Analyzing the Effect of Dual Long Memory Process in Forecasting Agricultural Prices in Different Markets of India

Ranjit Kumar Paul1, Bishal Gurung2, Sandipan Samanta3

Abstract

The potential presence of long memory (LM) properties in mean and volatility of the spot price of wheat and mustard in different markets of India has been investigated. The findings revealed the evidence of long range dependence in price series as well as in the volatility. Accordingly, Autoregressive fractionally integrated moving average (ARFIMA) with error following Fractionally integrated generalized autoregressive conditional heteroscedastic (FIGARCH) model has been applied for forecasting the price of commodities in different markets of India. To this end, evaluation of forecasting is carried out with root mean squares error (RMSE), mean absolute error (MAE) and relative mean absolute prediction error (RMAPE). The residuals of the fitted models were used for diagnostic checking.

Keywords: Long memory, conditional heteroscedastic, FIGARCH, wheat, mustard, stationarity

1. Introduction

Long run persistence or long memory in volatility has important implications for modelling and forecasting future volatility. In recent years, developing markets in general and emerging markets in particular has been the subject of close scrutiny by many looking to diversify their portfolios. Developing markets have increasingly attracted the interest of many foreign investors not only because of the relatively higher returns that they offer but also because of the low correlations with developed markets that lead to better diversification benefits. Therefore, a study of long memory that indicates predictability in the long horizon for key developing markets of the world should be of use to many investors and financial practitioners. There has been a large amount of research on long memory in economic and financial time series. The presence of long memory in asset returns has important implications for many of the models used in modern financial economics. Long memory tests that have been frequently used in the literature include the classical periodogram based estimator [1], the modified rescaled range (R/S) statistic [2], the rescaled variance (V/S) statistic [3], and the robust semiparametric procedure [4]. For modelling the time series in presence of long memory, the autoregressive fractionally integrated moving-average (ARFIMA) model is used. Many studies have empirically examined the long memory property of stock return volatility. The existence of non-zero \( d \) is an indication of long memory and its departure from zero measures the strength of long memory. ARFIMA model for forecasting of agricultural commodity prices has been applied in literature [5], [6] and [7]. However, ARFIMA model is based on some crucial assumptions like linearity, stationarity and homoscedastic errors. Further, time series data quite often exhibits features like long memory in volatility; which cannot be explained by ARFIMA model. Sometime asymmetric phenomena arises with economic series, which tend to behave differently when economy is moving into recession rather than when coming out of it. Many financial time series shows periods of stability followed by unstable periods with high volatility. To take care of the volatility, autoregressive conditional heteroscedastic (ARCH) model was proposed [8]. But, ARCH model has the property that the unconditional autocorrelation function of squared

1ICAR-Indian Agricultural Statistics Research Institute New Delhi, India
2ICAR-Indian Agricultural Statistics Research Institute New Delhi, India
3ICAR-Indian Agricultural Statistics Research Institute New Delhi, India
residuals; if it exists; decay very rapidly compared to what is typically observed unless maximum lag is large. To overcome the weaknesses of ARCH model, the Generalized ARCH (GARCH) model was developed [9] and [10]. Huge amount of empirical and theoretical research work has been already done for GARCH and related models. It seems a long memory model is more flexible than an ARCH model in terms of capturing irregular behaviour.

Fractional Integrated Generalized Autoregressive Conditional Heteroskedastic (FIGARCH) model [11] was applied in some studies [12], [13], [14], [15], [16], [17], [18] and [19] by using one or more of the long memory tests. They test for long memory in return volatility using high frequency daily data mainly for aggregate stock indices for both developed and less developed markets. With this type of estimation, it is not necessary to proxy the volatility series of returns. Instead, volatility is modeled from the data itself and the focus is on the significance and the magnitude of the fractional integration parameter $d$. FIGARCH model is capable of explaining and representing the observed temporal dependencies of the financial market volatility in a much better way than other types of GARCH models. FIGARCH model was applied for describing fourteen agricultural future price series [20]. In the present investigation, an attempt has been made to apply FIGARCH model for modelling and forecasting of long memory volatile price data of Mustard and wheat in different markets of India. Autoregressive fractionally integrated moving average (ARFIMA) model with error follows fractionally integrated GARCH (FIGARCH) model that accounts for long memory in the mean equation as well as long memory in volatility equation is used for the empirical estimations. There is strong evidence of long run persistence in volatility for all the markets.

