Available online at www.isas.org.in/jisas
JOURNAL OF THE INDIAN SOCIETY OF
AGRICULTURAL STATISTICS 75(3) 2021 187–192

An Improved Cointegration based Time Delay Neural Network Model for Price Forecasting

Pankaj Das[1], Girish Kumar Jha[2] and Achal Lama[1]

[1]*ICAR- Indian Agricultural Statistics Research Institute, New Delhi*
[2]*ICAR- Indian Agricultural Research Institute, New Delhi*

Received 27 August 2021; Revised 15 December 2021; Accepted 29 December 2021

SUMMARY

Cointegration among the prices of different commodities plays a pivotal role in the price decision mechanism. In this study, we have attempted to improve the existing time delay neural network (TDNN) by incorporating the error correction term (ECT) as an auxiliary information in the model. The R package "ECTTDNN" has been developed for carrying out the analysis using the proposed model. The empirical study using monthly wholesale price indices of fruit and crude oil for the period January 2005 to November 2020, clearly demonstrated the superiority in terms of forecasting ability of the proposed hybrid model as compared to the usual TDNN model. This study adds to the rich literature of hybrid models and can be used for other cointegrated agricultural price series.

Keywords: Cointegration, ECM, TDNN, Hybrid model, ECT.

1. INTRODUCTION

Fruits and their products are one of the essential commodities in our daily life. The price forecasting of fruits is important for the local as well as the global market. The present production status of fruits and its products are not sufficient to meet the market demand. The agricultural market environment is changing with unprecedented speed both locally and globally. The dynamic nature of the market affects farm prices and thereby farm income. Most of the rural farmers are unable to understand and interpret the market and price behaviour to their advantage (Anjaly *et al.,* 2010). Market integration with different products also affects the price of fruit products. Fruit prices are volatile due to seasonality, inelastic demand, production uncertainty, and also because they are perishable. Besides, agricultural marketing is quite complicated, large numbers of marketing intermediaries are involved that adds the marketing cost and eventually increases the price. The effect of crude oil prices on fruits is a well-established phenomenon now. For fruit production and transportation of products to different places requires several types of machinery and farm equipment which are mostly operated through diesel or petrol. Thus the fluctuations in the price of crude oils affect the price of fruits. Zhang *et al.* (1998) reviewed the applications of neural networks in time series forecasting. The study showed a detailed survey for artificial neural network (ANN) modeling in time series forecasting along with future aspects. Jha and Sinha (2013) demonstrated the superiority of the ANN model in agricultural price forecasting. Paul *et al.* (2016) investigated the market integration of major pulses in fives zones of India. They tried to find out the wholesale and retail market integration using the vector error correction model (VECM). Kumar and Jha (2017) investigated co-movement and causality between prices of agricultural commodities and energy using the Johansen cointegration approach. David *et al.* (2019) worked on the cointegration between ethanol and agricultural commodity price series. They also showed how this relationship affects the predictability and efficiency of the cointegrated price series.

Corresponding author: Pankaj Das
E-mail address: pankaj.iasri@gmail.com

Our focused review of literature has revealed that most of the researchers have restricted themselves to either study the market integration or price forecasting of agricultural commodities independently. The incorporation of market cointegration information in the model for price forecasting of the product has not been attempted yet. In the present study, we have attempted to address this gap by proposing a cointegration based time-delay neural network (TDNN) hybrid model. This model employs the concept of the auxiliary variable for incorporating the cointegrating relationship among the study variables into the model for efficient price forecasting. The proposed hybrid model has been illustrated successfully on real data set on monthly price indices of crude oil and fruits and forecasting performances were compared using different statistical measures. The remaining portion of the paper is divided into materials and methods, results and discussions followed by the conclusion section.

2. MATERIALS AND METHODS

2.1 Data source

Monthly price indices of fruits and crude oil starting from January 2005 to November 2020 were used to develop a time-delay neural network-based hybrid forecasting model considering cointegration. The price index of crude oil was obtained from the International Monetary Fund (IMF) website (https://www.imf.org). The monthly price indices of fruits were collected from the Office of the Economic Advisor, Ministry of Commerce, Government of India (https://eaindustry.nic.in). The data sets contain 191 data points (January 2005 to November 2020) in Fig. 1.

