed overdifference behaviour seems not to give a significant impact toward neither ARIMA nor ARFIMA models. As mentioned by Karia et al. (2013) and Maqsood and Burney (2014), the first order difference is resposible for the loss of important informationof specific time series data, but we are not certain of which of the time series data that will be affected. Moreover, for the case of five selected world edible oil prices, we found the performance from both models demonstrated almost similar result.

In this study, we also found that there is no evidence that the ARFIMA model shows poor in-sample and out-sample prediction considering for specific time-span in our analysis. Even though the ARIMA model suffers from the possible over difference, while ARFIMA model proven stationary with the perspective of Coleman and Sirichand (2012) and Tkacz (2001),ARIMA is the healthier model in predicting five selected world edible oil prices. However the performances from both of these models demonstrated decent and slightly different results. Therefore this study is strongly recommends that the ARIMA model is implemented due to its simplicity in predicting the world edible oil prices. Moreover, this study also suggests that the tendency of over difference be seriously studied in future work as means improving the existing model.

Acknowledgement

We thank the anonymous referee for the valuable comments and suggestions. The usual disclaimers apply.

References

Arouri, M. E. H., Hammoudeh, S., Lahiani, A., & Nguyen, D. K. (2012). Long memory and structural breaks in modeling the return and volatility dynamics of precious metals. *The Quarterly Review of Economics and Finance, 52*, 207-218.

Baillie, R. T., & Chung, S. K. (2002). Modeling and forecasting from trend-stationary long memory models with applications to climatology. *International Journal of Forecasting, 18*, 215-226.

Bhardwaj, G., & Swanson, N. R. (2006). An empirical investigation of the usefulness of ARFIMA models for predicting macroeconomic and financial time series. *Journal of Econometrics, 131*, 539-578.

Box, G. E. P., & Jenkins, G. M. (1976). *Time series analysis forecasting and control* (2nd ed.). San Francisco, CA: Holden-Day.

Chortareas, G., Jiang, Y., & Nankervis, J. C. (2011). Forecasting exchange rate volatility using high-frequency data: Is the euro different? *International Journal of Forecasting, 27*, 1089-1107.

Chu, F. L. (2008). A fractionally integrated autoregressive moving average approach to forecasting tourism demand. *Tourism Management, 29*, 79-88.

Coleman, S., & Sirichand, K. (2012). Fractional integration and the volatility of UK interest rates. *Economics Letters, 116*, 381-384.

Dickey, D. A., & Fuller, W. A. (1981). Likelihood ratio statistics for autoregressive time series with a unit root. *Econometrica, 49*(4), 1057-1072.

Diebold, F. X., & Inoue, A. (2001). Long memory and regime switching. *Journal of Econometrics, 105*, 131-159.

Doornik, J. A., & Ooms, M. (2004). Inference and forecasting for ARFIMA models with an application to US and UK inflation. *Studies in Nonlinear Dynamics & Econometrics, 8*(2), 1-23.

Ellis, C., & Wilson, P. (2004). Another look at the forecast performance of ARFIMA models. *International Review of Financial Analysis, 13*, 63-81.

Erfani, A., & Samimi, A. J. (2009). Long memory forecasting of stock price index using a fractionally differenced ARMA model. *Journal of Applied Sciences Research, 5*(10), 1721-1731.

Hosking, J. R. M. (1981). Fractional differencing. *Biometrika, 68*, 165-176.

Hurvich, C. M., & Ray, B. K. (1995). Estimation of the memory parameter for nonstationary or noninvertible fractionally integrated processes. *Journal of Time Series Analysis, 16*, 17-41.

Ibrahim, M., Jamil, M., Akhtar, A. M., Mir, Z.-u.-H., Akbar, S., & Ahmad, M. (2015). Patient's influx at outpatient medical laboratory (OPML); forecasting of patient's influx at mayo hospital, Lahore using ARIMA model. *The Professional Medical Journal, 22*(6), 767-775.

Kang, S. H., & Yoon, S.-M. (2013). Modeling and forecasting the volatility of petroleum futures prices. *Energy Economics, 36*, 354–362.

Karia, A. A., Bujang, I., & Ahmad, I. (2013). Fractionally integrated ARMA for crude palm oil prices prediction: case of potentially overdifference. *Journal of Applied Statistics, 40*(12), 2735-2748.

Koopman, S. J., Jungbacker, B., & Hol, E. (2005). Forecasting daily variability of the S&P 100 stock index using historical, realised and implied volatility measurements. *Journal of Empirical Finance, 12*(3), 445-475.

Kwan, W., Li, W. K., & Li, G. (2012). On the estimation and diagnostic checking of the ARFIMA-HYGARCH model. *Computational Statistics and Data Analysis 56*, 3632-3644.

Kwiatkowski, D., Phillips, P. C. B., Schmidt, P., & Shin, Y. (1992). Testing the null hypothesis of stationarity against the alternative of a unit root. *Journal of Econometrics, 54*, 159-178.

Maqsood, A., & Burney, S. M. A. (2014). Long memory forecasting: An application to KSE share index. *Pak. J. Statist., 30*(3), 333-344.

Mostafaei, H., & Sakhabakhsh, L. (2011). Modelling and forecastng of OPEC oil prices with ARFIMA model. *International Journal of Academic Research 3*(1), 817-822.

Ooms, M., & Doornik, J. (1999). Inference and forecasting for fractional autoregressive integrated moving average models, with an application to US and UK inflation. *Econometric Institute Report 9947/A, Erasmus University Rotterdam.*

Perron, P., & Qu, Z. (2007). An analytical evaluation of the log-periodogram estimate in the presence of level shifts. *Boston University.*

Reisen, V. A., & Lopes, S. (1999). Simulations and applications of forecasting long-memory time series models. *Journal of Statistical Planning and Inference 80*, 269-287.

Sowell, F. (1987). Maximum likelihood estimation of fractionally integrated time series models. *Discussion Paper 87-07, Department of Economics, Duke University*, 1-31.

Tan, P. P., Galagedera, D. U. A., & Maharaj, E. A. (2012). A wavelet based investigation of long memory in stock returns. *Physica A 391*, 2330-2341.

Tkacz, G. (2001). Estimating the fractional order of integration of interest rates using wavelet OLS estimator. *Studies in Nonlinear Dynamics and Econometrics, 5*, 1-21.

Xiu, J., & Jin, Y. (2007). Empirical study of ARFIMA model based on fractional differencing. *Physica A, 377*, 138-154.

Zhang, X., Pang, Y., Cui, M., Stallones, L., & Xiang, H. (2015). Forecasting mortality of road traffic injuries in China using seasonal autoregressive integrated moving average model. *Annals of Epidemiology, 25*(2), 101–106.