2. Long Memory Process

Long memory in time-series can be defined as autocorrelation at long lags [21]. Memory means that observations are not independent (each observation is affected by the events that preceded it). The acf of a time-series $y_t$ is defined as

$$
\rho_k = \frac{\text{cov}(y_t, y_{t-k})}{\text{var}(y_t)} \quad (1)
$$

for integer lag $k$. A covariance stationary time-series process is expected to have autocorrelations such that \( \lim_{k \to \infty} \rho_k = 0 \). Most of the well-known class of stationary and invertible time-series processes have autocorrelations that decay at the relatively fast exponential rate, so that $\rho_k \approx |m|^k$, where $|m|<1$ and this property is true, for example, for the well-known stationary and invertible ARMA($p,q$) process. For long memory processes, the autocorrelations decay at an hyperbolic rate which is consistent with $\rho_k \approx Ck^{2d-1}$, as $k$ increases without limit, where C is a constant and $d$ is the long memory parameter.

Suppose that \( \{Y_t\} \) is a stationary process with the spectral density function (SDF) denoted by $S_Y(\cdot)$, then \( \{Y_t\} \) is a stationary long memory process if there exist constants $a$ and $C_S$ satisfying $-1 < a < 0$ and $C_S > 0$ such that

$$
\lim_{f \to 0} S_Y(f)/(C_S |f|^a) = I \quad (2)
$$

In other words, a stationary long memory process has an SDF $S_Y(\cdot)$ such that $S_Y(\cdot) \approx C_S |f|^a$, with the approximation improving as $f$ approaches zero. An alternative definition can be stated in terms of the autocovariance sequence (ACVS) \( \{S_{Y,t}\} \) for \( \{Y_t\} \). \( \{Y_t\} \) is a stationary long memory process if there exist constants $b$ and $C_S$ satisfying $-1 < b < 0$ and $C_S > 0$ such that

$$
\lim_{\tau \to 0} S_{Y,t}(\tau)/(C_S \tau^b) = I \quad (3)
$$

where $b$ is related to $a$ in (2) via $b = -a - 1$. Standard time-series models such as stationary autoregressive processes have ACVSs such that $S_{Y,t} \approx C \phi^t$, for large $\tau$, where $C \geq 0$ and $|\phi| < 1$. For a long memory process $S_{Y,t} \approx C_S \tau^b$ for large $\tau$. In both the cases, $S_{Y,t} \to 0$ as $\tau \to \infty$, but the rate of decay toward zero
is much slower for a long memory process, implying that the observations that are widely separated in time can still have a non negligible covariance i.e. the current observations retain some ‘memory’ of the distant past.

We test for long memory components in the price series and volatility of mustard and wheat using the Geweke and Porter-Hudak (GPH) [1] statistic. The test has been extensively used in the literature. For long memory in the volatility process, this test is applied to the squared errors of fitted mean model, which is commonly regarded as a proxy of conditional volatility ([22], [23]).

Let $r_t$ be the price series. The GPH estimator of the long memory parameter $d$ for $r_t$ can be then determined using the following periodogram:

$$
\log[I(w_j)] = \beta_0 + \beta_1 \log\left[\frac{\sin\left(\frac{w_j}{2}\right)}{w_j}\right] + \epsilon_j
$$

where $w_j = 2\pi j/T$, $j = 1, 2, \ldots, n$; $\epsilon_j$ is the residual term and $w_j$ represents the $n = \sqrt{T}$ Fourier frequencies. $I(w_j)$ denotes the sample periodogram defined as

$$
I(w_j) = \frac{1}{2\pi T} \left| \sum_{t=1}^{T} r_t e^{-i w_j t} \right|^2
$$

where $r_t$ is assumed to be a covariance stationary time series. The estimate of $d$, say $\hat{d}_{GPH}$, is $-\hat{\beta}_1$.

**Garch Model**

The ARCH($q$) model for the series $\{\epsilon_t\}$ is defined by specifying the conditional distribution of $\epsilon_t$ given the information available up to time $t - 1$. Let $\psi_{t-i}$ denote this information. ARCH ($q$) model for the series $\{\epsilon_t\}$ is given by

$$
\epsilon_t | \psi_{t-1} \sim N(0, h_t)
$$

$$
h_t = a_0 + \sum_{i=1}^{q} a_i \epsilon_{t-i}^2
$$

where $a_0 > 0$, $a_i \geq 0$ for all $i$ and $\sum_{i=1}^{q} a_i < 1$ are required to be satisfied to ensure nonnegativity and finite unconditional variance of stationary $\{\epsilon_t\}$ series.

