International Journal of Agriculture Sciences
ISSN: 0975-3710 & E-ISSN: 0975-9107, Volume 14, Issue 9, 2022, pp.-11713-11715.
Available online at https://bioinfopublication.org/pages/jouarchive.php?id=BPJ0000217

Research Article

LENGTH OF SERIES AND FORECASTING ACCURACY OF ARIMA MODELS: ILLUSTRATION WITH CRUDE PETROLEUM PRICES AND EUR/USD EXCHANGE RATES

VEDAMURTHY K.B.*, RANGEGOWDA R. AND NIRMALA S.R.
Department of Dairy Business Management, Dairy Science College, Hebbal, 560024, Karnataka Veterinary, Animal and Fisheries Sciences University, Bidar, 585401, Karnataka, India
*Corresponding Author: Email - rangegowda33@yahoo.com

Received: September 03, 2022; Revised: September 26, 2022; Accepted: September 27, 2022; Published: September 30, 2022

Abstract: The study was made to see whether the currency fluctuations had any effect on crude petroleum prices through the transfer function, variant of ARIMA. The lag structure was found to be one and hence lagged exchange rates were used as an exogenous variable in the ARIMA model. The results showed a significant effect of exchange rate on crude petroleum prices, which implies that an increase in the exchange rate of the dollar would increase the petroleum prices significantly. The forecasted values for Dec. 2019 about $ 55 and in May 2020 about $ 62.

Keywords: *ARIMA, GARCH, Transfer Function, Forecasting, Mean, Variance, Crude Petroleum Prices, Exchange Rates*

Citation: Vedamurthy K.B., *et al*., (2022) Length of Series and Forecasting Accuracy of ARIMA Models: Illustration with Crude Petroleum Prices and EUR/USD Exchange Rates. International Journal of Agriculture Sciences, ISSN: 0975-3710 & E-ISSN: 0975-9107, Volume 14, Issue 9, pp.- 11713-11715.

Copyright: Copyright©2022 Vedamurthy K.B., *et al*., This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution and reproduction in any medium, provided the original author and source are credited.

Academic Editor / Reviewer: D D Pardhe, Dr Hemangi Mehta, Dr Vijaya Lakshmi V

Introduction

Time series forecasting, one of the most applied data science techniques remain something like a black horse. Among the many issues in timeseries forecasting, the length of the series required to build a model continuous to be challenging. Generally, the availability of sufficient data was itself a problem in forecasting. However, with increasing availability of data, the new question is, whether to include all the available data or include only recent data. It is unrealistic to assume the same seasonal pattern remains for long period. Therefore, it makes sense to include most recent data. However, what should be that recent period? This paper is a modest attempt to find objectively the length of the series to be included for building a model when dealing with long series.

Material and Methods

Monthly crude petroleum price from Feb, 1960 to May 2019 and the exchange rates (EURO/USD) from Feb,1975 to May 2019. This is not a big data for machine learning algorithms. However, when dealing with time series data, it is unrealistic to assume the seasonal pattern to remain same for more than four decades. Therefore, it is important to find the structural changes in the series.

Test of linearity

Both the series were tested for linearity using the R package nonlinearity Test which employs Teraesvirta's neural network and White neural network tests. For crude petroleum series, Teraesvirta's neural network test with x-squared value of 8.58 was significant which is evident from the lower p-value of 0.0136. X squared value for exchange rate series is 5.9 with 0.05 p value. Thus, it is concluded that the both series were nonlinear.

Structural Change and Breakpoints

The structural break points are obtained by strucchange package of R. strucchange features tests/methods from the generalized fluctuation test framework as well as from the F test (Chow test) framework. This comprises methods to fit, plot and test fluctuation processes (e.g., CUSUM, MOSUM, recursive/moving estimates) and F statistics, respectively.

It is possible to monitor incoming data online using fluctuation processes. Finally, the breakpoints in regression models with structural changes can be estimated together with confidence intervals [1].

