Indian Journal of Agricultural Sciences **88** (2): 268–72, February 2018/Article

Statistical modelling for forecasting volatility in potato prices using ARFIMA-FIGARCH model

DIPANKAR MITRA[1], RANJIT KUMAR PAUL[2] and A K PAUL[3]

ICAR-Indian Agricultural Statistics Research Institute, New Delhi 110 012

Received: 21 June 2017 ; Accepted: 05 October 2017

ABSTRACT

This paper investigates the presence of long memory both in mean and volatility in the potato prices in Agra and Amritsar markets of India, using the Autoregressive fractionally integrated moving average (ARFIMA) and Fractionally integrated generalized autoregressive conditional heteroscedastic (FIGARCH) models. Long memory tests are carried out both for the returns and squared return series. The results of GPH estimator indicate the existence of long memory in the price data.The ARFIMA model with error following FIGARCH process is fitted to return prices of potato for each of the two markets. At the end, the forecasting performance of fitted ARFIMA-FIGARCH models are carried out in terms of RMAPE and RMSE and the residuals are also examined to check adequacy of the fitted models.

Key words: ARFIMA, FIGARCH, GPH, Long memory, Potato,Volatility

In econometric study, analysis and modeling of several applied time-series is very important to capture their salient features. During the analysis of time-series it may happen to choose the models from different classes. In recent years modeling and forecasting of returns and volatility concerning the long run persistence or long memory is an emerging area of scientific research. The presence of long memory in the returns and volatility implies that there is a strong association between the observations widely separated in time. Granger and Joyeux (1980) and Hosking (1981) showed that fractionally integrated series could produce long memory property, and proposed the Autoregressive fractionally integrated moving average (ARFIMA) model. Paul *et al.* (2014) have applied ARFIMA for modelling and forecasting of daily retail price of pigeon pea (*Cajanas cajan*) in Karnal, Haryana. The ARFIMA model is based on the assumptions of linearity, stationarity and homoscedasticity of error variance. Under these assumptions it is quite impossible to deal with series exhibiting high volatility or periods of instability such as agricultural commodity price series. Many economic series show periods of stability followed by the periods of instability in volatility, to take care this Autoregressive conditional heteroscedastic (ARCH) model was developed.

But the feature of ARCH to give satisfactory forecast only with large number of parameters has necessitated the emergence of more parsimonious model that is Generalized ARCH (GARCH) model (Bollerslev 1986). The main limitation of GARCH model is that it cannot handle the long memory present in the volatility. For modeling and forecasting of agricultural commodities prices with long memory in volatility, Fractionally integrated GARCH (FIGARCH) model (Baillie 1996) is generally used. Paul *et al*. (2015a) studied the effect of long range dependence in modelling and forecasting volatility for the spot prices of mustard and wheat in different markets of India. Paul *et al*. (2016a) have studied long memory in conditional variance and applied FIGARCH model to forecast the volatility of spot price of gram in Delhi market, India.

Potato (*Solanum tuberosum* L.) is the 3rd most important food crop after rice and wheat in the world. It is consumed by more than 1 billion people worldwide. Two emerging Asian economies, viz. India and China together contribute almost 1/3rd of world's production. Keeping in mind the potential features of potato in diet Food and Agricultural Organization (FAO) has declared potato as "food for future". In India potato is cultivated in approximately 18-19 lakh ha it accounts nearly 26% of total vegetable production. Potato is generally shown during the month of October and November and harvested in month of January to February. Major potato producing states of India are- Uttar Pradesh, West Bengal, Punjab, Gujarat, Madhya Pradesh etc. For the present investigation, Agra market of Uttar Pradesh and Amritsar market of Punjab here considered. Consumer prices of essential commodities used in daily lifeare a very sensitive issue. The consumer suffers in case prices of consumable items are high. On the other hand, producer suffers when prices are too low to recover the cost of production. The price of consumable items is determined by the demand and supply of that particular commodity. The

[1]Ph D Scholar (e mail:dipankar28mitra@gmail.com), [2]Scientist (e mail: ranjit.paul@icar.gov.in, ranjitstat@gmail.com), ICAR-IASRI, [3]Principal Scientist (e mail: pal@iasri.res.in)

supply of agricultural products depends mainly on nature so there is a great element of uncertainty. If supply is in short of demand then price will increase. The other main factor influencing the market arrival or supply of potato is the amount stored in coldstorage. Such large fluctuation in the potato prices is a matter of concern. So there is need to study the price volatility of potato to help the producers as well as consumers. In the present study, an attempt has been made to model the long range persistent structure of volatile price of potato of different markets in India using FIGARCH model. The ARFIMA-FIGARCH model provides a useful way of explaining the relationship between the conditional mean and conditional variance of a process exhibiting the long memory property.