In Generalized ARCH (GARCH) model, conditional variance is also a linear function of its own lags and has the following form

$$
\epsilon_t = \xi_t h_t^{1/2}
$$
$$
h_t = a_0 + \sum_{i=1}^{q} a_i \epsilon_{t-i}^2 + \sum_{j=1}^{p} b_j h_{t-j}
$$

$$
= a_0 + a(L)\epsilon_t^2 + b(L)h_t
$$

where $\xi_t \sim \text{IID}(0,1)$. A sufficient condition for the conditional variance to be positive is

$$
a_0 > 0, \ a_i \geq 0, \ i = 1, 2, \ldots, q. \ b_j \geq 0, \ j = 1, 2, \ldots, p \ \text{and} \ a(L) \text{and} \ b(L) \text{are lag operator such that} \ a(L) = a_1 L + a_2 L^2 + \ldots + a_q L^q \ \text{and} \ b(L) = b_1 L + b_2 L^2 + \ldots + b_p L^p. \ \text{For} \ p = 0, \ \text{the process reduces to an ARCH}(q) \ \text{and for} \ p = q = 0, \ \epsilon_t \ \text{is simply a white noise process.}

The GARCH ($p$, $q$) process is weakly stationary if and only if
The conditional variance defined by (3) has the property that the unconditional acf of \( \varepsilon_t^2 \), if it exists, can decay slowly. Huge amount of empirical and theoretical research work has been already done for GARCH and related models [24].

**Testing for ARCH Effects**

Let \( \varepsilon_t \) be the residual series. The squared series \( \{ \varepsilon_t^2 \} \) is then used to check for conditional heteroscedasticity, which is also known as the ARCH effects. To this end, two tests, briefly discussed below, are available. The first one is to apply the usual Ljung-Box statistic \( Q(m) \) to the \( \{ \varepsilon_t^2 \} \) series. The null hypothesis is that the first \( m \) lags of autocorrelation functions of the \( \{ \varepsilon_t^2 \} \) series are zero. The second test for conditional heteroscedasticity is the LM test, which is equivalent to usual F-statistic for testing \( H_0: a_i = 0, i=1,2,\ldots,q \) in the linear regression

\[
\varepsilon_t^2 = a_0 + a_1 \varepsilon_{t-1}^2 + \ldots + a_q \varepsilon_{t-q}^2 + \varepsilon_t, \quad t = q+1,\ldots,T
\]

where \( \varepsilon_t \) denotes error term, \( q \) is prespecified positive integer, and \( T \) is sample size. Let

\[
SSR_0 = \sum_{t=q+1}^{T} (\varepsilon_t^2 - \bar{\varepsilon})^2
\]

where \( \bar{\varepsilon} = \sum_{t=q+1}^{T} \varepsilon_t^2 / T \) is sample mean of \( \{ \varepsilon_t^2 \} \), and \( SSR_j = \sum_{t=q+1}^{T} \hat{\varepsilon}_t^2 \), where \( \hat{\varepsilon}_t \) is least squares residual of (8). Then, under \( H_0 \):

\[
F = \frac{SSR_0 - SSR_j}{q} \frac{T-q-1}{SSR_j}
\]

is asymptotically distributed as chi-squared distribution with \( q \) degrees of freedom. The decision rule is to reject \( H_0 \) if \( F > \chi^2_q(\alpha) \), where \( \chi^2_q(\alpha) \) is the upper \( 100(1-\alpha) \)th percentile of \( \chi^2_q \) or, alternatively, the \( p \)-value of \( F \) is less than \( \alpha \).

**Figarch Process**

The Garch\((p, q)\) process may also be expressed as an ARMA\((m, p)\) process in \( \varepsilon_t^2 \)

\[
[1-a(L)-b(L)]\varepsilon_t^2 = a_0 + [1-b(L)] \nu_t
\]

where \( m = \max(p, q) \) and \( \nu_t = \varepsilon_t^2 - h_t \). The \( \{ \nu_t \} \) process can be interpreted as the “innovations” for the conditional variance, as it is a zero-mean martingale. Therefore, an integrated GARCH\((p, q)\) process can be written as

\[
[1-a(L)-b(L)] [1-L] \varepsilon_t^2 = a_0 + [1-b(L)] \nu_t
\]

The fractionally integrated GARCH or FIGARCH class of models is obtained by replacing the first difference operator \( (1-L) \) in (9) with the fractional differencing operator \( (1-L)^d \), where \( d \) is a fraction \( 0 < d < 1 \). Thus, the FIGARCH class of models can be obtained by considering

\[
[1-a(L)-b(L)] (1-L)^d \varepsilon_t^2 = a_0 + [1-b(L)] \nu_t
\]

Such an approach can develop a more flexible class of processes for the conditional variance that are capable of explaining and representing the observed temporal dependencies of the financial market volatility in a much better way than other types of GARCH models [25].

