v. 3, n. 6, 2023

Journal of
Engineering
Research

USE OF NON-OBSERVABLE COMPONENT DECOMPOSITION MODELS TO PREDICT NON-STATIONARY TIME SERIES OF AGRICULTURAL COMMODITIES

Lucas Valle Mielke
ICMC — USP

Paulino Ribeiro Villas Boas
Embrapa (Instrumentation)
ICMC — USP

INTRODUCTION

Prediction of time series is often done by Autoregressive Moving Average Models (ARMA), the main disadvantage of these models being the requirement that the time series studied be stationary, which is often not the case [1]. To work around this problem, data differentiation processes are usually performed, which can be done directly on the dataset by calculating the difference of the value of the variable in step t with the value of the previous step t – 1 [4], or by setting the models to successively differentiate the analyzed series until it becomes stationary, as in the case of Autoregressive Integrated Moving Average Models (ARIMA, Autoregressive Integrated Moving Average in English) and their variations [3].

Despite the possibility of differentiation, the results of the ARIMA models continue to be sensitive to disturbed data that do not develop around a constant mean [3]. In view of this, the Unobserved Component Models (UCM) emerge as a promising alternative to these models because they do not assume stationarity of the data, in addition, the UCM models can be understood in trend components, cycle and perturbation, which makes them especially useful for analyzing series that have a cycle, as is the case of agricultural crops [1]. In view of this, the objective of this study was to compare the performance of Decomposition Models into Non-Observable Components with the performance of Autoregressive Moving Average Models for the prediction of non-stationary time series, in this case the price of the commodities of Rice, Coffee, Corn, Soy and Wheat.

MATERIALS AND METHODS

DATABASE

To elaborate the proposed model, spot prices in reais (Brazilian currency) and Dollars of Rice, Coffee, Corn, Soybeans and Wheat were collected from the price survey carried out by the Center for Advanced Studies in Applied Economics (CEPEA) of ''Universidade de São Paulo" (USP) to the period between June 30, 2012 and June 30, 2022, totaling 10 years. The spot price in reais was used as a predictive or dependent variable, while the dollar price was used to calculate the dollar rate for each sample, to be used as an explanatory or independent variable.

ANALYTICS AND MACHINE LEARNING

To ensure that the analyzed time series are not stationary, we perform the Augmented Dickey–Fuller unit root test (ADF), where the null hypothesis (p-value > 0.05) is that the series has roots and belongs to the non-stationary series and, the alternative (p-value ≤ 0.05), that there is no unit root and the series is a stationary sequence. Table 1 displays the test result for each of the analyzed cultures. We also decomposed the series into trend, cycle and perturbation in order to verify the behavior of these components for all series and identify behavior patterns [2].

For machine learning, we express the problem in the form of an equation where we have a Y variable of interest that can be explained by X vectors, called explanatory or independent variables in the way it occurs in a classical linear model. The dataset with the vectors was prepared for machine learning training and testing. The original series sequence was maintained, of which 85% of the oldest data were used for training and 15% of the most recent data for testing.

The data were trained and tested using the UCM and seasonal ARIMA models with exogenous SARIMAX (Seasonal Auto-Regressive Integrated Moving Average with eXogenous factors) models. The second allows working with Seasonal series, as is the case with agricultural crops, and including exogenous variables in the equation, in this case the dollar rate. Then, the result of each of the models was compared for each of the cultures by measuring the Mean Absolute Percentage Error (MAPE). Table 1 displays the MAPE of both models for all cultures studied.

RESULTS AND DISCUSSION

The ADF Test obtained a p-value greater than 0.05 for all agricultural crops analyzed, indicating that the price of none of them is stationary, as expected. Table 1 presents the results of the ADF test for all series. The decomposition of the series generated trend, cycle and noise curves. The trend curves allowed us to verify relative price stability for all commodities between the beginning of the analyzed period and the end of 2019, when there was a sharp increase in prices, especially between the beginning of 2020 and the middle of 2021 for Corn, Soy and Wheat. Rice also showed a sharp growth trend from 2020 onwards, but since January 2021 it has shown a downward trend, while coffee had a later tendency to grow, from December 2020 onwards demonstrated the annual seasonality of the prices of the crops analyzed, with a price peak occurring in the second half, with the exception of Corn, whose period of greater seasonality occurs very close to the moment of lesser seasonality for Soybean. Figure 1 presents the trend curves of all analyzed series, while figure 2 presents the cycle curves for Corn and Soybean.

