# Some Comments On Fractionally Integration Processes Involving Two Agricultural *Commodities*

***Lucas Renato Trevisan***
***Sergio Adriani David***
University of São Paulo – Brazil

**Abstract**

This paper investigates time series of soybean and corn, which are two important Brazilian commodities. Long-range dependence or persistence is a behavior seen on times series and currently there is an increasing interest regarding the application of long memory concepts in areas such as economics and finances. A very know type of long memory model is named ARFIMA (Auto Regressive Fractionally Integrated Moving Average) which derives from the ARIMA (Auto Regressive Integrated Moving Average) model. The present work aim to analyze soybeans and corn time series to compose the spot price and forecast future prices for the aforementioned commodities. In order to test the better model for prices prediction, the ARIMA and ARFIMA models were compared. The comparison between the two models has shown that for prices forecasting, ARFIMA model has higher efficiency then ARIMA models.

**Keywords:** ARFIMA, ARIMA, forecast, commodities, prices

## Introduction

Time series analysis can be applied to a variety range of scientific fields and can be related to different sorts of data, information or phenomenon being observed. This analysis has shown to be very useful for many problems and the effectiveness of the models depends on what it is being used for. Jaynes (1982) emphasizes that there is no conflict in between analyzes methods such as Maximum-Entropy, Bayesian, Schuster, Autoregressive models and others. The distinction among this models are the uses of it, where each one has its better field of application.

Box & Jenkins (1970) stated studies that led to the method known as ARIMA (Autoregressive Integrated Moving Average), this models as centered in the idea that time series is a natural stochastic process which can be represented by a mathematic model.

Within its concepts and applications, Autoregressive models can be classified distinctly in some situations and can be presented as particular models, called: AR (Autoregressive) ARMA( Autoregressive and Moving Average), ARIMA (Auto Regressive Integrated Moving Average) and ARFIMA (Auto Regressive Fractionally Integrated Moving Average), as the case may be.

The applications of long memory processes were first introduced by Granger and Joyeux (1980) and later by Hosking (1981), and has become a successful tool for studies in areas such as hydrology, climatology, geophysics, economics and finances.

In this scope the agricultural commodities can be noted. According to Geman (2005), a commodity can be defined as a physical asset that presents standard features, with extensive trading in various locations, which can be transported and stored for a long period of time.

Marques et al (2006), states that the interesting of knowing future markets has increased, either applied on risks administrations, proper profits, or even to lead negotiations. Great progress has been made in understanding the links between government policy, interest rates, exchange rates, economic blocks, barriers to free trade and prices of the various commodities (agricultural, energy, gold and dollar).

In this work two Brazilian agricultural commodities are investigated, namely soybean and corn. The time series of prices for this commodities were obtained from CEPEA/USP (Center for Advanced Studies on Applied Economics/University of São Paulo). For the model development the time range for the spot price was from January 2009 to December 2013 and from January 2013 to December 2014 for the forecast process. The free software R was used to compose the ARIMA and ARFIMA model. The two models are tested to evaluate the reliability and effectiveness of the better model.

**ARIMA(p,d,q) models**

As a time series presents its values collected sequentially over time, it is expected to present a serial correlation in time (WERNER e RIBEIRO, 2003). This fact is reflected also as expected behavior of dependence between a current value and the previous values to this.

The models proposed by Box e Jenkins (1978) are widely known in sciences as ARIMA, which are mathematical models that intend to capture the autocorrelation behavior between the values of a time series and, once its behavior has been described, it is used to make predictions of future values in this series. If this correlation structure is well modeled, a good forecast can be provided (WERNER e RIBEIRO, 2003).

Fava (2000) presents that the ARIMA type models are the result of three distinct elements related to each other, which are the autoregressive component (AR), the moving averages component (MA), and the integration component (I). The result from the model in may have three parts, or only a subset of them. The values for the components of the ARIMA model are formally represented in the literature by the letters *p*, *q* and *d*. The *p* value refers to the AR component of the model, while the parameter *q* is related to the level of MA component and finally the parameter *d* shows the number of integrations on the model.