It may be noted that the fractional differencing operator \( (1-L)^d \) can be written in terms of hypergeometric function.
\[(1 - L)^d = F(-d, 1, 1; L) = \sum_{k=0}^{\infty} \Gamma(k-d) \Gamma(k+1)^{-1} \Gamma(-d)^{-1} L^k\]

The ARFIMA\((p, d, q)\) class of models for the discrete time real-valued process \(\{y_t\}\) [26]; [27, 28] and [29] is defined by

\[a(L)(1 - L)^d y_t = b(L)\xi_t\]  \hspace{1cm} (11)

where \(a(L)\) and \(b(L)\) are polynomials in the lag operator of orders \(p\) and \(q\) respectively, and \(\xi_t\) is a mean-zero serially uncorrelated process. For the ARFIMA models, the fractional parameter \(d\) lies between \(-1/2\) and \(1/2\), [29]. The ARFIMA model is nothing but the fractionally integrated ARMA for the mean process. Analogous to the ARFIMA\((p, d, q)\) process defined in (11) for the mean, the FIGARCH\((p, d, q)\) process for \(\varepsilon_t^2\) can be defined as

\[a(L)(1 - L)^d \varepsilon_t^2 = a_0 + \left[1 - b(L)\right] v_t\]  \hspace{1cm} (12)

where \(0 < d < 1\), and all the roots of \(a(L)\) and \([1 - b(L)]\) lie outside the unit circle. In the case of ARFIMA model, the long memory operator is applied to unconditional mean \(\mu\) of \(y_t\) which is constant. But this is not true in the case of FIGARCH model, where it is not applied to \(\omega_0\), but on squared errors.

Rearranging the terms in (8), an alternative representation for the FIGARCH\((p, d, q)\) model may be obtained as

\[
\begin{align*}
[1 - b(L)] h_t &= a_0 + \left[1 - b(L) - a(L)(1 - L)^d \right] v_t^2,
\end{align*}
\]

where, \(v_t = \varepsilon_t^2 - h_t\).

**Estimation of Figarch Model**

The estimation of parameters of FIGARCH model is generally carried out using the maximum likelihood method (which is most efficient) with normality assumption for \(z\). But the normality assumption can be questioned with some empirical evidence and therefore the use of quasi-maximum likelihood estimator is preferred.

The FIGARCH model is estimated by using the quasi-maximum likelihood (QML) estimation method allowing for asymptotic normality distribution, based on the following log-likelihood function

\[
LL(x, \theta) = -\frac{1}{2} \log(2\pi) - \frac{1}{2} \sum_{t=1}^{T} \log(h_t) + \frac{\varepsilon_t^2}{h_t}\]  \hspace{1cm} (14)

where \(\theta' = \{a_0, d, b_1, b_2, \ldots, b_p, a_1, a_2, \ldots, a_q\}\).

The likelihood function is maximized conditional on the start-up values. For the FIGARCH\((p, d, q)\) model with \(d > 0\), the population variance does not exist. In most practical applications with high frequency financial data, the standardized innovations \(\tilde{\varepsilon}_t = h_t^{-1/2} \varepsilon_t\) are leptokurtic and not normally distributed through time. In these situations the robust quasi-MLE (QMLE) procedures [30] and [31] may give better results while doing inference. When estimating the parameters of a FIGARCH model, generally, the value of parameter \(d\) is estimated first and one uses these estimates to obtain the estimation of other parameters ([32]; [33]).

**Forecasting by Figarch Model**

The one-step ahead forecast [34] of \(h_t\) by FIGARCH model is given by

\[h_t(1) = a_0 \left[1 - b_1\right]^{-1} + \lambda_1 \varepsilon_t^2 + \lambda_2 \varepsilon_{t-1}^2 + \ldots\]

where, \(\lambda_k \approx [1 - b_1] \Gamma(d)^{-1} k^{d-1}\)

Similarly, the two-step ahead forecast is given by
R. K. Paul et al

\[ h_t(2) = a_0 \left(1 - b_1 \right)^{-1} + \lambda_1 \epsilon_{t+1} + \lambda_2 \epsilon_t^2 + \ldots \]

Here \( \epsilon_{t+1} \) is unobservable and to be estimated by its conditional expectation \( h_t(1) \), which is a function of past \( \epsilon_t^2 \).

