Journal of Environmental and Agricultural Studies
Website: www.al-kindipublisher.com/index.php/jeas/index
Al-Kindi Center for Research and Development

JAES

Volatility in Agricultural Commodities Prices, Case of Sugar Prices: Evidence of ARIMA-GARCH Family Models

CHELLAI Fatih
Faculty of Economics Commerce and Management. Ferhat Abbas, University, Setif, Algeria.
Corresponding Author: CHELLAI Fatih, E-mail: fatih.chellai@univ-setif.dz

ARTICLE INFORMATION

Received: April 02, 2020
Accepted: June 10, 2020
Volume: 1
Issue: 1

KEYWORDS

Volatility, Sugar prices, Heteroscedasticity; ARIMA-GARCH

ABSTRACT

This article focused on analyzing the volatility in agricultural commodities prices, where the class of ARMA models with ARCH errors were used. Maximum Likelihood and Least Squares estimates of the parameters of the model and their covariance matrices are noted and incorporated into techniques for the model building based upon the application of the usual Box-Jenkins methodology of identification, estimation, and diagnostic checking to the ARMA equation, the ARCH equation, and the full model. The techniques are applied to sugar prices daily time series over the period (1962-2020). It is seen that ARIMA (4,1,0)-GARCH(1,2) fits well the data among other competitive models.

1. Introduction

In financial markets of agricultural products, such: Wheat, Coffee, Sugar,… the prices are the engine indicator; furthermore, these prices are the best indicator for measuring and knowing the state of world, regional and domestic markets, and they are very important in helping to elaborate economic strategies as well as making appropriate and timely decisions regarding production (for the 'farmer), marketing as well as Consumption (especially for countries where the ratio of agricultural products to total imports is very high)(Apergis & Rezitis, 2011). This is why the phenomenon of hyper-volatility of these prices is one of the serious problems which must be studied.

An abundant and variety of studies have been carried on agricultural prices-analysis and forecasting. Heady and Kaldor(1954) firstly focalized on expectation and errors in forecasting. Acharya and Agarwal (1994) analyzed agricultural prices and related policies. Xiong *et al* (2015) presented a new combination method for interval forecasting of agricultural commodities prices.

We work on time series data, which are a branch of econometrics whose object is the study of variables over time. Among its main objectives is the determination of trends within these series as well as the stability of values (and their variation) over time. In particular, linear models (mainly AR and MA, for Auto-Regressive and Moving Average), Box-Jenkins, (1976), distinguish conditional models (notably ARCH, for Auto-Regressive Conditional Heteroskedasticity) (Engel, 1982). Unlike traditional econometrics, the purpose of time series analysis is not to relate variables to one another, but to focus on the "dynamics" of a variable. In this study, we deal with ARIMA-ARCH family models to estimate and forecast the sugar prices over the period (1962M12-2020M2), by using the ARIMA models (called Box-Jenkins Approach) developed by, Box-Jenkins, (1976), and autoregressive conditional heteroscedastic (ARCH) first introduced by Engel (1982).

The rest of the article is divided as section (2) presents the statistical methods used (ARIMA and GARCH models), in section (3) we showed the dynamic of sugar prices over the period (1962-2020), the estimation results of fitted models, discussion results and forecasting of future prices for 30 periods, finally the section (4) concludes and summarize the study.

2. Methods

2.1. ARIMA Models presentation

2.1.1. Auto-Regressive Model, AR (p)

The conditional approach in Equation (1) provides a decomposition prediction error, according to which:

$$\boldsymbol{y_t} = \mathrm{E}(\boldsymbol{y_t} \setminus \boldsymbol{y_{t-p}}) + \boldsymbol{\epsilon_t} \Leftrightarrow \boldsymbol{y_t} = \sum_{i=1}^{p} \boldsymbol{\beta_i}\, \boldsymbol{y_{t-i}} + \boldsymbol{\epsilon_t} \tag{2}$$