Granger and Newbold (1986) report that most economic series shows to be non-stationary, as in general, its average and variance does not remain constant over time.

According to the work of Sartoris (2008), ($y_t$) follows an ARIMA (p, d, q) where the letter (I) in the middle (and also the number *d*) refers to the integration order. That is, ($y_t$) is integrated of order *d*, and its *d*-th difference follows a combined autoregressive process (order *p*) and moving average (order *q*).

Lima, Goes and Ulysses (2007) points out that the ADF test sets the entire differentiation level (d) of the time series model.

As the value of the parameter *m* is such that $d = 0$, the model is estimated as an ARMA (p, q), since there are no differentiation. If the series is not stationary, the difference will be applied as often as needed to acquire a stationary series, in this case, for example, *d* may assume values equal to 1 or 2 or n, and in this case the model is estimated as an ARIMA (p , d, q).

**ARFIMA(p,d,q) models**

According to Franco e Reisen (2007), ARIMA (p,d,q) in many cases is classified as a general process called fractional differentiation when a non-integer value for the parameter d (degree of difference) is adopted. In these cases, new way of modeling is created and which can bring great benefits to the study of various fields, such as engineering, economics, chemistry, physics, etc. These models are known as ARFIMA.

The ARFIMA models can be described as a generalization of the ARIMA model, being responsible for capture and shaping processes with long serial dependence, which are popularly called long-term memory processes (Souza et al. 2010).

Franco e Reisen (2007) has also shown that the most important feature of ARFIMA model is the long dependence also named as long memory, found to *d* values in the range of 0.0 to 0.5. Another very important feature of the model is the small outbuilding, or short memory, which infers to *d* values between -0.5 and 0.0.

Regarding to ARFIMA model, Lima, Gois and Ulysses (2007) stated that, formally, the entire differentiation assumption of ARIMA model is arbitrary. Thus it can be said that it is possible to carry out modeling a temporal series considering that *d* can assume non-integral values.

Lima, Goes and Ulysses (2007) further describes that the fractional difference parameter *d* on ARFIMA, can be estimated by semi-parametric procedure proposed by Geweke and Hudack-Porter (1983).

In addition Lima, the fractional differentiation can be performed using binomial expansion of the form proposed by Diebold and Rudebusch (1989):

**Estimation, verification and prediction**

To obtain an ARIMA/ARFIMA model is required to identify the value for the coefficients $\theta_P$ and $\varphi_p$. The autocorrelation function (ACF) and the partial autocorrelation function (PACF) are responsible for explicit which class of model the time series has, in other words, through the behavior of ACF and PACF functions is possible to identify the parameter AR(p) and MA(q) for the model. This functions are merely the correlation in between a present value and its past values.

The correlation between a current value $Y_t$ in the time series Y and its previous, named $Y_{t-1,}$ is known as the autocorrelation of the series. In the same sense if the value of $Y_{t\ -1}$ correlates with its past value $Y_{t-2}$, it is expected that $Y_t$ also has a correlation with $Y_{t-2.}$ The resulting correlation in between $Y_t$ and $Y_{t-1}$ is called frequently as lag 1, and following the same idea the correlation between $Y_t$ and $Y_{t-2}$ is called lag 2. Generalizing, the correlation between $Y_t$ and $Y_{t-n}$ is the lag n of the time series.

The relationship in between a lag n and its next lag n+1, is such that lag n+1 should be exactly the squared value of lag n. This implies that the correlation of a lag n propagates throughout the higher-order lags.

An appropriate model should behavior in a way that the residuals of ACF and PACF has no statistical significant values implying that the values of the residuals does not influence in any value of the model. Another evaluation taken to the final model refers to its Bayesian information criterion value (BIC). Another way of evaluating the efficiency of a prediction model is looking through its errors values.