Therefore,

\[ h_t(2) = a_0 \left(1 - b_1 \right)^{-1} + \lambda_1 h_t(1) + \lambda_2 \epsilon_t^2 + \ldots \]

In general, the \( l \)-step ahead forecast is

\[ h_t(l) = a_0 \left(1 - b_1 \right)^{-1} + \sum_{i=1}^{l-1} \lambda_i h_t(l-i) + \sum_{j=0}^{M} \lambda_{t+i} \epsilon_{t+j}^2 + \ldots \]

For all practical purpose, we stop at a large \( M \) and this leads to the forecasting equation

\[ h_t(l) \approx a_0 \left(1 - b_1 \right)^{-1} + \sum_{i=1}^{l-1} \lambda_i h_t(l-i) + \sum_{j=0}^{M} \lambda_{t+i} \epsilon_{t+j}^2 \]

The parameters will have to be replaced by their corresponding estimates

3. Results and Discussions

Daily time series data for spot prices of Wheat in Najafgarh, Narela, Agar, Aujha, Akbarpur and Alampur and spot price of Mustard in Najafgarh, Narela, Agar, Aujha and Mumbai during 1 January, 2009 to 31 December, 2013 has been considered. The data is collected from Ministry of Consumer’s Affairs, Government of India. The data for the period January 1, 2009 to October 31, 2013 have been used for model building and the remaining data have been used for model validation. The summary statistics for the wheat and mustard price series have been computed and reported in Table 1 and Table 2 respectively. A perusal of table 1 indicates that all the series are positively skewed and platy-kurtic. The daily unconditional volatility, as measured by standard deviations, ranges between 140.213 to 314.379 for wheat and 394.442 to 542.970 for mustard. In order to test for stationarity, two tests namely Augmented Dickey-Fuller unit root test and Philips-Peron unit root test are used. The results of the tests are also reported in Table 1 and Table 2 for wheat and mustard respectively. The tables indicate that the wheat price in Narela, Agar, Akbarpur and Alampur market are stationary and the remaining are nonstationary; whereas for mustard, except the price in Aujha, all the market price are stationary.

Table 1: Descriptive Statistics for Wheat Price

<table>
<thead>
<tr>
<th></th>
<th>N.garh</th>
<th>Narela</th>
<th>Agar</th>
<th>Aujha</th>
<th>Akbarpur</th>
<th>Alampur</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>1287.33</td>
<td>1300.906</td>
<td>1232.228</td>
<td>1185.367</td>
<td>1177.315</td>
<td>1223.633</td>
</tr>
<tr>
<td>Median</td>
<td>1230</td>
<td>1250</td>
<td>1180</td>
<td>1140</td>
<td>1120</td>
<td>1140</td>
</tr>
<tr>
<td>Maximum</td>
<td>1801</td>
<td>2900</td>
<td>2165</td>
<td>1485</td>
<td>1835</td>
<td>3551</td>
</tr>
<tr>
<td>Minimum</td>
<td>1020</td>
<td>1000</td>
<td>950</td>
<td>930</td>
<td>925</td>
<td>860</td>
</tr>
<tr>
<td>Std. Dev.</td>
<td>174.281</td>
<td>206.373</td>
<td>185.755</td>
<td>140.213</td>
<td>146.098</td>
<td>314.379</td>
</tr>
<tr>
<td>Skewness</td>
<td>0.539</td>
<td>2.079</td>
<td>0.829</td>
<td>0.497</td>
<td>0.996</td>
<td>4.336</td>
</tr>
<tr>
<td>Kurtosis</td>
<td>2.038</td>
<td>13.409</td>
<td>2.993</td>
<td>2.096</td>
<td>3.176</td>
<td>28.327</td>
</tr>
<tr>
<td>CV(%)</td>
<td>13.538</td>
<td>15.864</td>
<td>15.075</td>
<td>11.829</td>
<td>12.409</td>
<td>25.692</td>
</tr>
<tr>
<td>Jarque-Bera</td>
<td>132.592</td>
<td>7983.947</td>
<td>174.674</td>
<td>114.793</td>
<td>254.274</td>
<td>45539.890</td>
</tr>
<tr>
<td>Probability</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>Observations</td>
<td>1525</td>
<td>1525</td>
<td>1525</td>
<td>1525</td>
<td>1525</td>
<td>1525</td>
</tr>
</tbody>
</table>
Testing Stationarity