Where $\mathbf{E}(\boldsymbol{y_t} \setminus \boldsymbol{y_{t-p}})$, is the component of $\boldsymbol{y_t}$, that can give rise to a forecast, when the history of the process, $\boldsymbol{y_{t-1}}, \boldsymbol{y_{t-2}} \dots, \boldsymbol{y_0}$ are known. And $\boldsymbol{\epsilon_t}$, represents unpredictable information. We suppose, $\boldsymbol{\epsilon_t} \sim WN\ (0, \sigma^2)$, is a white noise process. The equation (2) represents an autoregressive model (AR) of order $\boldsymbol{p}$. As an example an autoregressive process of order 1, AR (1) is defined:

$$\boldsymbol{y_t} = \boldsymbol{c} + \boldsymbol{\alpha y_{t-1}} + \boldsymbol{\epsilon_t} \tag{3}$$

The value $\boldsymbol{y_t}$ depends only on its predecessor. Its properties are functions of $\boldsymbol{\alpha}$ which is a factor of inertia. Autoregressive processes AR(p) assume that each observation $\boldsymbol{y_t}$ can be predicted by the weighted sum of a set of previous observations $\boldsymbol{y_{t-1}}, \boldsymbol{y_{t-2}}, \dots \boldsymbol{y_{t-p}}$, plus a random error term. the other type of process of the Box-Jenkins approach is Moving Average, MA(q).

2.1.2. Moving-Average process MA (q)

The moving average processes assume that each observation $\boldsymbol{y_t}$ is a function of the errors in the preceding observations, $\boldsymbol{\epsilon_{t-1}}, \boldsymbol{\epsilon_{t-2}}, \dots, \boldsymbol{\epsilon_{t-p}}$, plus its error. A moving average process is given as:

$$\boldsymbol{y_t} = \sum_{i=1}^{q} \boldsymbol{\theta_i}\, \boldsymbol{\varepsilon_{t-i}} + \boldsymbol{c} \tag{4}$$

The combination of the two models, AR (p) in equation (3) and MA(q) in equation (4) is an ARMA(p, q) process; which is the most popular models of the Box Jenkins for its flexibility and suitability for various data types. The model is designed as follow:

$$\mathbf{ARMA(p,q)} \quad : \quad \sum_{i=1}^{p} \boldsymbol{\beta_i}\, \boldsymbol{y_{t-i}} = \sum_{i=1}^{q} \boldsymbol{\theta_i}\, \boldsymbol{\varepsilon_{t-i}} \tag{5}$$

With: $\boldsymbol{\beta_i}\,(\boldsymbol{i} = \mathbf{1}, \dots, \boldsymbol{p}),\quad \boldsymbol{\theta_i}\,(\boldsymbol{i} = \mathbf{1}, \dots, \boldsymbol{q}) \in \mathbb{R} \quad , \quad \boldsymbol{\epsilon_t} \backsim \boldsymbol{WN}(\mathbf{0}, \boldsymbol{\delta_\varepsilon})$.

The time series $\boldsymbol{y_t}$ must be stationary to be fitted by an ARMA model. We take the case of weak stationary, and we put its definition:

Definition: a time process $\boldsymbol{y_t}$ with real values and discrete-time $\boldsymbol{y_1}, \boldsymbol{y_2}, \dots \boldsymbol{y_t}$It is stationary in the weak sense (or "second-order", or "in covariance") if:

- $\mathrm{E}(\boldsymbol{y_i}) = \boldsymbol{\mu} \qquad \forall\ \boldsymbol{i} = \mathbf{1}, \dots, \boldsymbol{t}.$
- $\mathrm{Var}(\boldsymbol{y_i}) = \boldsymbol{\sigma^2} \neq \infty \qquad \forall\ \boldsymbol{i} = \mathbf{1}, \dots, \boldsymbol{t}.$
- $\mathrm{Cov}(\boldsymbol{y_i}, \boldsymbol{y_{i-k}}) = \boldsymbol{f(k)} = \boldsymbol{\rho_k} \qquad \forall\ \boldsymbol{i} = \mathbf{1}, \dots, \boldsymbol{t}.$

When one or more stationary conditions are not met, the series is said to be non-stationary. This term, however, covers many types of non-stationary, (non-stationary in trend, stochastically non-stationary,), we focused on the latter. Thus, if $\boldsymbol{y_t}$ is a stochastically non-stationary, a difference stationary technique should be applied. Consequently, a series is stationary in difference if the series obtained by differentiating the values of the original series is stationary. Generally, we used the KPSS test, Kwiatkowski, et *al,* (1992), and Leybourne & McCabe test, (1994).