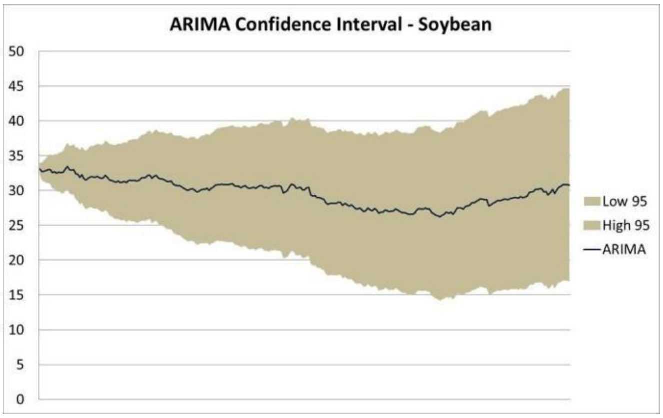
Once the specific values of parameter are known, through the observation of ACF and PACF behavior, the forecast for the ARIMA and ARFIMA model can be taken and the coefficients for the models can be calculated. The results should lead to a linear equations composed by lags of dependent values and forecast error. In a general way, the process consists on prediction a desired number of ahead observations that are taken considering the past values of the time series.

**ARIMA and ARFIMA prediction and comparison**

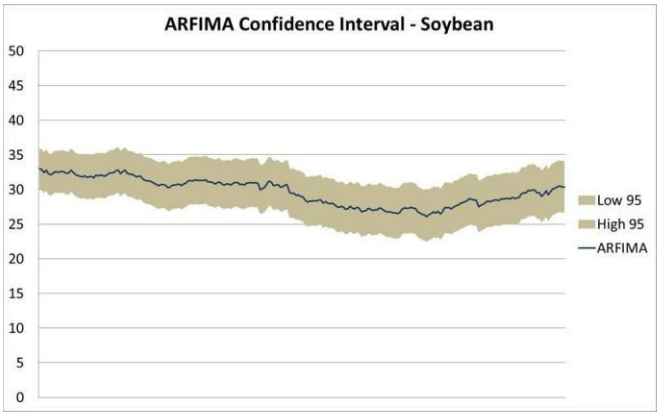
The ACF and PACF plots are obtained from the original time series, and then evaluated about its stationarity. For the cases where the original series is not stationary, the first difference is taken. After differencing, the series is analyzed again and if it is still non-stationary, the second difference is taken, and so on.

Figure 1 and 2 show the ARIMA and ARFIMA model for soybean, as well as the confidence interval for each model. A comparison between these two models is taken in order to evaluate the best option for a future prediction for each commodity. Table 1 to 3 shows the evaluation BIC values and the errors calculates to each model.