Given the number of breakpoints, the function 'breakpoints' in r computes the optimal breakpoints. The procedure is concerned with testing or assessing deviations from stability in the classical linear regression model

$yi = x > i\, \beta + ui.$

It is reasonable to assume that there are m breakpoints, where the coefficients shift from one stable regression relationship to a different one. Thus, there are m + 1 segments in which the regression coefficients are constant, and the model can be rewritten as

$yi = x > i\, \beta j + ui \qquad (i = ij-1 + 1, \ldots, ij\,, j = 1, \ldots, m + 1)$

where j denotes the segment index. The algorithm for computing the optimal breakpoints given the number of breaks is based on a dynamic programming approach. The underlying idea is that of the Bellman principle. The main computational effort is to compute a triangular RSS matrix, which gives the residual sum of squares for a segment starting at observation i and ending at I' with i < I'. Finally, the number of optimal breakpoints is selected based on minimum BIC [2].

This resulted in four optimal breakpoints, hence five segments for crude petroleum series and five optimal breakpoints and six segments for exchange rate series which is shown in [Table-1&2]:

Table-1 *Breakpoints corresponding to break dates: Crude Petro Series*

Breakpoints	Breakyear					BIC
m=1					2005 (2)	6875
m=2		1978 (11)			2005 (2)	5952
m=3		1978 (11)	1987 (9)		2005 (2)	5717
m=4		1978 (11)	1987 (9)	1996 (7)	2005 (5)	5707
m=5	1970 (1)	1978 (11)	1987 (9)	1996 (7)	2005 (5)	5708

Fit:

m	0	1	2	3	4	5
BIC	6875	5952	5717	5709	5707	5708

Table-2 *Breakpoints corresponding to break dates: EUR/USD Exchange rate Series*

Breakpoints	Break year					BIC
m=1				2003 (11)		-403.32
m=2	1981 (9)	1988 (5)				-518.45
m=3	1981 (9)	1988 (5)		2003 (11)		-599.34
m=4	1981 (9)	1988 (5)	1997 (3)	2003 (10)		-701.05
m=5	1981 (9)	1988 (5)	1997 (3)	2003 (11)	2012 (11)	-759.00

Fit:

m	0	1	2	3	4	5
BIC	-330.87					

Dividing the data into training and validation datasets

Based on the optimal breakpoints the data was divided into different segments for training and testing as given in table. First training dataset for both series is from the beginning of the period to May, 2017. Then onwards, the training dataset is selected based the beginning of the corresponding breakpoint to May, 2017. Thus, the five models are obtained for crude and six models for the exchange rate series [Table-3].

Table-3 *Training and Test datasets used in building the models*

Data Segments	Crude Petroleum Series		Exchange Rate Series	
	From	To	From	To
Test	June, 2017	May, 2019	June, 2017	May, 2019
Train-1	Feb, 1960	May, 2017	Jan, 1975	May, 2017
Train-2	Dec, 1978	May, 2017	Oct, 1981	May, 2017
Train-3	Oct, 1987	May, 2017	June, 1988	May, 2017
Train-4	July, 1996	May, 2017	Apr, 1997	May, 2017
Train-5	June, 2005	May, 2017	Dec, 2003	May, 2017
Train-6			Dec, 2012	May, 2017

Forecasting and comparing for accuracy

ARIMA models were used for forecasting the time series models. The models were built to different training sets and validated using the test data. The best model was selected using the MAPE and Janus Coefficient as accuracy measures. Janus coefficient is a measure of forecast accuracy, calculated as the ratio of average of squared forecast errors for extra sample data to the comparable average for in sample data [3].

Augmenting the Forecast

The training datasets were tested for GARCH effect using the residuals of the ARIMA model. Wherever, there was GARCH effect, the ARIMA forecasts were augmented using the forecasts of estimated GARCH model.

*Augmented_forecast = arima_forecast + sqrt(sigma)*2*

Sigma is obtained from the GARCH model.

Results and Discussion

The oil and currency are two most important commodities which affect the global economy and polity and hence been the favourite for time series studies. From Fig-1, it can be observed that oil prices did not fluctuate during 1960 to 1980 except for the sharp spike in mid 70s. Till the beginning of 2000 there was a gradual decline in the prices. However, the real fluctuation started from 2000 onwards. However, unlike crude, the currency has been fluctuating since from 1975 (Fig-2). As discussed earlier, the linearity tests showed that the both series were nonlinear. Hence the strucchange package in R applied to find the breakpoints. This resulted into five and six optimal segments in crude and exchange rate respectively.