The difference operator is given by: $\Delta(y_t) = y_t - y_{t-1}$. if the series is differentiated **d** times, we say that it is integrated of order **I (d).** The process will be noted as $\boldsymbol{ARIMA(p, d, q)}$, defined by the equation:

$$\boldsymbol{\beta}(\boldsymbol{L})(\mathbf{1}-\boldsymbol{L})^{\boldsymbol{d}}\,\boldsymbol{y}_{\boldsymbol{t}} = \boldsymbol{\theta}(\boldsymbol{L})\,\boldsymbol{\varepsilon}_{\boldsymbol{t}} \qquad (\mathbf{6})$$

With, L: is the lag *operator* (L) or backshift *operator* (B); If the time series $\boldsymbol{X_t} = (\mathbf{1}-\boldsymbol{L})^{\boldsymbol{d}}\,\boldsymbol{y_t}$ is stationary, then, estimating an $\mathbf{ARIMA}(\boldsymbol{p},\boldsymbol{d},\boldsymbol{q})$ process on $\boldsymbol{y_t}$ is equivalent to estimating an **ARMA (p, q)** process on $\boldsymbol{X_t}$**.**

2.1.3. Estimation of ARIMA models

Box and Jenkins (1970) proposed a prediction technique for a univariate series that is based on the notion of the ARIMA process. This technique has three stages: identification, estimation, and verification. The *first step* is to identify the ARIMA model (p, d, q) that could spawn the series. It consists, first of all, in transforming the series to make it stationary (the number of differentiations determines the order of integration: d), and then to identify the ARMA model (p, q) of the series transformed with the correlogram and partial correlogram. The graph of autocorrelation (correlogram) and partial autocorrelation coefficients (partial correlogram) give information on the order of the ARMA model. Thus, if we observe that the first two autocorrelation coefficients are significant, we will identify the following model: MA (2). The *second step* is to estimate the ARIMA model using a non-linear method (nonlinear least squares or *maximum likelihood*). These methods are applied using the degrees p, d and q found in the identification step.

Generally, we use the *Likelihood Maximum method*; by consider that the errors $\boldsymbol{\varepsilon_t}$ follow a normal distribution, $\boldsymbol{N}(\mathbf{0},\boldsymbol{\sigma}_{\boldsymbol{\varepsilon}}^{\mathbf{2}})$**.** The log-likelihood function of an $\boldsymbol{ARMA}(\boldsymbol{p},\boldsymbol{q})$ process is defined as:

$$\mathbf{Log}\,\boldsymbol{L_t} = -\frac{\boldsymbol{T}}{\mathbf{2}}\,\mathbf{log2}\boldsymbol{\pi} -= -\frac{\boldsymbol{T}}{\mathbf{2}}\,\mathbf{log}\boldsymbol{\sigma}_{\boldsymbol{\varepsilon}}^{\mathbf{2}} - \frac{\mathbf{1}}{\mathbf{2}}\mathbf{log}(\boldsymbol{det}[\boldsymbol{\psi}'\boldsymbol{\psi}]) - \frac{\boldsymbol{\varpi}(\boldsymbol{\beta},\boldsymbol{\theta})}{\mathbf{2}\boldsymbol{\sigma}_{\boldsymbol{\varepsilon}}^{\mathbf{2}}} \qquad (\mathbf{6})$$

With:

- **T:** number of observations**,**
- $\boldsymbol{\psi}$ a matrix of $(\boldsymbol{p}+\boldsymbol{q}+\boldsymbol{T},\boldsymbol{p}+\boldsymbol{q})$ dimensions, dependent of $\boldsymbol{\beta_i}\,(\boldsymbol{i}=\mathbf{1},\ldots,\boldsymbol{p})$ and $\boldsymbol{\theta_i}\,(\boldsymbol{i}=\mathbf{1},\ldots,\boldsymbol{q})$**,**
- $\boldsymbol{\varpi}(\boldsymbol{\beta},\boldsymbol{\varphi}) = \sum_{t=-\infty}^{T}\left(\mathbb{E}\left[\boldsymbol{\varepsilon_t}\setminus\boldsymbol{X_t},\,\boldsymbol{\beta_i},\boldsymbol{\theta_j},\boldsymbol{\sigma}_{\boldsymbol{\varepsilon}}^{\mathbf{2}}\right]\right)^{\mathbf{2}}$, with: $\boldsymbol{i}=\mathbf{1},\ldots,\boldsymbol{p}\,;\ \boldsymbol{j}=\mathbf{1},\ldots,\boldsymbol{q}$**.**