MATERIALS AND METHODS

Long memory process

Let $\{X_t\}$; (t = 0, 1, 2,....) be a time-series process with the autocorrelation function for any integer lag k

$$\rho_k = cov(x_t, x_{t-1})/var(x_t) \tag{1}$$

For a stationary time-series process it is expected to have $\lim_{k\to\infty} \rho_k = 0$. Most of the well-known class of stationary and invertible (ARMA) time-series process have autocorrelations that decay at an exponential rate, so that $\rho_k \approx |m|^k$, where $|m| < 1$.

For any stationary long-memory process the autocorrelation function decays at an hyperbolic rate having the form $\rho_k \approx Ck^{2d-1}$, as k increases indefinitely, where c is a nonzero positive constant and is the long-memory parameter controlling the speed of decline of the autocorrelation.

The ARFIMA model

A time-series process $\{X_t\}$; (t = 0,1,2,...) is called a fractionally differenced or I(d) process

$$(1-L)^d x_t = \in_t,\ t = 1, 2 \tag{2}$$

where L denotes the lag operator and d is the fractional difference parameter and $\in_t$ is white noise with $E(\in_t) = 0$, $E(\in_t^2) = \sigma^2$ and $E(\in_t \in_s) = 0\ \forall\ t \neq s$. The process $\{X_t\}$ is stationary and long memory process if $d \in (1/2,0)$ and called nonstationary long memory process if $d \in (1/2,0)$.

A more flexible parametric model called ARFIMA (p, d, q) which incorporates both long and short memory process is given as

$$(1-L)^{d} \varphi(L) y_t = \theta(L) \in_t, \in_t \pitchfork i.i.d\ (0, \sigma^2) \tag{3}$$

where $\varphi(L) = 1 - \varphi_1 L - \varphi_2 L^2 - \cdots - \varphi_2 L^p$ and $\theta(L) = 1 - \theta_1 L - \theta_2 L^2 - \cdots - \theta_q L^q$ represent the AR and MA polynomial of order p and q respectively. $(1 - L)^d$ is the fractional difference operator (filter) which is essentially is the binomial expansion as

$$(1-L)^d = 1 - dL + \frac{d(d-1)}{2!}L^2 + \frac{d(d-1)(d-2)}{3!}L^3 - \cdots = \sum_{k=0}^{\infty} \frac{\Phi(k-d)L^k}{\Phi(-d)\Phi(k+1)} \tag{4}$$

where Φ (.) represents the gamma function.

The long memory parameter d in ARFIMA model can be estimated by two methods under the maximum likelihood estimation (MLE) technique, namely Exact maximum likelihood estimation (EMLE) and Approximate maximum likelihood estimation (AMLE). The main feature of MLE is that it estimates the short memory parameters (p and d) and long memory parameter d simultaneously to fit an ARFIMA (p, d, q) model.

The GARCH model

The process $\{\in_t\}$ is ARCH(q) if the conditional distribution of $\{\in_t\}$ given the available ψ_{t-1} information is

$$\in_t | \psi_{t-1} \sim N(0, h_t) \text{ and } \in_t = h_t^{1/2} \xi_t \tag{5}$$

where is $\{\xi_t\}$ a white noise process i.e., $\xi_t \pitchfork iid(0,1)$.

In Generalized ARCH (GARCH) model, the conditional variance is a linear function of its own lags as well as past squared residuals. The GARCH (p,q) process has the following form provided $a_0 > 0$, $a_i \geq 0\ \forall$ i; bj $\geq 0\ \forall\ j$

$$h_t = a_0 + \sum_{i=1}^{q} a_i\, \varepsilon_{t-1}^2 + \sum_{j=1}^{p} a_j\, h_{t-1} \tag{6}$$

Paul *et al.* (2016b) have considered volatility in prices of onion for three markets of Delhi, viz. Azadpur, Keshopur and Shahdara using different foresting techniques.

Testing for ARCH effects

Checking the possible presence of the conditional heteroscedasticity in the squared residual series $\{e_t^2\}$ is known as the test for ARCH effects. There are two tests available in the literature for testing the possible presence of ARCH effect. The first one is the usual Ljung-Box statistic where the hypothesis is to test the absence of autocorrelation in the $\{e_t^2\}$ series. The second one is the Lagrange multiplier test by Engle (1982) which is equivalent to usual F-test for testing $H_0 : a_i = 0$, i = 1, 2,..., q in the linear regression

$$e_t^2 = a_0 + a_1 e_{t-1}^2 + \ldots + a_q e_{t-q}^2 + u_t, \quad t = 1, 2, \ldots, T \tag{7}$$

where is the error term, is the prespecified nonzero positive integer and is the sample size.

