Indian Journal of Agricultural Sciences **85** (1): 69–72, January 2015/Article

Modelling and forecasting of retail price of arhar dal in Karnal, Haryana

RANJIT KUMAR PAUL[1], BISHAL GURUNG[2] and A K PAUL[3]

Indian Agricultural Statistics Research Institute, New Delhi 110 012

Received: 7 October 2013; Revised accepted: 31 October 2014

ABSTRACT

Forecasting retail price of agricultural commodities is of utmost importance for planning in advance to resist any abnormalities. In this regard, the autoregressive fractionally integrated moving-average (ARFIMA) model is employed. ARFIMA model searches for a non-integer differencing parameter *d* to difference the data to capture long memory. The model is applied for modelling and forecasting of daily retail price of pigeonpea (*Cajanas cajan*) in Karnal during January, 2011 to July, 2013. Augmented Dickey-Fuller (ADF) test and Philips Peron (PP) test are used for testing the stationarity of the series. Autocorrelation (ACF) and partial autocorrelation (PACF) functions showed a slow hyperbolic decay indicating the presence of long memory. In the present price series, long memory parameter is found to be significant. On the basis of minimum AIC values, the best model is identified. To this end, evaluation of forecasting is carried out with root mean squares prediction error (RMSPE), mean absolute prediction error (MAPE) and relative mean absolute prediction error (RMAPE). The residuals of the fitted models were used for diagnostic checking. It is found that ARFIMA model has been able to capture the long memory present in the data set. The R software package has been used for data analysis.

Key words: ADF test, ARFIMA model, Long Memory, PP test, Stationarity

Retail forecasting methods anticipate the future purchasing actions of consumers by evaluating past revenue and consumer behaviour over the previous months or year to discern patterns and develop forecasts for the upcoming months. Forecasting helps the retailer to meet the demands of the customer by understanding consumer purchase patterns better. Accurate forecasts that meet the forthcoming consumption demands of customers help retail business owners and management to maximize and extend profits over the long term. Forecasting permits price adjustments to correspond with the current level of consumer spending patterns. Therefore, forecasting retail price of agricultural commodities is of utmost importance by using advanced statistical techniques. Long-memory models have become increasingly popular as a tool to describe economic time series. Useful entry points to the literature are the surveys by Robinson (1995) and Baillie *et al*. (1996), who considers the developments in the econometric modelling of long memory, and Beran (1995), who reviews long-memory modelling in other areas, viz. economics, finance etc. The monograph of Beran (1994) discusses most of the central issues, including forecasting. Although most economic time series are nonstationary and do require differencing of some kind, it is not necessarily true that taking first differences and then using an ARMA model will be the best remedy. In Box-Jenkins analysis, it is assumed that if the series is nonstationary, it is hoped that differenced series will have rapidly decaying autocorrelations and be free of trend-like behaviour, so that it can be well described by a stationary invertible ARMA model. But this is not always the case. The autoregressive fractionally integrated moving-average (ARFIMA) model searches for a non-integer parameter, *d*, to difference the data to capture long memory. The existence of non-zero *d* is an indication of long memory and its departure from zero measures the strength of long memory. Long memory is also called fractal structure because of noninteger *d*. The long memory study on agricultural futures markets is at the beginning. The empirical work of Helms *et al*. (1984) is an isolated study, which analyzed the short series (about 230 observations) of one commodity (the soybean complex) using only the classical R/S techniques. So there is need to investigate the long memory behaviour in the agricultural price. In the present investigation, an attempt is made to investigate the structure of long memory in daily retail price of tur or arhar dal in Karnal, India. The present study will take advantage of the new developments in statistical methods to analyze the time series data of retail price of tur or arhar dal in Karnal, India for modelling and forecasting purposes.

MATERIALS AND METHODS

Long memory in time-series can be defined as autocorrelation at long lags (Robinson 1995). According to Jin and Frechette (2004), memory means that observations

[1]Scientist (e mail: ranjitstat@gmail.com), [2]Scientist (e mail: vsalrayan@gmail.com), [2]Senior Scientist (e mail: pal@iasri.res.in), Division of Statistical Genetics

are not independent (each observation is affected by the events that preceded it). The acf of a time-series y_t is defined as

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

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.