The *third step* is to check whether the estimated model reproduces the model that generated the data. For this purpose, the residuals obtained from the estimated model are used to check whether they behave like white noise errors using a "portmanteau" test (a global test that makes it possible to test the hypothesis of independence of residues). The common tests are based on residuals analysis for normality, and autocorrelation: Box and Pierce (1970), Ljung and Box (1978), Durbin and Watson (1950, 1951). Homoskedasticity: Test of Breusch and Pegan (1979), ARCH Test, Engel (1982). The last point under this step is the prediction of future values of y_t by the selected model.

2.2. GARCH models

In this section, we briefly introduce the conditional heteroscedastic family models including the autoregressive conditional heteroscedastic (ARCH) model of Engle (1982), the generalized ARCH (GARCH) model of Bollerslev (1986). Engel (1982) developed the ARCH models to allow the variance of a time series to depend on the set of available information, and in a particular time. This class of models aims to overcome the inadequacy of ARMA models; several phenomena are characterized by variable volatility and asymmetry that cannot be taken by the ARMA novelizations. The fact that the understanding and modeling of volatility a major priority in the economic, social, and political.

In $ARCH(q)$ model, we define ε_t as a process that established:

$$\begin{cases} E\left(\varepsilon_t \setminus \underline{\varepsilon_{t-1}}\right) = 0 \\ V\left(\varepsilon_t \setminus \underline{\varepsilon_{t-1}}\right) = \boldsymbol{\sigma}_{\boldsymbol{t}}^{\mathbf{2}} \end{cases} \qquad (1)$$

With: $\underline{\varepsilon_{t-1}} = (\varepsilon_{t-1},\varepsilon_{t-2},\varepsilon_{t-3},\ldots)$, and $\boldsymbol{\sigma}_{\boldsymbol{t}}^{\mathbf{2}}$ is the conditional variance of the ε_t process. It's so clear that this variance can vary over time unlike in ARMA models. $ARCH(q)$ models are based on a quadratic parameterization of the conditional variance $\boldsymbol{\sigma}_{\boldsymbol{t}}^{\mathbf{2}}$**.** We define an $ARCH(q)$ model as:

$$\boldsymbol{\sigma}_{\boldsymbol{t}}^{\mathbf{2}} = \boldsymbol{\beta}_{\mathbf{0}} + \sum_{\boldsymbol{i}=\mathbf{1}}^{\boldsymbol{q}} \boldsymbol{\beta_i}\boldsymbol{\varepsilon}_{\boldsymbol{t}-\boldsymbol{i}}^{\mathbf{2}} \qquad (\mathbf{2})$$

Where: $\boldsymbol{\beta_0} > 0$ and $\boldsymbol{\beta_i} \geq \mathbf{0}, \forall\, \boldsymbol{i}$.

General Autoregressive Conditional Heteroscedastic $GARCH(p,q)$ models

A generalization of the ARCH model had been proposed by **Bollerserv (1986),** a $GARCH(p,q)$ model is an extension allowed the introduction of lagged values of conditional variance $\boldsymbol{\sigma_t^2}$, a simple way to define a $GARCH(p,q)$ process is:

$$\boldsymbol{\sigma_t^2 = \beta_0 + \sum_{i=1}^{q} \beta_i \varepsilon_{t-i}^2 + \sum_{j=1}^{p} \alpha_i \sigma_{t-i}^2} \quad (\mathbf{3})$$

Where: $\boldsymbol{\beta_0} > 0$ and $\boldsymbol{\beta_i} \geq \mathbf{0}, \boldsymbol{\alpha_i} \geq \mathbf{0}, \forall\, \boldsymbol{i}. \forall\, \boldsymbol{j}$. usually, these constraints of parameters are to guarantee the positivity of conditional variance.