2.2 Methods

In the study cointegration based time-delay neural network (TDNN) model has been proposed. For this, traditional cointegration or error correction model proposed by Granger (1981) has been used. Engle and Granger (1987)concentrated on the classical *I(0)/I(1)* cointegration framework where *d=1* and *b=1* and proposed a two-step method, where the model

$$(y_t - \beta x_t - \alpha) = \varepsilon_{1,t}$$

$$\Delta x_t = \varepsilon_{2,t} \tag{1}$$

where Δ indicates first differences, $\varepsilon_{1,t}$ and $\varepsilon_{2,t}$ are zero mean *I(0)* residuals an $\vec{\beta} = (1, -\alpha, -\beta)$ is the cointegration vector with an intercept α. In the second steps the error correction model

$$\Delta y_t = \phi_0 + \sum_{j}^{p} \phi_j \Delta y_{t-j} + \sum_{h=0}^{q} \theta_h \Delta x_{t-h} + \lambda \varepsilon_{1,t-1} + u_t \tag{2}$$

can be estimated with λ as the adjustment coefficient, ϕ_0 as constant and ϕ_j, θ_h as coefficients of short-run relationship with *p* lags of endogenous and

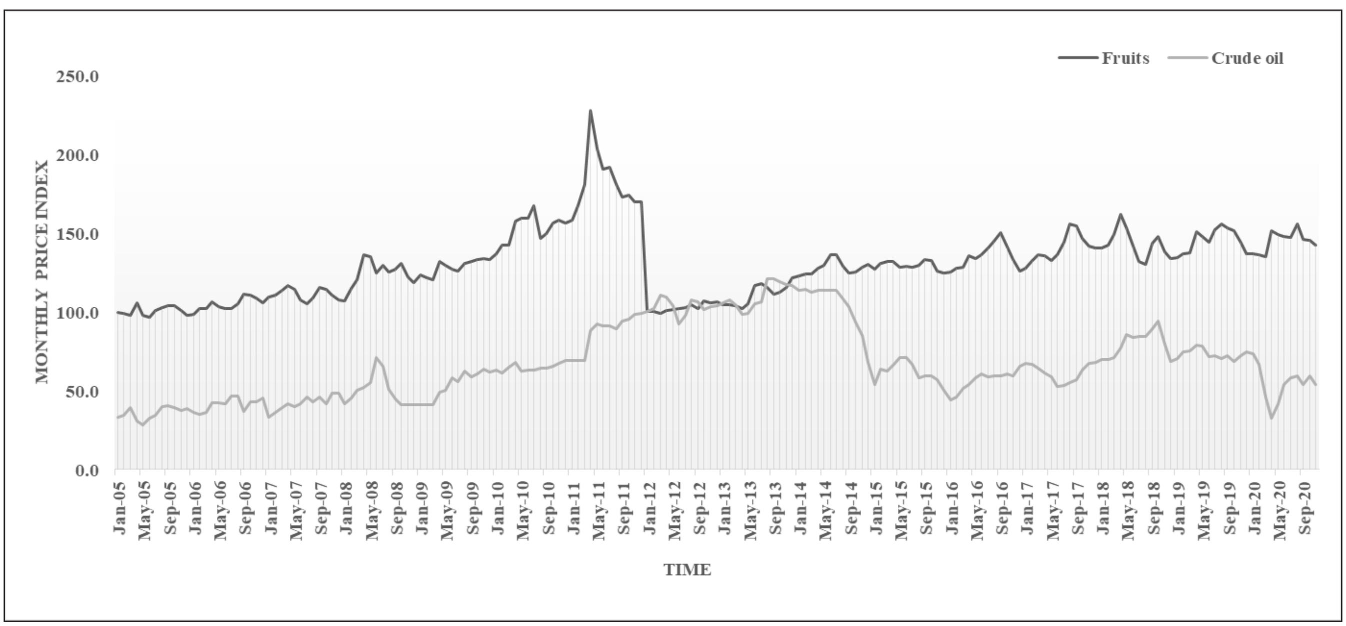

Fig. 1. Time plot of fruits and crude oil indexed price (nominal 2011-12=100).

q lags of exogenous variable. Thus the *Eq. 1* can be written as:

$$\begin{cases} \Delta^{d-b}(y_t - \beta x_t - \alpha) = \varepsilon_{1,t} \\ \Delta^d x_t = \varepsilon_{2,t} \end{cases} \quad (3)$$

where $\Delta^d = (1-B)^d$ and B is a backshift operator *i.e.* $Bx_t = x_{t-1}$; $0<b<1$.