Fractional integration is the primary conceptual framework for describing long memory in financial time-series. Fractional integration is a generalization of integer integration, under which time-series are usually presumed to be integrated of order zero or one. For example, an autoregressive moving-average process integrated of order d [denoted by ARFIMA(p, d, q)] can be represented as

$$(1 - L)^d\ \varphi(L)y_t = \theta(L)u_t$$

where u_t is an independently and identically distributed (i.i.d.) random variable with zero mean and constant variance, L denotes the lag operator; and $\varphi(L)$ and $\theta(L)$ denote finite polynomials in the lag operator with roots outside the unit circle. For $d = 0$, the process is stationary, and the effect of a shock to $u(t)$ on $y(t + j)$ decays geometrically as j increases. For $d = 1$, the process is said to have a unit root, and the effect of a shock to $u(t)$ on $y(t + j)$ persists into the infinite future. In contrast, fractional integration defines the function $(1 - L)^{-d}$ for noninteger values of the fractional differencing parameter d. It turns out that for $-0.5< d< 0.5$ the process $y(t)$ is stationary and invertible. A detail description of ARFIMA model can be found in Robinson (2003).

We deal with some well known estimation methods of the long memory parameter d. The first one is the semiparametric method based on an approximated regression equation obtained from the logarithm of the spectral density function of a model. This method is proposed by Geweke and Porter-Hudak (1983). The second is the Gaussian semiparametric method developed by Robinson (1995).

The unit root test described by Dickey and Fuller (1979) is valid if the time series y_t is well characterized by an AR(1) with white noise errors. Many financial time series, however, have a more complicated dynamic structure than is captured by a simple AR(1) model. Said and Dickey (1984) augment the basic autoregressive unit root test to accommodate general ARMA(p, q) models with unknown orders and their test is referred to as the augmented Dickey-Fuller (ADF) test. The ADF test tests the null hypothesis that a time series y_t is I(1) against the alternative that it is I(0), assuming that the dynamics in the data have an ARMA structure. The ADF test is based on estimating the test regression

$$\Delta y_t = \beta' D_t + \pi y_{t-1} + \sum_{j=1}^{p} \psi_j \Delta y_{t-j} + \varepsilon_t$$

where D_t is a vector of deterministic terms (constant, trend etc.). The p lagged difference terms, Δy_{t-j}, are used to approximate the ARMA structure of the errors, and the value of p is set so that the error ε_t is serially uncorrelated. The error term is also assumed to be homoskedastic. Under the null hypothesis, Δy_t is I(0) which implies that $\pi = 0$. The ADF t-statistic is then the usual t-statistic for testing $\pi = 0$.

Phillips and Perron (1988) developed a number of unit root tests that have become popular in the analysis of financial time series. The Phillips-Perron (PP) unit root tests differ from the ADF tests mainly in how they deal with serial correlation and heteroscedasticity in the errors. In particular, where the ADF tests use a parametric autoregression to approximate the ARMA structure of the errors in the test regression, the PP tests ignore any serial correlation in the test regression. The test regression for the PP tests is

$$\Delta y_t = \beta' D_t + \pi y_{t-1} + u_t$$

where u_t is I(0) and may be heteroskedastic. The PP tests correct for any serial correlation and heteroscedasticity in the errors ut of the test regression by directly modifying the test statistics. Under the null hypothesis that $\pi = 0$, the PP statistics have the same asymptotic distributions as the ADF t-statistic. One advantage of the PP tests over the ADF tests is that the PP tests are robust to general forms of heteroscedasticity in the error term u_t. Another advantage is that the user does not have to specify a lag length for the test regression.

Steps for fitting of ARFIMA model are: (i) Estimate d using periodogram based estimate. (ii) If d is significantly different from zero, difference the series. 3. (iii) Find estimates for the order of p and q (max=2) possibly on the differenced data 4. (iv) Fit an ARFIMA(p,d,q) model on the undifferenced data using the order found in step 3. (v) If d in step 2 was significantly lower than 0 ARFIMA is fitted with a drange=c(-1,0) otherwise drange=c(0,1). (vi) The best Fitting model is selected by selecting lowest AIC.

RESULTS AND DISCUSSION

Data set

For the present investigation, the daily retail price of tur or arhar dal in Karnal, India, during the period 1 January, 2011 to 31 July, 2013 is used. The data is collected form Ministry of Consumer's Affairs, Government of India. The time series plot of above dataset has been exhibited in Fig 1.