3. Results and Discussion

3.1 Data Description of Sugar production and Prices

World sugar production is estimated at around 180 million tonnes. Production is mainly dominated in order of magnitude by Brazil, India, the European Union, and China, see Figure (1) below. Among the diversity of sugars present on the market, the production of cane sugar represents 80% and comes mainly from regions such as Asia, South America, and even Central America. Sugar beet production continues to be controlled by the European Union, followed by Russia and the United States.

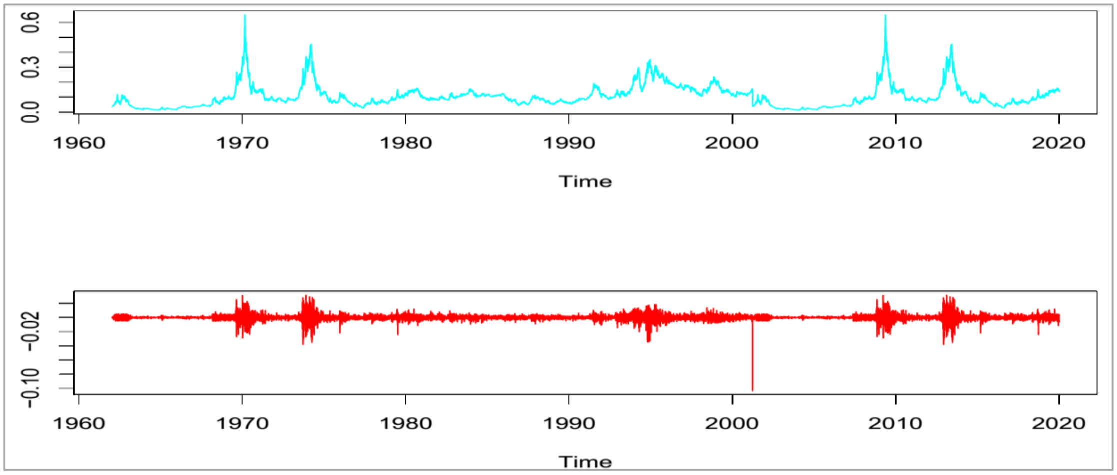
Figure.1 Geographical distribution of the World's biggest sugar-producing regions**.**

Source: https://commodity.com/soft-agricultural/sugar/ , retrevied (2020/02/15).

In this very liberalized sector, it will also be necessary to count on the Asian giant of agribusiness which took hold in March 2017 of record quantity via the American futures market. Based in Singapore, the company has quickly established itself as a leader in trade and trading in this market.

The daily sugar prices are showed in Figure2, the range prices over the whole period are (0.0125-0.652) $ per kilogram, with 0.11 $/kg as average prices. The filtered series (*in red*) is stationary on average, but not stationary in variance (it is very volatile: we can even distinguish groups of large variations or small variations on the series): which justifies the use of heteroscedastic modeling to study our filtered series.

Figure.2 Graphical representation of the sugar time series (in blue) and filtered/differentiated series of order 1 (in red) over the period(1962-2020).



Source: Plotted using R program. Data source: https://www.macrotrends.net/2537/sugar-prices-historical-chart-data, (free download).

For the identification $ARIMA$ process, we used the augmented Dickey-Fuller (ADF), information criterions tests**,** the raw time series is not stationary, and that is shown on Figure (1). The prices is integrated of first-order(*i.e.*) $(\mathbf{1}-\boldsymbol{L})\boldsymbol{y_t} = \Delta(\boldsymbol{y_t}) = \boldsymbol{y_t} - \boldsymbol{y_{t-1}},\ \ \boldsymbol{t} \coloneqq 1,2, \dots, 14312.$ The best $ARIMA(p, d, q)$ models were selected through the criteria (LL, AIC, BIC…etc) lead us to select the model in Table (2) to fit the dynamics of the sugar prices time series, the full results are in the Table (2) (*green lines*) .

3.2 ARIMA-GARCH estimation results

We retain the $ARIMA(4,1,1)$ process without drift which models the filtered series better than the other candidate models.