Nonlinear error correction model was defined by Kapetanios *et al*. (2006)

$$\Delta Y_t = \phi_0 + \Lambda \vec{\beta}' y_{t-1} + \Psi(\vec{\beta}' y_{t-1}) + \sum_i^p \phi_i \Delta y_{t-1} + U_t \quad (4)$$

Where Y_t is a matrix containing the time series y_t, ϕ_0 is a vector of constant, ϕ_i is a matrix of the coefficients of lagged endogenous variables, Λ is the adjustment coefficient and U_t is a matrix of white noise residuals and ψ is a nonlinear function. Instead of a defined functional form of ψ a NN function η has been used. The modified equation can be written as:

$$\Delta^d Y_t = \eta(\Delta^d Y_{t-1}, ..., \Delta^d Y_{t-p}, \Delta^{d-b} \vec{\beta}' y_{t-1}) + U_t \quad (5)$$

here, all variables are assumed to be stationary.

The general expression of a TDNN with single hidden layer is given by Jha and Sinha (2014) -

$$\hat{y}_t = g\left(\alpha_0 + \sum_{j=1}^{q} \alpha_j f\left(\beta_{0j} + \sum_{i=1}^{p} \beta_{ij} y_{t-i}\right)\right) \quad (6)$$

where $\hat{y}_t$ is the predicted value for y_t at time t, p input and q hidden nodes, i, j is i^{th} node of input layer, j^{th} node of hidden layer respectively, y_{t-i} ;$(i = 1,2,\dots p)$ are network input nodes. β_{ij} $(j = 1,2, \ . \ . \ ., q)$ refer to the connection weight between i^{th} and j^{th}neuron. α_j refer the weight between j^{th} neuron of hidden node and output node. α_0 *and* β_{0j} are bias term for output layer and j^{th} hidden node. f and g are respectively hidden and output layer activation function, mainly logistic $f(v_j) = \frac{1}{1+e^{-v_j}}$ and g is an identity function. The above mentioned TDNN model has been used for forecasting.

3. RESULTS AND DISCUSSION

The summary statistics of the datasets are given in the Table 1. It was observed that both the price series were positively skewed. Further Jarque-Bera test (Jarque and Bera, 1987) was tested to check the normality behaviour of the both crude oil and fruit dataseries (Table 1). The p values of the test was less than 0.01 for the dataseries. It indicated that the variables follow non-normal distribution and datasets were leptokurtic in nature.

Table 1. Descriptive statistics and Normality test of the monthly data series

Descriptive statistics	Crude oil	Fruits
Mean	67.50	130.22
Median	62.90	129.70
Maximum	121.10	227.70
Minimum	28.30	96.61
Std. Dev.	23.91	22.20
CV (%)	35.53	17.08
Skewness	0.56	0.86
Kurtosis	-0.68	-1.63
Jarque-Bera Test	68.01	42.02

3.1 Cointegration test

In the case of nonstationarity of the time series, cointegration provides an appropriate statistical technique to investigate if there is a statistically significant relationship between the time series. Accordingly, the first step is determination of nonstationary nature of the price series. Nonstationarity behaviour of the crude oil and fruits price series was checked using the ADF (Dickey and Fuller, 1979) and PP test (Phillips and Perron, 1988). The null hypothesis of both test is the series is nonstationary. The results (Table 2) highlighted that both the price series were nonstationary at the level and stationary after first differencing. These results indicated that both the data series were integrated of order one *I(1)* and suitable for cointegration analysis.

Table 2. Stationarity test of data series

Series		Augmented Dickey-Fuller		Phillip-Perron	
		t-statistic	Prob.	t-statistic	Prob.
Crude oil	Level	-1.14	0.81	-1.156	0.78
	1st difference	-14.086	<0.001	-9.952	<0.001
Fruits	Level	-2.80	0.24	-17.29	0.13
	1st difference	-13.16	<0.001	-22.229	<0.001

The nonlinearity of the data series was checked using BDS (Brock *et al.,* 1996) test (Table 3). The results indicated that both the fruits and crude oil data series were nonlinear. The cointegration of the price series was checked using Johansen's cointegration test. For the optimal