A perusal of the Fig 1 indicates that the dataset is stationary. In order to test for stationarity, two tests namely Augmented Dickey-Fuller unit root test (Said and Dickey, 1984) and Philips-Peron unit root test (Philips and Peron, 1988) are conducted. The results of the tests are given in Table 1. A perusal of Table 1 reveals that n all the three lags both the test statistics rejects the null hypothesis of presence

Table 1 Stationarity testing

Lag	ADF Test Statistic		PP Test Statistic	
	Test statistic	Probability	Test statistic	Probability
0	–3.29	0.0161	–3.29	0.0161
1	–3.33	0.0144	–3.32	0.0149
2	–3.29	0.0164	–3.32	0.0150

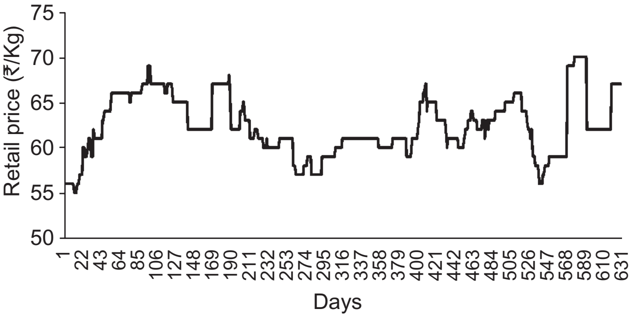

Fig 1 Time series plot of retail prices of tur in Karnal

of unit root indicating that series is stationary.

Structure of autocorrelations

For a linear time series model, typically an autoregressive integrated moving average (ARIMA(p,d,q)) process, the patterns of autocorrelations and partial autocorrelations could indicate the plausible structure of the model. At the same time, this kind of information is also very important for modelling nonlinear dynamics. The basic property of a long memory process is that the dependence between the two distant observations is still visible. For the present daily retail price series, 150 autocorrelations were estimated. The autocorrelation functions of above series are plotted in Fig 2. A perusal of Fig 2 indicates that, the autocorrelations do not decay exponentially over time span, rather, there is hyperbolic decay of the autocorrelations functions towards zero and they show no clear periodic patterns. There is no evidence that the magnitude of autocorrelations become small as the time lag, *j*, becomes large. No seasonal and other periodic cycles were observed.

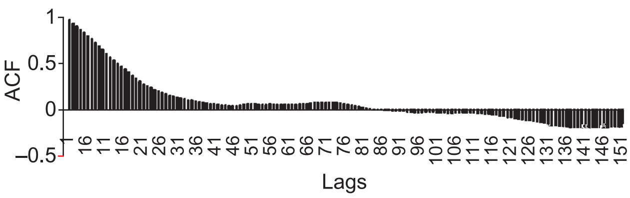

Fig 2 Correlogram of time series data of retail prices of Tur in Karnal

The most common method for estimating the fractional integration parameter *d* is the ARFIMA time series method (Robinson 2003). We estimate different ARFIMA specifications, as described previously. Based on the smallest AIC value, the best ARFIMA model was selected. AIC values and Log likelihood are reported in Table 2. Estimates for the selected ARFIMA models are provided in Table 3, along with t-statistics.

Table 2 Log likelihood and AIC values of different ARFIMA models

Models	Log -likelihood	AIC
ARFIMA(1,d,1)	–765.8	1539.68
ARFIMA(1,d,0)	–765.6	1537.764
ARFIMA(0,d,1)	–848.4	1702.771
ARFIMA(2,d,1)	–766.0	1542.013
ARFIMA(2,d,0)	–766.3	1540.695
ARFIMA(2,d,2)	–765.8	1543.688

Table 3 Parameter Estimates of ARFIMA Model

Market	Parameters	Estimate	S. E.	Z-Value	Probability
Amritsar	d	0.030	0.008	3.407	< 0.001
	AR1	0.961	0.011	88.220	< 0.001

These estimates indicate evidence of long memory (high degree of predictability) in the study price series with $0 < d < 0.5$. Positive values of the fractional differencing parameter indicate a short of long-memory known as persistence. Persistence is characterised by positive autocorrelations, and exhibit low variance at low frequencies. Note that, when *d* parameter is positive and significant, then the series may have infinite conditional variance.

The results show that the value of long memory parameter *d* is 0.03 and is highly significant indicating the presence of long memory in daily retail price of tur in Karnal. Hence, empirical evidence shows that the lag length increases the autocorrelations decay hyperbolically to zero.