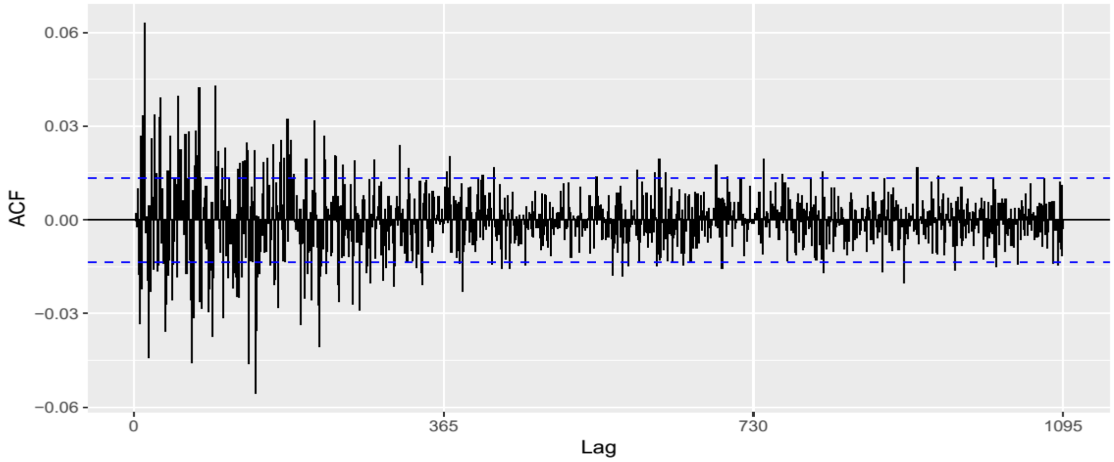
Table2: Model fitting for daily Sugar prices over the period (1962-2020).

Variable	**Coefficient**	90% CI Low	90% CI High	95% CI Low	95% CI High	99% CI Low	99% CI High
AR(1)	**-0.760570**	-0.797045	-0.724094	-0.804034	-0.717105	-0.817694	-0.703445
AR(2)	**0.058648**	0.052620	0.064675	0.051465	0.065830	0.049208	0.068087
AR(3)	**0.046676**	0.040626	0.052725	0.039467	0.053885	0.037202	0.056150
AR(4)	**0.059558**	0.054644	0.064473	0.053702	0.065414	0.051861	0.067255
MA(1)	**0.816659**	0.780297	0.853021	0.773330	0.859988	0.759712	0.873606

Source: Estimation of the mean equation of sugar prices.

We see that the variation of sugar prices negatively depends on the previous period, and positively for the periods $(t-2, t-3, and\ t-4)$. We note also that the previous innovation (though $MA(1) = 0.816$) affect positively the current variation of the sugar prices. After the residuals analysis of the mean equation ($ARIMA(4,1,1)$, the ARCH effects are attached in the residual time series, graphically, this so showed in the auto-correlation function of the residuals and squared residuals. (*See Figure3 below*); furthermore, The Q statistic of Ljung-Box indicates many terms statistically different from; Which leads us to assume the presence of ARCH effects.

Figure.3. Auto-correlation functions of the squared Residuals of $ARIMA(4,1,1)$ model.



Source: Plotting from R program.

We applied a test to confirm the presence of ARCH effects on the estimated ARIMA model, as is shown in the table below, the critical probability being < 0.05 for an ARCH (6), we accept the hypothesis of the presence of ARCH effects (for lags: 6, 7..., the parameters of the autoregressive terms were found also to be insignificant). (see Figure 3).

Table.3 ARCH-LM test for heteroscedasticity.

F-statistic	1053.282	Prob. F(6,14243)	0.0000
Obs*R-squared	4379.563	Prob. Chi-Square(6)	0.0000

After integration of the GARCH effects in our model, the optimal model for the conditional variance was a $GARCH(1,2)$; the estimation results of the mean equation have been highly changed (so the ARMA model was not stable at the first time); the coefficients of the equation of the mean (except the MA(1) component) are statistically significant; they are therefore far from the model presented in **table 2**