The FIGARCH model

Engle and Bollerslev (1986) considered a particular class of GARCH models known as integrated GARCH (IGARCH) models whose unconditional variance does not exist. This occurs when $\sum_{i=1}^{q} a_i + \sum_{j=1}^{p} b_j = 1$

The IGARCH model implies infinite persistence of the conditional vari (p, q) a shock in squared returns. An integrated GARCH process can be written as

$$[1 - a(L) - b(L)]\ (1 - L)\ \varepsilon_1^2 = a_0 + [1 - b(L)]\ \gamma_t$$

The fractionally integrated GARCH or FIGARCH class of models (Baillie *et al*, 1996) is obtained by replacing the first difference operator with the fractional differencing operator $(1 - L)^d$, where d is a fraction $0 < d < 1$.The FIGARCH class of models can be obtained by considering

$$[1 - a(L) - b(L)]\,(1 - L)^{d}\,\varepsilon_1^2 = a_0 + [1 - b(L)]\,\gamma_t \quad (9)$$

where $\gamma_t = \varepsilon_1^2 - h_t$.

This model includes a more flexible class of process for the conditional variance in explaining and representing the observed temporal dependencies of financial market volatility in a better way the other types of GARCH model. The parameters of an FIGARCH (p, d, q) model are estimated by using the quasi-maximum likelihood (QML) method. Details of FIGARCH model can be found in Tayafi and Ramanathan (2012) and Paul *et al*. (2015b).

RESULTS AND DISCUSSION

Data description and descriptive statistics

For the present study daily time-series data regarding tomodal wholesale prices of potato for Agra, Amritsar markets for the period 1 January 2010 to 13 May, 2017, are collected from National Horticulture Research and Development Foundation (NHRDF) (http://nhrdf.org/en-us/). The data series is divided into two parts: training set and testing set (holdout set). The training data set (1 January, 2010 to 16 February, 2017) is used for parameter estimation and the last 60 observations i.e. from 17 February 2017 to 13 May, 2017 considered as testing set is used for model validation purpose and also for obtaining out-of-sample forecast.

The descriptive statistics of potato price for two markets are reported in Table 1. A perusal of Table 1 indicates that average potato price is higher in Agra market than Amritsar. The variability is more in the Amritsar market than Agra in terms of coefficient of variation. Higher value of CV is an indication of period of instability or volatility in the original data.The series under consideration are positively skewed and leptokurtic.

Test for stationarity

Augmented Dickey Fuller (ADF), Phillips-Perron (PP) and Kwiatkowski–Phillips–Schmidt–Shin (KPSS) tests have been applied to see the presence of non-seasonal unit root in the return series and the results are found significant at 5% level of significance indicating stationarity of the return series.

Table 1 Descriptive statistics of potato price in different markets

Statistics	Agra	Amritsar
Observations	1909	1441
Mean (₹/quintal)	756.78	639.96
Median	615.00	550.00
Minimum	260.00	140.00
Maximum	2190.00	1700.00
Standard deviation	395.16	362.09
Coefficient of variation (%)	48.25	56.58
Skewness	1.28	0.85
Kurtosis	4.38	3.10

Auto correlation function of the series

The autocorrelation function (ACF) and partial autocorrelation (PACF) function of the return and squared return seriesare studied to investigate the distributional characteristic of the series. The ACF and PACF plots of the return series and squared return series for both the markets are shown in Fig 1. Since the autocorrelation functions and partial autocorrelation functions are significantat distant lags (even after 800 lags), there is a clear indication of long memory in the both return and squared return series. The dotted lines in the ACF and PACF plots represent the 95% critical values of the test statistic.