Table 3. Brock- Dechert-Scheinkman (BDS) test

Series	Embedding dimension				Conclusion
	2		3		
	Statistics	Probability	Statistics	Probability	
Fruits	112.794	< 0.001	199.717	< 0.001	Nonlinear
	51.822	< 0.001	63.787	< 0.001	
	40.849	< 0.001	43.339	< 0.001	
	37.268	< 0.001	36.669	< 0.001	
Crude oil	63.750	< 0.001	102.304	< 0.001	Nonlinear
	83.517	< 0.001	106.389	< 0.001	
	53.964	< 0.001	58.675	< 0.001	
	40.097	< 0.001	40.129	< 0.001	

lag length for Johansen's cointegration test, the vector autoregression (VAR) model was applied at one to ten lags. For selecting the optimal length, four criteria were used i.e. Akaike information criterion (AIC), Hannan-Quinn information criterion (HQ), Schwarz information criterion (SIC), and Final prediction error (FPE). Based on the information criteria, the lag length of 3 was fixed for the model with fruits as the dependent variable and crude oil as the independent variable. All the computations were carried out in R Studio and the required packages were "vars" and "urca". Both Johansen trace-based (Johansen, 1988), as well as maximum eigen value based tests, were used to find cointegration. The test results indicated that there was cointegration between crude oil price and fruit price (Table 4).

Table 4. Johansen cointegration rank test

Fruits vs Crude oil	Test statistic	Prob.	Remarks
λ_{trace} $H_0: r = 0\, vs\, H_1: r \geq 1$ $H_0: r \leq 1\, vs\, H_1: r \geq 2$	12.31 0.62	0.14 0.42	*r=1* not rejected. Co-integration occurs.
λ_{max} $H_0: r = 0\, vs\, H_1: r \geq 1$ $H_0: r \leq 1\, vs\, H_1: r \geq 2$	11.68 0.62	0.12 0.42	

The detailed estimated parameters of VAR models are given in Table 5 with the speed adjustment factor and the cointegrating vector β. The estimated value of β was -0.28 for the fruits vs crude oil model. This provided strong evidence of a long-run relationship among the variables.

Table 5. Estimated parameter value of fitted VAR models

Model	Regressors	Parameter estimates	t-test	p-value
Fruits vs Crude Oil	β ECT_{t-1}	-0.28 -0.13	-5.66 -4.40	<0.01 <0.01

3.2 Proposed TDNN model

The estimated value of the error correction term (ECT) was used for building the proposed time-delay neural network (TDNN) model. The main concept of the proposed TDNN model was the development of a model that uses cointegration behaviour among the data series when the data is nonlinear and nonstationary. The estimated ECT of the fitted VAR models was incorporated in the TDNN model as auxiliary information (Fig. 2). The fitting of the proposed TDNN model was done in R-Studio with the help of our developed package "ECTTDNN" (Das *et al.*, 2021). The developed "ECTTDNN" package replaces necessity of packages"vars" "urca" and "nnet". It first find out the cointegration and ECT from a cointegarting data series. Later it fits a TDNN model with an auxiliary information ECT on the data series.

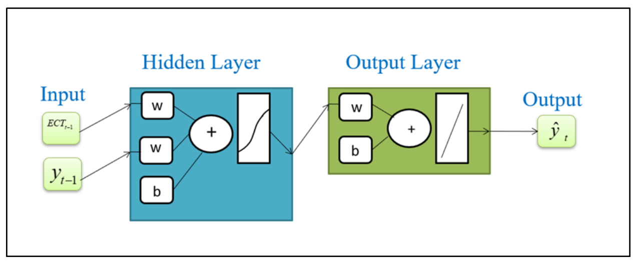

Fig. 2. Proposed hybrid TDNN model framework

3.3 Performance of the proposed TDNN model

For training of the proposed model, the first 179 observations were used and the remaining last 12 observations were kept to check the generalization power of the model. As mentioned earlier, in practice, a simple neural network structure with a small number of parameters is preferred due to better generalization ability for out of the sample data. Accordingly, we varied input lags from one to five, and the number of hidden nodes from one to ten. TDNN model with three input lags and five hidden nodes was found as the best model in terms of overall accuracy criterion such as the root mean squared error, mean absolute error, etc. Repeats were tried from 10 to 30 for obtaining the best forecast from the TDNN model. Repeat means the number of networks that were averaged for getting the output. In this study, repeats = 26 i.e. 26 neural networks were averaged to get the desired forecast.

Table 6. Parameter values used for TDNN model fitting

	Lag	Hidden node	Repeats	xreg	Maxit	BoxCox Parameter value
Fruits	3	5	26	-0.13	200	0.5

Further, the estimated value of ECT (Table 5) was taken as a subset i.e. subset of the variable. Maximum numbers of iterations "maxit" for neural network fitting was checked from 100 to 250. The best result was obtained at 200 iterations. The parameter values used for fitting are reported in Table 6. The performance criteria like maximum error (ME), root mean squared error (RMSE), mean absolute error (MAE), mean absolute percentage error (MAPE) (Das *et al.* 2019) are considered for evaluation of the fitted models Table 7 describes the in-sample performance of the ECT based TDNN model and usual TDNN model. It has been noticed that ECT based TDNN model performs better as compared to the standard TDNN model. The percentage of improvement concerning MAPE for the TDNN model with ECT was 9.00.