<table>
<thead>
<tr>
<th></th>
<th>ADF</th>
<th>PP</th>
</tr>
</thead>
<tbody>
<tr>
<td>Value</td>
<td>-4.303</td>
<td>-4.914</td>
</tr>
<tr>
<td>Value</td>
<td>-3.179</td>
<td>-5.119</td>
</tr>
<tr>
<td>Value</td>
<td>-1.536</td>
<td>-3.771</td>
</tr>
<tr>
<td>Value</td>
<td>-3.771</td>
<td>-3.742</td>
</tr>
</tbody>
</table>

Critical value for ADF and PP test at 5% level of significance is 2.8639

Table 2: Descriptive Statistics for Mustard Price

<table>
<thead>
<tr>
<th></th>
<th>AGRA</th>
<th>AHUJA</th>
<th>MUMBAI</th>
<th>NAJAFGARH</th>
<th>NARELA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>2629.601</td>
<td>2703.345</td>
<td>2963.522</td>
<td>2522.714</td>
<td>2532.193</td>
</tr>
<tr>
<td>Median</td>
<td>2450</td>
<td>2470</td>
<td>2900</td>
<td>2380</td>
<td>2465</td>
</tr>
<tr>
<td>Maximum</td>
<td>3980</td>
<td>3830</td>
<td>4410</td>
<td>4146</td>
<td>3962</td>
</tr>
<tr>
<td>Minimum</td>
<td>1570</td>
<td>2000</td>
<td>1800</td>
<td>1175</td>
<td>1145</td>
</tr>
<tr>
<td>Std. Dev.</td>
<td>394.442</td>
<td>542.970</td>
<td>429.951</td>
<td>490.500</td>
<td>406.479</td>
</tr>
<tr>
<td>Skewness</td>
<td>1.397</td>
<td>0.653</td>
<td>0.544</td>
<td>0.297</td>
<td>0.072</td>
</tr>
<tr>
<td>Kurtosis</td>
<td>4.724</td>
<td>2.003</td>
<td>4.528</td>
<td>3.868</td>
<td>4.201</td>
</tr>
<tr>
<td>CV(%)</td>
<td>15.000</td>
<td>20.085</td>
<td>14.508</td>
<td>19.443</td>
<td>16.052</td>
</tr>
<tr>
<td>Jarque-Bera</td>
<td>636.565</td>
<td>159.274</td>
<td>207.598</td>
<td>65.279</td>
<td>86.413</td>
</tr>
<tr>
<td>Probability</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>Observations</td>
<td>1417</td>
<td>1417</td>
<td>1417</td>
<td>1417</td>
<td>1417</td>
</tr>
</tbody>
</table>

Testing Stationarity

<table>
<thead>
<tr>
<th></th>
<th>ADF</th>
<th>PP</th>
</tr>
</thead>
<tbody>
<tr>
<td>Value</td>
<td>-3.238</td>
<td>-4.547</td>
</tr>
<tr>
<td>Value</td>
<td>-4.819</td>
<td>-6.056</td>
</tr>
<tr>
<td>Value</td>
<td>-3.584</td>
<td>-4.066</td>
</tr>
<tr>
<td>Value</td>
<td>-4.538</td>
<td>-7.547</td>
</tr>
</tbody>
</table>

Critical value for ADF and PP test at 5% level of significance is -2.864

Auto Correlation

The distributional characteristics of the series can be investigated further by analyzing the behavior of their autocorrelation functions. The results, displayed in Fig. 1 and 2, show that the autocorrelation functions of the series are decaying very slowly indicating the possible presence of long memory.
ACF Lags

Ahuja

Akbarpur

Alampur

Nai afzarh
Fig. 1: Autocorrelation Function for Wheat Price in Different Markets

- **Narela**
- **Agra**
- **Ahuja**
- **Najafgarh**
On investigating the acf and testing for presence of long memory in the price series by GPH estimator, it was found that in case of mustard, the long memory was found to be significant in Mumbai, Najafgarh and Agra; whereas in case of wheat, long memory was found in Akbarur and Alampur market. Accordingly, ARIMA (for short memory) and ARFIMA (for long memory) models were applied for modelling the price series. The squared residuals of fitted ARIMA and ARFIMA were used for testing the presence of ARCH effect by ARCH LM test and it reveals that in all the markets for both wheat and mustard there is a significance presence of conditional volatility. In most of the markets, the volatility was observed to follow long memory model. Intuitively, this volatility persistence can be appropriately modeled by a FIGARCH process because it allows for long memory behavior and slow decay of the impact of a volatility shock.