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 ARFIMA, which has been done through examining the autocorrelations and partial autocorrelations of the residuals of various orders. 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 model was an appropriate model for forecasting the data under study.

Validation

In the present study, the data set during 1 January,

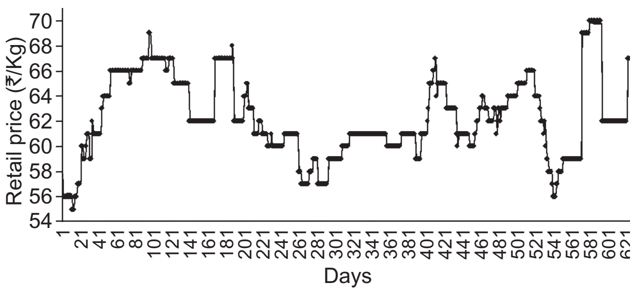

Fig 3 Observed vs fitted plot of retail prices of Tur in Karnal

Table 4 Validation of models

MAPE	RMSPE	RMAPE (%)
0.798	0.891	1.23

2011 to 3 July, 2013 have been used for model building and remaining 20 data points, i e 04-31 July, 2013 (working days) for validation purpose. One-step ahead forecasts of retail price along with their corresponding standard errors, upper confidence interval and lower confidence interval for the period 04-31 July, 2013 in respect of above fitted model are computed.

For measuring the accuracy in fitted time series model, Mean absolute error (MAE), Mean absolute percentage error (MAPE) and Relative mean absolute prediction error (RMAPE) are computed by using the formulae given below and are reported in Table 4.

$$\text{MAE} = 1/20\sum_{i=1}^{20}\left|y_{t+i} - \hat{y}_{t+i}\right|$$

$$\text{MAPE} = 1/20\sum_{i=1}^{20}\left|y_{t+i} - \hat{y}_{t+i}\right|$$

$$\text{RMAPE} = 1/20\sum_{i=1}^{20}\left\{\left|y_{t+i} - \hat{y}_{t+i}\right| / y_{t+i}\right\}\times 100$$

A perusal of above table indicates that in all the price series data, RMAPE is less than 2% indicating the accuracy of the models. The observed (dots) and fitted graph (line) of the data is depicted in Fig 3.

CONCLUSION

Long memory systems are characterised by their ability to remember events in the long history of time series data and their ability to make decisions on the basis of such memories. The price of an asset (or commodity) determined in an efficient market should follow a martingale process in which each price change is unaffected by its predecessor and has no memory. So, if the price series exhibit long memory (or long-term dependence), they display significant autocorrelation between distant observations. Therefore, the series realisations are not dependent over time, thus violating the market efficiency hypothesis. The model demonstrated a good performance in terms of explained variability and predicting power. The findings of the present study provided direct support for the potential use of accurate forecasts in decision making for the retailers as well as consumers.

REFERENCES

Baillie R T, Bollerslev T and Mikkelsen HO. 1996. Fractionally integrated generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics* **74**: 3–30.

Beran J. 1994. Maximum likelihood estimation of the differencing parameter for invertible short and long memory autoregressive integrated moving average models. *Journal of the Royal Statistical Society* **B 57** (4): 659–72.

Beran J. 1995. *Statistics for Long-Memory Processes.* Chapman and Hall Publishing Inc. New York.

Dickey D and Fuller W. 1979. Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American Statistical Association* **74**: 427–31.

Geweke J and Porter-Hudak S. 1983. The estimation and application of long-memory time-series models. *Journal of Time series Analysis* **4**: 221–38.

Helms B P, Kaen F R and Rosenman R E. 1984. Memory in commodity futures contracts. *Journal of Futures Markets* **10**: 559–67.

Jin H J and Frechette D. 2004. Fractional integration in agricultural futures price volatilities. *American Journal of Agricultural Economics* **86**: 432–43.

Phillips P C B and Perron P. 1988. Testing for unit roots in time series regression. *Biometrika* **75**: 335–46.

Robinson P M. 1995. Log-periodogram regression of time-series with long-range dependence. *Annals of Statistics* **23**: 1 048–72.

Robinson P M. 2003. *Time series with long memory*. Oxford university Press, Oxford.

Said S E and Dickey D. 1984. Testing for unit roots in autoregressive moving-average models with unknown order. *Biometrika* **71**: 599–607.