Fig 1:Structural Changes in Crude Oil Prices (in current U.S. dollars per barrel)

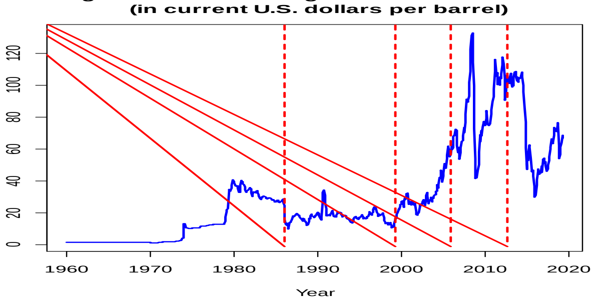
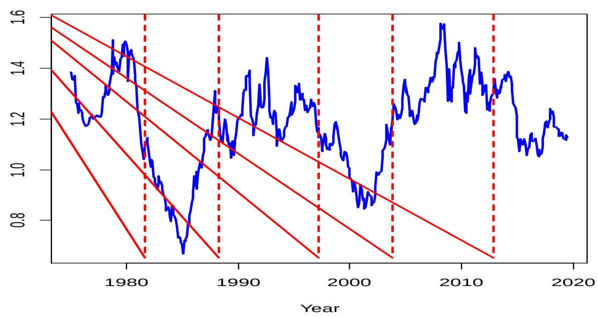


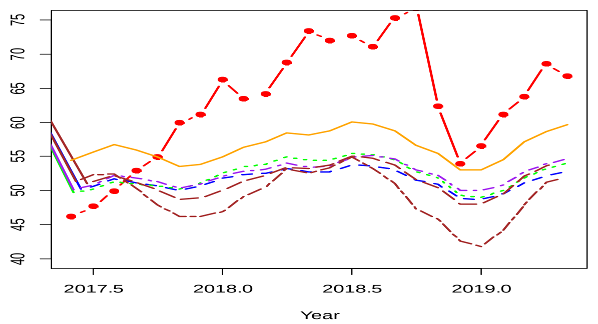
Fig 2: Structural Changes EURO/USD Prices



The training dataset for the both the series were decided based on the breakpoints and are depicted in the [Table-3]. The ARIMA models were employed on the training datasets and tested using the test dataset. The forecasts obtained for crude petro are plotted against the test data and are depicted in the Fig-3 and the results of accuracy measures are given in [Table-4].

The results show that Model-2 scored better on all the accuracy measures. However, a glance at the [Fig-3] shows that all the models have underestimated the actual crude prices, but the model-5 could capture the turning points better. Therefore, the series was tested for the GARCH effect using the residuals of the model. The forecasts of the estimated GARCH model were used to augment the ARIMA forecasts that yielded better forecasts. Thus, using the recent segments of the given long series has resulted in better forecasts compared to the models with earlier segments.

Fig 3:Forecasted V/S Actual:Crude Petro



To reinforce the findings the forecasts were obtained for the exchange rate series using the training datasets with different segments. The results are presented in the [Table-5]. The model-3 estimated on recent three regimes yielded better ex-ante forecasts evidenced by the measures of accuracy. Since there was no perceptible GARCH effect, it was not incorporated in the forecasts as in the case of crude petroleum series.

Conclusion

For long period time series data, the challenge is to choose the best length of series for obtaining accurate forecasts. Monthly crude petroleum price and the exchange rates (EURO/USD) from Feb, 1975 to May 2019 was used to decide optimum length of the series (www.investing.com). Both the series were tested for linearity using the R package nonlinearity Test which employs Teraesvirta's neural network and White neural network tests and found to be non-linear. This resulted in four and five structural break points for Crude Petroleum and the Currency prices respectively. ARIMA models were fitted for different lengths of period based on the indicated breakpoints. The series was divided into two parts a test and a train. The train data covered the respective periods being considered. Testing data spanned the period from June 2017 to May 2019. Models were built on train dataset and validated on the test dataset using accuracy measures such as RMSE, MAPE and Janus Quotient. For the crude petroleum, the model with recent three structural regimes (Jan, 2006 to May 2017) scored better on all accuracy measures. However, the model with the recent two structural regimes captured the turning points better than the other models, but scored low on the accuracy measures as the forecasts were under estimated.