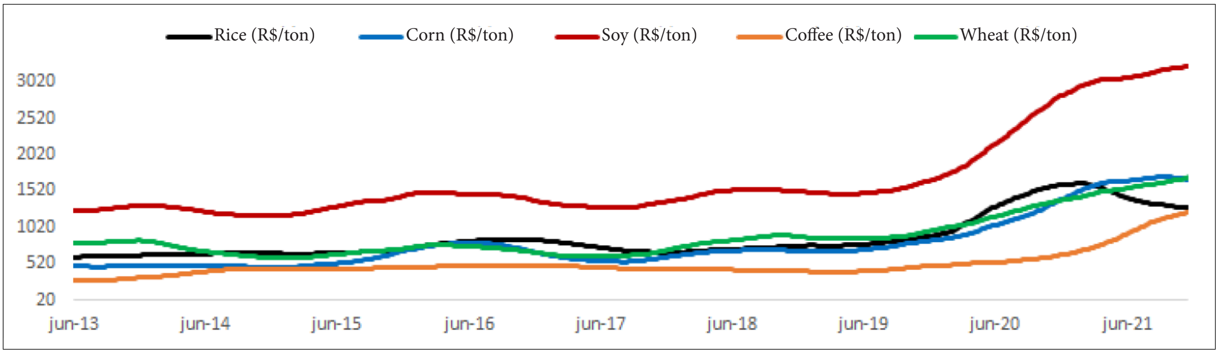

Figure 1: Trend curves of all analyzed price series.

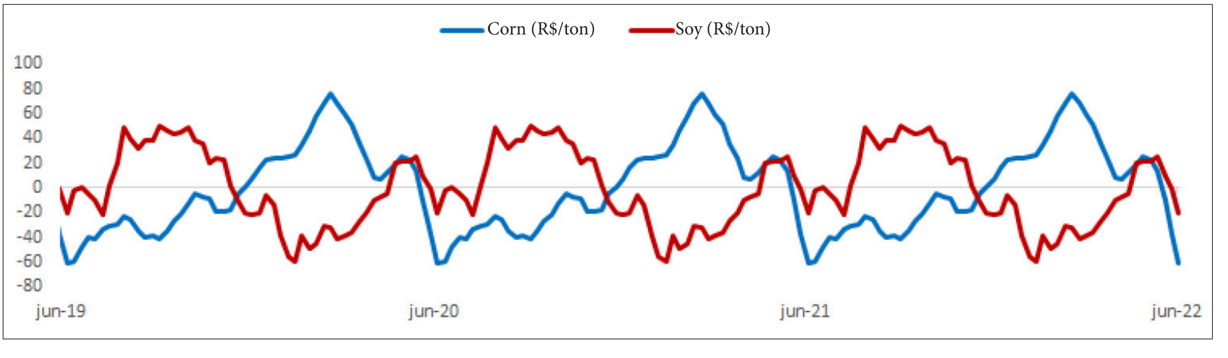

Figure 2: Corn and Soybean price cycle curves.

When analyzing the results of the machine learning models, it was found that the MAPE was lower with the use of the UCM model in all cultures, indicating that this model may be more promising for the type of analysis in this work. However, it must be noted that the error was greater than 18% for all cultures, indicating that both models did not reasonably describe the predictive variable as a function of the explanatory variable and that more studies are needed with the inclusion of more explanatory variables. Table 1 presents the MAPE for all studies performed.

Culture	ADF Test	MAPE%	
		UCM	SARIMA
Rice	0,5424	18,7%	19,6%
Coffee	0,9968	31,4%	41,7%
Corn	0,9623	32,5%	42,3%
Soy	0,9933	25,7%	31,5%
Wheat	0,9978	20,8%	22,9%

Table 1: Tests of ADF and MAPE.

CONCLUSIONS

In this work, we verified that the price of Rice, Coffee, Corn, Soybean and Wheat crops are not stationary by the ADF test and showed a tendency of accentuated growth in some periods, as shown in figure 1. We also verified that the UCM Model presented lower MAPE than the Model SARIMAX for all time series, indicating that this may be more promising for the type of analysis performed in this work. However, the MAPE was greater than 18% in all cases, indicating that both models did not reasonably describe the predictive variable, making it necessary to explore other variables in order to improve the ability of these models to describe the predictive variable 'price' of commodities studied.

REFERENCES

1. N. K. K. Brintha et al.Use of Unobserved Components Model for Forecasting Non-stationary Time Series: A Case of Annual National Coconut Production in Sri Lanka. Disponível em: https://pdfs.semanticscholar.org/761e/0ac6a53020b7bee95751e489b16275397cdd.pdf. 2014.

2. J. Brownlee. How to Decompose Time Series Data into Trend and Seasonality. 2017. *Ma chine Learning Mastery*. https://machinelearningmastery.com/decompose-time-series-data-trend-seasonality/

3. J. E. C. Lima et al. Aplicação do Modelo SARIMA na Previsão de Demanda no Setor Calçadista. *Revista Multidisciplinar de Psicologia*. 2019. Disponível em: https://idonline.emnuvens.com.br/id/article/view/1875.

4. L. Rabelo. Princípios básicos para criar previsões de Séries Temporais. 2019. *Ensina.AI*. Disponível em: https://medium.com/ensina-ai/princ%C3%ADpios-b%C3%A1sicos-para-criar-previs%C3%B5es-de-s%C3%A9ries-temporais-e58c451a25b