Table 7. In-sample performance of proposed model for fruits data

Training set	ME	RMSE	MAE	MAPE	% of improvement
TDNN	0.34	5.58	3.66	0.46	9
TDNN with ECT	0.29	4.61	2.78	0.37	

Further, the generalization power of this model was checked using a hold-out sample of the last 12 observations. The results of the TDNN and TDNN model with ECT are given in Table 8. The performance measures showed that the proposed TDNN model with ECT performed better than the standard TDNN model. The rate of improvement of proposed model in out-sample was 18.18%. The percentage of improvement for ECT based TDNN has also improved in out-sample.

Table 8. Out-sample performance of proposed model for fruits data

Testing set	ME	RMSE	MAE	MAPE	% of improvement
TDNN	0.25	6.12	4.19	0.33	18.18
TDNN with ECT	0.23	5.89	3.92	0.27	

The out-sample performance of the fitted models is also shown in Fig. 3. The square dots described the original time series values. The triangle dots line and diamond dots denote the forecasted value of TDNN with ECT model and standard TDNN model respectively. Fig. 3 indicated that the forecasted values of the proposed TDNN model with ECT were closer to the original data point, while predicted values of standard TDNN model deviated from the original data points. It also observed the predicted values of standard TDNN model almost stable after 6th predicted value where as the proposed ECT-TDNN model predicts not

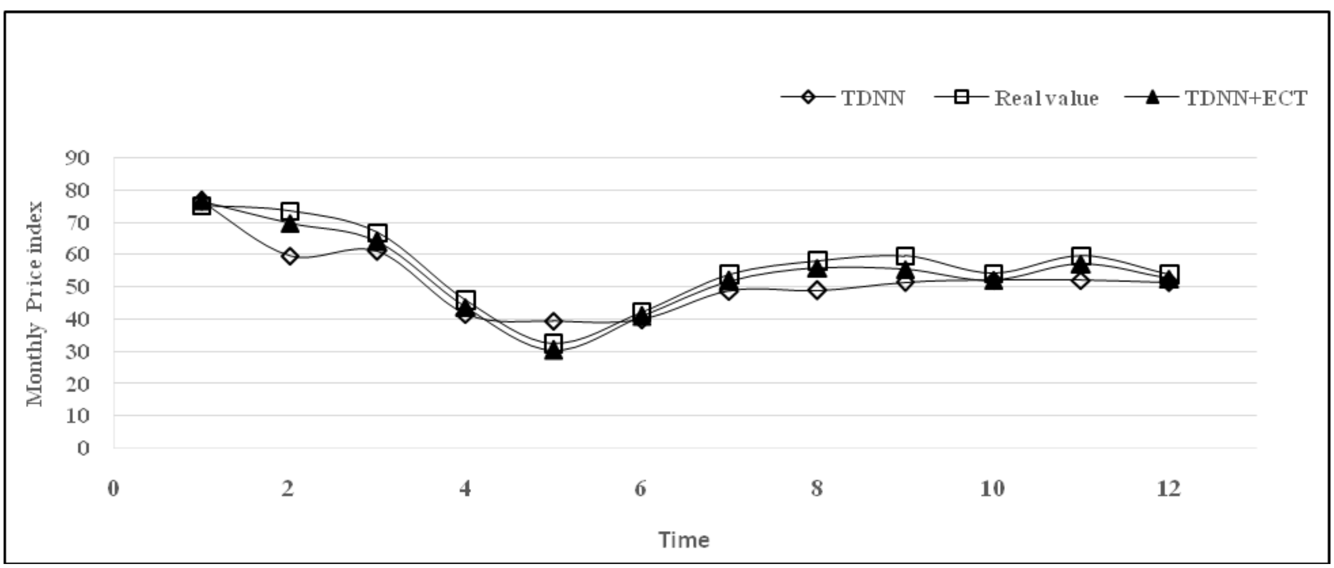

Fig. 3. Original values with forecasted values of proposed TDNN model for fruits price

same. Hence, the prediction of the proposed model is more accurate than the standard TDNN model.