**a-**




**b-**



**c-**

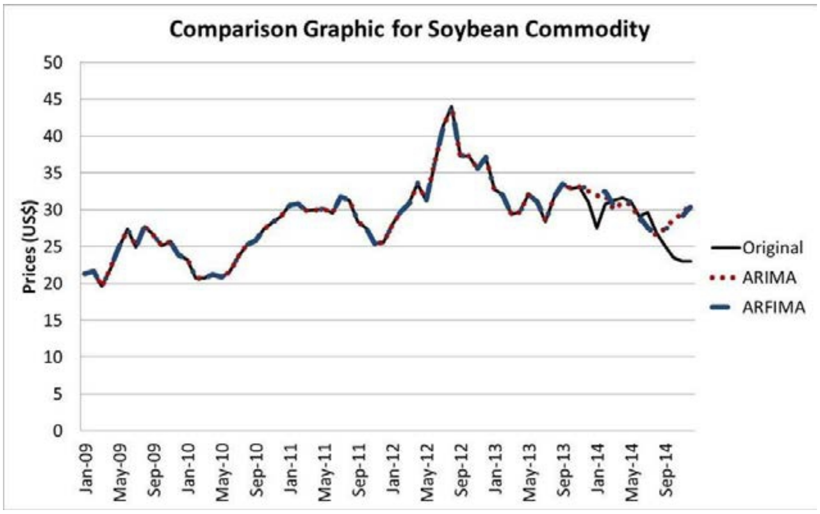
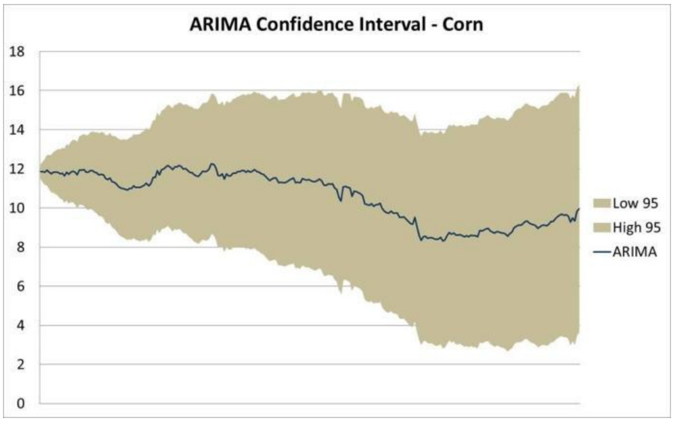
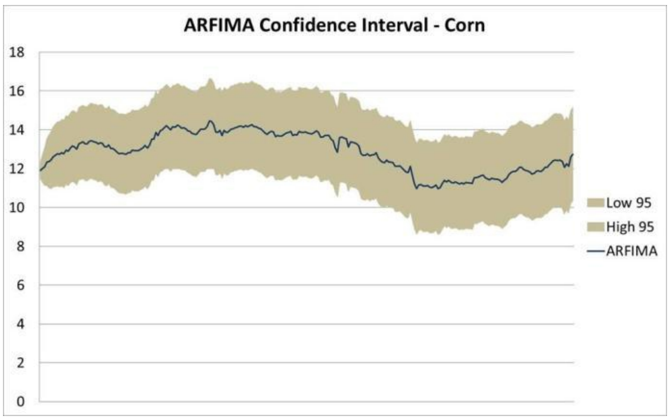



Figure 1. Soybean commodity: (a) ARIMA, (b) ARFIMA, (c) Comparison

**a-**




**b-**



**c-**

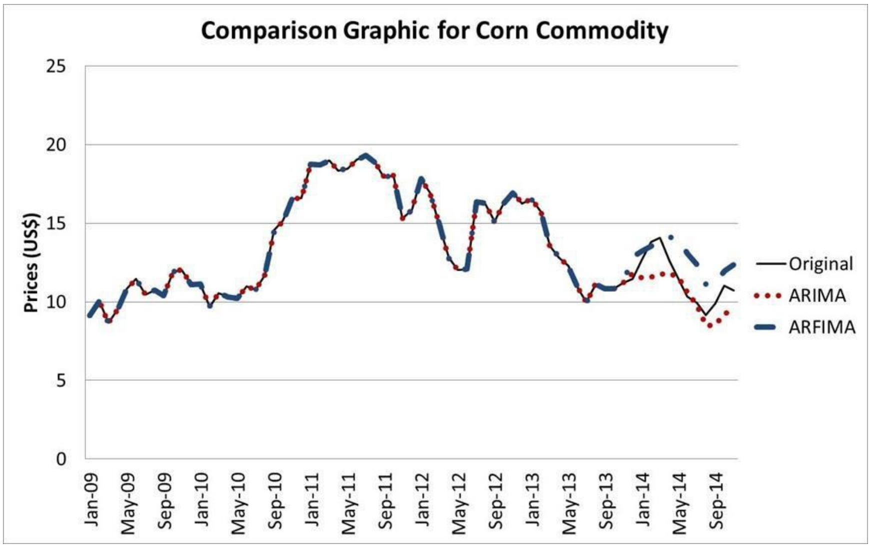



Figure 2. Soybean commodity: (a) ARIMA, (b) ARFIMA, (c) Comparison

Table 1. BIC values - ARIMA

| COMMODITY | PAPER SIZE | BIC |
|---|---|---|
| SOYBEAN | STL+ARIMA(3,2,4) | 1888,31 |
| CORN | STL+ARIMA(2,1,0) | -1185,048 |