Table-4 *Measures of Accuracy: Crude Petro*

Models	Length	Period	(p,d,q) (P,D,Q)	AIC	RMSE	MAPE	Janus Coeff
Model-1	689	1960-02 -2017-05	(1,1,3) (3,1,2)	3401	13.8	23.3	0.119
Model-2	462	1978-12 -2017-05	(2,1,2) (3,1,2)	2442	12.8	21.2	0.110
Model-3	356	1987-10 -2017-05	(1,1,2) (3,1,1)	1944	12.8	21.2	0.110
Model-4	251	1996-07 -2017-05	(1,1,2) (2,1,1)	1437	13.6	23.4	0.118
Model-5	144	2005-06 -2017-05	(1,1,2) (3,1,1)	860	15.9	29.7	0.141
Model-6	144				11.9	20.2	0.101

Table-5 *Measure of Accuracy: EURO/USD*

Models	Length	Period	ARIMA (p,d,q) (P,D,Q)	AIC	RMSE	MAPE	Janus Coeff.
Model-1	508	1975-01 -2017-05	(0,1,0) (4,1,1)	-1837	0.05	3.6	0.021
Model-2	428	1981-10 -2017-05	(0,1,0) (4,1,1)	-1542	0.04	3.5	0.020
Model-3	348	1988-06 -2017-05	(0,1,0) (4,1,1)	-1226	0.04	3.5	0.020
Model-4	242	1997-04 -2017-05	(0,1,0) (4,1,1)	-833	0.06	4.3	0.025
Model-5	162	2003-12 -2017-05	(0,1,0) (4,1,1)	-508	0.05	4.1	0.024
Model-6	54	2012-12 -2017-05	(0,1,0) (1,1,1)	-151	0.1	9.4	0.046

Therefore, the series was tested for the GARCH effect using the residuals of the model. The forecasts of the estimated GARCH model were used to augment the ARIMA forecasts that yielded better forecasts. To reinforce the above findings the procedure was repeated for exchange rates. Model estimated on recent two regimes yielded better ex-ante forecasts evidenced by the measures of accuracy. Since there was no perceptible GARCH effect, it was not incorporated in the forecasts as in the case of crude petroleum series.

Application of research: In a time series, seasonal pattern may change over a long period. Therefore, it becomes important to decide what part of the series to be included for building a time series model. Dividing the series into different segments by assessing deviations from stability and then building model and validating against the test data would help in deciding the length of the series to include. The most recent segments are more important in building a better model.

Research Category: Business Management

Acknowledgement / Funding: Authors are thankful to Department of Dairy Business Management, Dairy Science College, Hebbal, 560024, Karnataka Veterinary, Animal and Fisheries Sciences University, Bidar, 585401, Karnataka, India

****Research Guide or Chairperson of research: Dr K B Vedamurthy**
University: Karnataka Veterinary, Animal and Fisheries Sciences University, Bidar, 585401, Karnataka, India
Research project name or number: Research station study

Author Contributions: All authors equally contributed

Author statement: All authors read, reviewed, agreed and approved the final manuscript. Note-All authors agreed that- Written informed consent was obtained from all participants prior to publish / enrolment

Study area / Sample Collection: Dairy Science College, Hebbal, 560024

Cultivar / Variety / Breed name: Nil

Conflict of Interest: None declared

Ethical approval: This article does not contain any studies with human participants or animals performed by any of the authors.
Ethical Committee Approval Number: Nil

References

[1] Zeileis A., Leisch F., Hornik K., Kleiber C. (2002) *Journal of Statistical Software*, 7(2), 1–38.
[2] Zeileis A., Kleiber C., Krämer W., Hornik K. (2003) *Computational Statistics and Data Analysis*, 44, 109-123.
[3] Gadd and Wold H. (1964) *In "Econometric model building : essays on the causal chain approach". Amsterdam : North-Holland Publishing Company*, 229-235.