Generally, for a nonstationary time series, differencing of the data series lead to loss of some information. The ECT from cointegration model help to improve the performance of the TDNN model by incorporating this lost information. As a result, the proposed TDNN model with ECT generated superior result than the single TDNN model. So, the TDNN model with ECT can be used for further forecasting.

4. CONCLUSION

This study has put on concentrated efforts to improve the prediction ability of the standard time-delay neural network (TDDN) model by incorporating the error correction term (ECT) obtained from the co-integration analysis. The ECT of the error correction model is used as auxiliary information in TDNN model for forecasting the monthly wholesale price index of fruits. The performance of the model is evaluated on basis of fit statistics like RMSE, MAD, MAPE and ME. The study suggests that the researchers should also focus on the cointegration analysis for better forecasting accuracy of the agricultural commodities. The proposed approach can be applied to the variety of agricultural price series where cointegrating relationship exists.

ACKNOWLEDGEMENTS

The authors are thankful to anonymous reviewers for their valuable comments and the Director, ICAR-IASRI for providing facilities for carrying out the present research.

REFERENCES

Anjaly K.N., Surendran S., Babu S.K. and Thomas J.K. (2010). Impact Assessment of Price Forecast: A Study of Cardamom Price Forecast by AMIC, KAU. *NAIP on Establishing and Networking of Agricultural Market Intelligence Centres in India*. College of Horticulture, Vellanikkara, India.

Bishop M.C. (1995). *Neural networks for Pattern recognition*. Oxford University Press, New York.

Brock W.A., Scheinkman J.A., Dechert W.D. and LeBaron B. (1996). A test for independence based on the correlation dimension. *Econom Rev*., **15**, 197-235.

Das P. (2019). *Study on machine learning techniques based hybrid model for forecasting in agriculture*. http://krishikosh.egranth.ac.in/handle/1/5810147805; 2019.

Das P., Lama A. and Jha G.K. (2021). Package ECTTDNN: https://CRAN.R-project.org/package=ECTTDNN.

David S.A., Inácio C.M.C. and Machado A.J.T. (2019). Quantifying the predictability and efficiency of the cointegrated ethanol and agricultural commodities price series. *Appl Sci*., **9(24)**, 1-14.

DES (Directorate of Economics and Statistics) Cost of cultivation of principal crops in India (various issues), Government of India, New Delhi.

Dickey D.A. and Fuller W.A. (1979). Distribution of the estimators for autoregressive time series with a unit root. *J Am Stat Assoc*., **84**, 427-431.

Engle R. and Granger C. (1987). Cointegration and error correction: representation, estimation and testing. *Econometrica*., **55**, 251-276.

Engle R.F. and Yoo B.S. (1987). Forecasting and testing in co-integrated systems. *J Econom*., **35**, 143-159.

Granger C.W.J. (1981). Some properties of time series data and their use in econometric model specification. *J Econom.*, **16**, 121-130.

Jarque C.M. and Bera A.K. (1987). A test for normality of observations and regression residuals. *Int Stat Rev*., **55(2)**, 163-172.

Jha G.K. and Sinha K. (2013). Agricultural price forecasting using neural network model: An innovative information delivery system. *AgriEconoRReview*., **26**, 229-239.

Jha G.K. and Sinha K. (2014). Time-delay neural networks for time series prediction: an application to the monthly wholesale price of oilseeds in India. *Neural Comput Appl*., **3-4**, 563-571.

Johansen S. (1988). Statistical analysis of cointegration vectors. *J Econ Dyn Control*., **12**, 231-254.

Kapetanios G., Shin Y. and Snell A. (2006). Testing for cointegration in nonlinear smooth transition error correction models. *Econ Theory*., **22**, 279-303.

Kumar R.R. and Jha G.K. (2017). Examining the co-movement between energy and agricultural commodity prices in India. *J Indian SocAgric Stat.,***70(3)**, 241-252.

Paul R.K., Saxena R. and Bhat S.A. (2016). How price signals in pulses are transmitted across regions and value chain? Examining horizontal and vertical market price integration for major pulses in India. *AgriEcono R Review,* **22**, 75-86.

Phillips P.C.B. and Perron P. (1988). Testing for a Unit Root in Time Series Regression. *Biometrika*., **75 (2)**, 335-346.

R Core Team (2018). R: A language and environment forstatistical computing. R Foundation for Statistical Computing, Vienna, Austria. (https://www.R-project.org/.)

Zhang G.P., Zhang B.E. and Hu M.Y. (1998). Forecasting with artificial neural networks: The state of the art. *Int J Forecast*., **14**, 35-62.