Testing for long memory

After visualizing the ACF and PACF of the return and squared return series, long memory tests are applied using GPH estimator in order to test the presence of long memory and the results are provided in Table 2. A perusal of Table 2 indicates that both the return and squared return series of both the markets have significant long memory pattern (as calculated Z statistic is greater than the critical value at 5% level of significance, i.e. 1.96). Accordingly, ARFIMA model is fitted to the return series. The best ARFIMA is selected based on minimum AIC and SBC values. Residuals computed from the best selected model are examined for

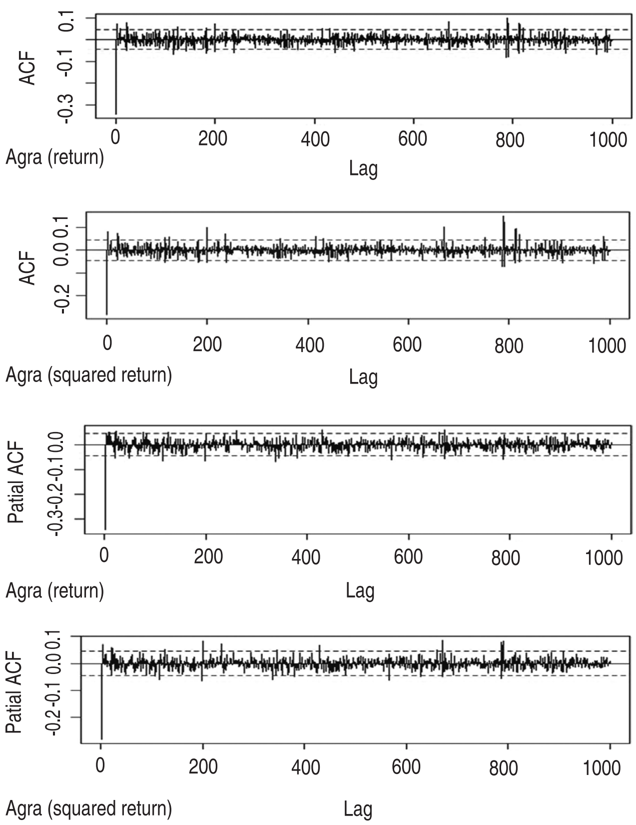

Fig 1 ACF and PACF plots of return and squared return series

Table 2 Results of long memory test for return and squared return series using GPH estimator

	Agra		Amritsar	
	Return	Squared returns	Return	Squared returns
d	0.096	0.091	0.048	0.046
SE	0.047	0.039	0.023	0.019
Z	2.043	2.333	2.087	2.421

possible presence of conditional heteroscedasticity. It is found that squared residuals are autocorrelated up to a long lags. Furthermore, the ARCH-LM testis carried out on the squared residual series and the tests reveals that in both the markets there is a significant presence of conditional volatility. As previously reported that squared return series is also having long memory, therefore, the persistency in the volatility can well be captured by fitting a FIGARCH model to the return series.

Fitting of ARFIMA-FIGARCH model

ARFIMA-FIGARCH model is fitted to the return series and the estimates of parameter along with standard error in the bracket are reported in Table 3. The best model is selected on the basis of two criteria, namely AIC and SBC values. From Table 3 it is seen that the long memory parameters d_1 and d_2 for both the markets are significant at least at 5% level of significance. Since the long memory parameters of return and squared return series for both the markets lie between 0 to 0.5, there is significant presence of stationary long memory both in mean and volatility.

Residual diagnosis

The model is verified by plotting the autocorrelations at various lags of the residuals obtained from the fitted ARFIMA-FIGARCH model. The ACF plots of residuals for the markets are given in the fig. 2. Since almost all the autocorrelations are insignificant, it has been proved that the selected ARFIMA-FIGARCH model was an appropriate

Table 3 Parameter estimates of fitted ARFIMA-FIGARCH models

Series	Agra	Amritsar
Mean equation		
Constant	-0.001 (0.002)	-0.003 (0.004)
AR1	-0.304** (0.036)	0.224** (0.071)
MA1	-	-0.728* (0.051)
d_1	0.071** (0.025)	0.151* (0.065)
Variance equation		
Constant	0.004** (<0.001)	0.201** (0.071)
GARCH	0.168**(0.068)	0.608** (0.066)
FIGARCH	0.409** (0.058)	0.245** (0.068)
d_2	0.013* (0.006)	0.617** (0.047)
AIC	-2.871	-1.166
BIC	-2.802	-1.137

**Represents significance at 1% level of significance level and * represents significance at 5% level of significance level.

Table 4 RMAPE and RMSE for each forecast horizons of potato price for different markets

Series	Forecast horizon (*h*)						
	10	20	30	40	50	60	Average
				RMAPE			
Agra	12.52	8.07	7.67	6.56	6.02	5.67	7.75
Amritsar	0.83	3.02	2.32	2.03	1.88	3.86	2.32
				RMSE			
Agra	58.48	44.09	39.14	34.67	32.07	30.09	39.76
Amritsar	2.51	13.02	10.66	9.27	8.34	20.87	10.78

model for capturing dual long memory and volatility present in the price data.