Table 2. BIC values -ARFIMA

| COMMODITY | PAPER SIZE | BIC VALUES |
|---|---|---|
| *SOYBEAN* | STL+ARFIMA(0,0.5,5) | 3061,716 |
| *CORN* | STL+ARFIMA(0,0.5,5) | 3137,638 |

**Table3.** Results and comparison of methods

| Commodity | ARIMA | | | | | ARFIMA | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | (p,d, q) | E | MAE | ACF1 | MASE | (p,d, q) | E | MAE | ACF1 | MASE |
| Soybean | (3,2, 4) | 8,3764 | 2,3525 | 0,9756 | 8,1392 | (0,0.5 ,5) | 8,1496 | 2,2859 | 0,9751 | 7,9091 |
| Corn | (2,1, 0) | 8,6800 | 1,1079 | 0,9797 | 9,0139 | (0,0.5 ,5) | 13,4429 | 1,6095 | 0,9846 | 13,0946 |

## Conclusion

Fractionally integrated processes motivated an increasing interest on the application in economics and finance. One important characteristic of fractionally integrated processes is to allow more flexibility than the extreme assumption of a unit root. The real advantage of fractional models may well be in terms of representing relationships between variables and the testing of forms of fractional cointegration. In this work, the ARIMA and ARFIMA models were applied in agricultural commodities using the R software for spot price composition and future price prediction. The results show that the ARFIMA has a better performance overall for the future prices forecasting when compared with ARIMA model.

## References:


Jaynes, E.T., On the rationale of maximum-entropy methods. Proceedings of the IEEE, Vol. 70, No 9, September, 1982, 939-952.

Granger, C.W.J.; Joyeux, R.. An introduction to long memory time series models and fractional differencing. Journal of time series analysis, 1, pp. 15-39, 1980.

Hosking, J.R.M..Fractional Differencing, Biometrika, 68, pp.165-176, 1981.

Geman H. Commodities and Commodity Derivatives. 1ed New York: John Wiley & Sons. Led, 2005.

Marques, P. V., Mello, P. C. Mercados futuros de commodities agropecuárias: exemplos e aplicações para os mercados brasileiros. São Paulo: Bolsa de Mercadorias & Futuros, 2006.

Werner, L.; Ribeiro, J. L. D. Previsão de demanda: uma aplicação dos modelos Box-Jenkins na área de assistência técnica de computadores pessoais. *Gestão e Produção*, v. 10, n. 1, p. 47-67, 2003.

Box, G. P., Jenkins, G. M. Time Series Analysis: Forecasting and Control. San Fracisco: Holden Day, 1978.

FAVA, V. L. Manual de econometria. In: VASCONCELOS, M. A. S.; ALVES, D. São Paulo: Editora Atlas, 2000.

Franco, G.C. e Reisen, V.A. (2007). Bootstrap approaches and confidence intervals for stationary and non-stationary long-range dependence processes. Physica A, 375, 546- 562.

Granger C W J, Newbold P 1986 Forecasting Economic Time Series, 2nd edn. Academic Press, Orlando, FL.

LIMA, R. C., GÓIS, M. R., ULISES, C. Previsão de preços futures de commodities agrícolas com diferenciação inteira e fracionária, e erros heteroscedásticos. Revista Economia e Sociologia Rural vol.45, n.3, Brasilia July/Sept. 2007.

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.