Finally, ARFIMA-FIGARCH model was applied to take care of the long memory presence in mean as well as in conditional variance simultaneously. The best model was selected based on AIC, SBC, Maximum Likelihood and $R^2$ values. The estimates of parameters along with standard error in bracket and values of AIC, SBC, Maximum Likelihood and $R^2$ are reported in Table 3 and Table 4 for mustard and wheat respectively. The estimates of conditional volatility in mustard and wheat price series in the studied markets are plotted in Fig. 3 and 4 respectively.

### Table 3: Parameter Estimates of ARFIMA-FIGARCH Models in Mustard

<table>
<thead>
<tr>
<th>Variable</th>
<th>Mumbai</th>
<th>Najafgarh</th>
<th>Narela</th>
<th>Ahuja</th>
<th>Agra</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>3075.091</td>
<td>2510.062</td>
<td>2412.158</td>
<td>2792.842</td>
<td>2373.529</td>
</tr>
<tr>
<td></td>
<td>(104.88)</td>
<td>(46.78)</td>
<td>(75.28)</td>
<td>(76.80)</td>
<td>(74.98)</td>
</tr>
<tr>
<td>AR1</td>
<td>0.596</td>
<td>0.853</td>
<td>-0.230</td>
<td>0.129</td>
<td>0.189</td>
</tr>
<tr>
<td></td>
<td>(0.089)</td>
<td>(0.138)</td>
<td>(0.041)</td>
<td>(0.097)</td>
<td>(0.053)</td>
</tr>
<tr>
<td>MA1</td>
<td>-0.289</td>
<td>-0.752</td>
<td>----</td>
<td>-0.312</td>
<td>---</td>
</tr>
<tr>
<td></td>
<td>(0.083)</td>
<td>(0.154)</td>
<td></td>
<td>(0.099)</td>
<td></td>
</tr>
<tr>
<td>d1</td>
<td>0.469</td>
<td>0.372</td>
<td>0.639</td>
<td>1</td>
<td>0.408</td>
</tr>
<tr>
<td></td>
<td>(0.055)</td>
<td>(0.038)</td>
<td>(0.048)</td>
<td></td>
<td>(0.039)</td>
</tr>
</tbody>
</table>
Table 4: Parameter estimates of ARFIMA-FIGARCH Models in Wheat

<table>
<thead>
<tr>
<th>Variable</th>
<th>Akbarpur</th>
<th>Alampur</th>
<th>Narela</th>
<th>Ahuja</th>
<th>Agra</th>
<th>Najafgarh</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>1023.581 (70.41)</td>
<td>1139.978 (17.08)</td>
<td>1126.073 (10.483)</td>
<td>949.999 (28.13)</td>
<td>1395.415 (15.82)</td>
<td>464.021 (14.75)</td>
</tr>
<tr>
<td>AR1</td>
<td>-0.583 (0.081)</td>
<td>0.542 (0.064)</td>
<td>----</td>
<td>---</td>
<td>0.780 (0.073)</td>
<td>---</td>
</tr>
<tr>
<td>MA1</td>
<td>0.273 (0.095)</td>
<td>0.521 (0.027)</td>
<td>-0.749 (0.021)</td>
<td>0.196 (0.038)</td>
<td>-0.577 (0.066)</td>
<td>-0.294 (0.044)</td>
</tr>
<tr>
<td>d1</td>
<td>0.469 (0.055)</td>
<td>0.346 (0.036)</td>
<td>1</td>
<td>1</td>
<td>0.533 (0.074)</td>
<td>1</td>
</tr>
<tr>
<td>ARCH Constant</td>
<td>939.528 (18.62)</td>
<td>6681.914 (121.23)</td>
<td>11851.866 (251.24)</td>
<td>232.541 (12.79)</td>
<td>6217.320 (123.52)</td>
<td>1172.4551 (29.32)</td>
</tr>
<tr>
<td>GARCH1</td>
<td>0.965 (0.112)</td>
<td>0.5760 (0.082)</td>
<td>0.7857 (0.118)</td>
<td>0.115 (0.043)</td>
<td>0.667 (0.140)</td>
<td>0.165 (0.044)</td>
</tr>
<tr>
<td>FIGARCH1</td>
<td>0.927 (0.305)</td>
<td>0.970 (0.303)</td>
<td>0.317 (0.028)</td>
<td>0.991 (0.401)</td>
<td>0.4657 (0.023)</td>
<td>0.652 (0.053)</td>
</tr>
<tr>
<td>d2</td>
<td>0.285 (0.021)</td>
<td>0.076 (0.010)</td>
<td>0.539 (0.026)</td>
<td>0.092 (0.040)</td>
<td>0.014 (0.096)</td>
<td>0.135 (0.032)</td>
</tr>
<tr>
<td>R² (%)</td>
<td>91.454 (89.28)</td>
<td>72.898 (86.43)</td>
<td>98.799 (86.43)</td>
<td>97.437 (97.437)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ML</td>
<td>-7491.812</td>
<td>-8067.684</td>
<td>-8563.050</td>
<td>-6109.206</td>
<td>-7898.360</td>
<td>-6897.383</td>
</tr>
</tbody>
</table>
Fig. 3: Estimates of Conditional Volatility in Mustard
Diagnostic Checking