Validation of models

One-step ahead forecast of price for the period 17 February 2017 to 13 May 2017 using the selected ARFIMA-FIGARCH model is calculated. The accuracy of the fitted models is measured in terms of relative mean absolute percentage error (RMAPE) and root mean square error (RMSE) for last sixty observations (i.e. for h = 60) using the following formulae.

$$\text{RMAPE} = \frac{1}{h}\sum_{t=1}^{n} | y_t - y | / y_t \times 100;$$

$$\text{RMSE} = \sqrt{\frac{1}{2}\sum_{t=1}^{n} (y_t - y_t)^2}$$

Table 5 Forecast for potato price in different markets using the fitted ARFIMA-FIGARCH models

Forecast horizon (h)	Agra		Amritsar	
	Actual value	Fitted value	Actual value	Fitted value
1	430	378.19	150	148.77
2	450	418.50	150	148.79
3	440	448.62	150	148.81
4	460	445.33	150	148.83
5	470	456.72	150	148.85
6	450	470.01	150	148.86
7	460	456.74	200	148.88
8	455	458.16	200	179.38
9	450	457.38	200	186.03
10	460	451.86	200	190.74
11	440	458.07	200	193.93
12	450	445.60	200	196.06
13	300	447.36	200	197.48
14	360	330.22	200	198.42
15	350	337.86	200	199.02
16	390	351.72	200	199.40
17	410	378.85	200	199.63
18	320	406.50	200	199.76
19	310	340.37	200	199.82
20	325	308.79	200	199.84

The computed RMAPE and RMSE values for each of the forecast horizons are listed in the Table 4 for all the price data respectively. Forecast evaluation is carried out for six moving windows (10-step, 20-step, 30-step, 40-step, 50-step, 60-stepahead). For each market, the final columns of Table 4 labeled as “Average” show average RMAPE and RMSE, across all the forecast horizons for all the markets respectively. Along with the actual value for last twenty observations the fitted value are listed in the Table 5 for the markets considered for the present study.

In the present study the relevance of capturing long-range dependence pattern in modeling and forecasting of potato price in India is investigated. Significant results of GPH test indicate the presence of long memory in mean and volatility as well. After fitting ARFIMA model to the return series the residuals are obtained from the fitted mean model and the squared residuals are tested for possible presence of ARCH effect. It is found that the ARCH effect is significant. Accordingly, ARFIMA-FIGARCH model is fitted and the best model is selected on the basis of minimum AIC and SBC value. The lesser values of RMAPE ensure good predictability of the fitted models. Finally forecasts are calculated for last twenty data points using the best fitted model and it has been observed that the forecasts are very closer to the actual values indicating better fitness of the model.

REFERENCES

Baillie R T. 1996. Long memory processes and fractional integration in econometrics. *Journal of Econometrics* **73:** 5–9.

Bolerslev T. 1986. Generalized autoregressive conditional heteroscedasticity. *Journal of Econometrics* **31:** 307–27.

Engle R F. 1982. Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica* **50**: 987–1007.

Engle R F and Bollerslev T. 1986. Modelling the persistence of conditional variances. *Econometric Reviews* **5**: 1–50.

Granger C W J and Joyeux R. 1980. An introduction to long-memory time series models and fractional differencing. *Journal of Time Series Analysis* **4**: 221–38.

Paul R K, Gurung B and Paul A K. 2014. Modelling and forecasting of retail price of arhar dal in Karnal, Haryana. *Indian Journal of Agricultural Sciences* **85(1)**: 69–72.

Paul R K, Gurung B and Samanta S. 2015a. Analyzing the effect of dual long memory process in forecasting agricultural prices in different markets of India. *International Journal of Empirical Finance* **4(4)**: 235–49.

Paul R K, Gurung B, Samanta S and Paul A K. 2015b. Modeling long memory in volatility for spot price of lentil with multi-step ahead out-of-sample forecast using AR-FIGARCH model. *Economics Affairs* **60(3)**: 457–66.

Paul R K, Gurung B, Paul A K and Samanta, S. 2016a. Long memory in conditional variance. *Journal of Indian Society of Agricultural Statistics* **70(3)**: 243–54.

Paul R K, Rana S and Saxena R. 2016b. Effectiveness of price forecasting techniques for capturing asymmetric volatility for onion in selected markets of Delhi. *Indian Journal of Agricultural Sciences* **86(3)**: 303–9.

Tayafi M and Ramanathan T V. 2012. An overview of FIGARCH and related time-series models. *Australian Journal of Statistics* **141:** 175–96.