The model verification is concerned with checking the residuals of the model to see if they contained any systematic pattern which still could be removed to improve the chosen FIGARCH Model. This has been done through examining the autocorrelations and partial autocorrelations of the residuals of various lags. For this purpose, autocorrelations of the residuals were computed and it was found that none of these autocorrelations was significantly different from zero at any reasonable level. This proved that the selected ARFIMA-FIGARCH model was an appropriate model for capturing the dual long memory and volatility present in the data under study.

Validation

One-step ahead forecasts of price for the period November 01, 2013 to December 31, 2013 in respect of above fitted model are computed. For measuring the accuracy in fitted time series model, Mean absolute error (MAE), Root Mean square error (RMSE) and Relative mean absolute prediction error (RMAPE) are computed by using the formulae given below and are reported in Table 5.

\[
\text{MAE} = 1/40 \sum_{i=1}^{40} |y_{t+i} - \hat{y}_{t+i}|
\]

\[
\text{RMSE} = \left[ 1/40 \sum_{i=1}^{40} \left( y_{t+i} - \hat{y}_{t+i} \right)^2 / y_{t+i} \right]^{1/2}
\]

\[
\text{RMAPE} = 1/40 \sum_{i=1}^{40} \left( y_{t+i} - \hat{y}_{t+i} \right) / y_{t+i} \times 100
\]
Table 5: Validation of Models

<table>
<thead>
<tr>
<th>Market for Wheat</th>
<th>Akbarpur</th>
<th>Alampur</th>
<th>Narela</th>
<th>Ahuja</th>
<th>Agra</th>
<th>najafgarh</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>RMSE</strong></td>
<td>28.683</td>
<td>212.914</td>
<td>27.206</td>
<td>16.180</td>
<td>94.385</td>
<td>34.969</td>
</tr>
<tr>
<td><strong>RMAPE (%)</strong></td>
<td>3.62</td>
<td>5.21</td>
<td>3.16</td>
<td>3.06</td>
<td>4.01</td>
<td>2.68</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Market for Mustard</th>
<th>Mumbai</th>
<th>Najafgarh</th>
<th>Narela</th>
<th>Ahuja</th>
<th>Agra</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>MAE</strong></td>
<td>46.939</td>
<td>35.967</td>
<td>85.705</td>
<td>84.909</td>
<td>37.878</td>
</tr>
<tr>
<td><strong>RMSE</strong></td>
<td>50.728</td>
<td>29.950</td>
<td>122.388</td>
<td>107.809</td>
<td>38.846</td>
</tr>
<tr>
<td><strong>RMAPE (%)</strong></td>
<td>4.26</td>
<td>3.41</td>
<td>5.16</td>
<td>5.19</td>
<td>3.94</td>
</tr>
</tbody>
</table>

4. Conclusion

The present investigation is aimed to test the relevance of long memory in modeling and forecasting volatility for the spot prices of mustard and wheat in different markets of India. GPH test indicated the existence of long-term memory in the mean and volatility processes. On the basis of minimum AIC and BIC values, the best ARFIMA-FIGARCH model was selected. We find that long memory is particularly strong and plays a dominant role in explaining the spot price of wheat and mustard. Finally, our out-of-sample analysis indicates that the FIGARCH-based model performs satisfactorily in terms of RMSE, MAE and RMAPE.

References