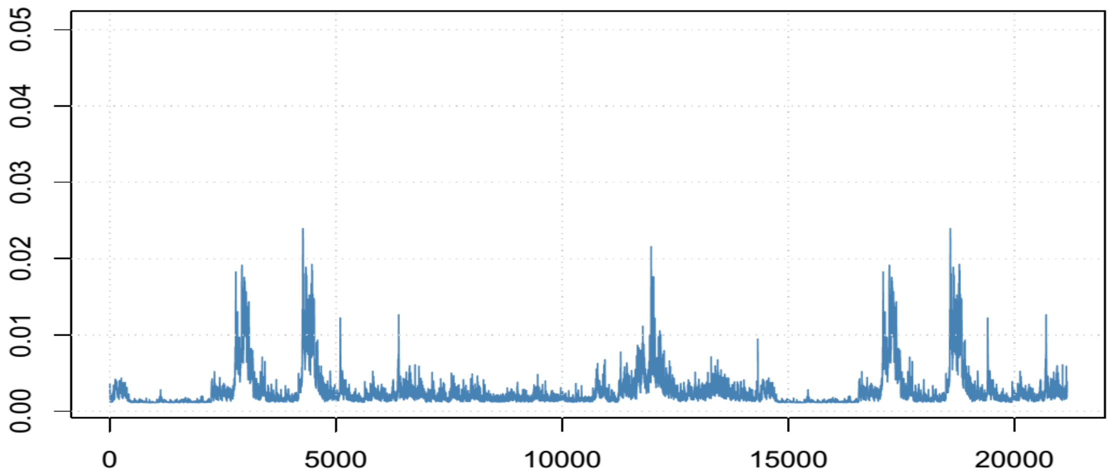
Table.4 Estimation of ARIMA-GARCH (1,2) model

		90% CI		95% CI		99% CI	
Variable	**Coefficient**	Low	High	Low	High	Low	High
AR(1)	**0.033622**	0.019067	0.048177	0.016278	0.050966	0.010827	0.056417
AR(2)	**-0.022533**	-0.036517	-0.008548	-0.039197	-0.005869	-0.044434	-0.000631
AR(3)	**0.023147**	0.009143	0.037152	0.006459	0.039835	0.001215	0.045080
AR(4)	**0.019966**	0.006300	0.033633	0.003682	0.036251	-0.001436	0.041369
C	**5.74E-09**	3.99E-09	7.49E-09	3.65E-09	7.83E-09	3.00E-09	8.48E-09
ARCH(1)	**0.092198**	0.083761	0.100636	0.082144	0.102252	0.078984	0.105412
GARCH(1)	**0.544610**	0.423082	0.666137	0.399798	0.689422	0.354286	0.734934
GARCH(2)	**0.368461**	0.253393	0.483528	0.231347	0.505575	0.188253	0.548668

Source: Estimation Results from R program. *Note*: the terms $AR(1) \dots AR(4), MA(1)$ – *in green*- are the **mean equation**, β_0, β_1 and α_1 ,α_2 are, *respectively*, the constant, ARCH(1) and GARCH(2) compounds in the GARCH equation models (see Equation 4).

Moreover, we have the sum of the coefficients ARCH (1) and GARCH (1) is very close to $1;(0.092 + 0.544 + 0.368)$. This testifies to a phenomenon of persistence in the conditional variance; this sum also confirms the stability of the model estimated. According to Chan (2010, persistence of volatility occurs when: $\sum_{i=1}^{q} \boldsymbol{\beta_i} + \sum_{j=1}^{p} \boldsymbol{\alpha_i} = 1$, (*see equation 3*) and thus $\boldsymbol{\sigma_t^2}$ is non-stationary process. This is also called IGARCH (Integrated GARCH). Under this scenario, unconditional variance becomes infinite.

Figure3. The conditional standard deviation of the estimated model



Source: Graphs using R program.

Based on the selected models, and through the theoretical part of this study, the almost objective of the Box-Jenkins method is to forecast the future dynamic of the times series. Among the candidates' models, the best model selected is an $ARIMA\ (4,1,0) - GARCH(1,1)$, the forecast equation according to this model is :

$$\begin{cases} \boldsymbol{\hat{y}_t = 0.034\, y_{t-1} - 0.023\, y_{t-2} + 0.022\, y_{t-3} + 0.019\, y_{t-4}} \\ \boldsymbol{\hat{\sigma}_t^2} = 5.74\mathrm{E} - 09 + 0.092\boldsymbol{\hat{\varepsilon}_{t-1}^2} + 0.544\, \boldsymbol{\sigma_{t-1}^2} + 0.368\ \boldsymbol{\sigma_{t-2}^2} \end{cases}$$

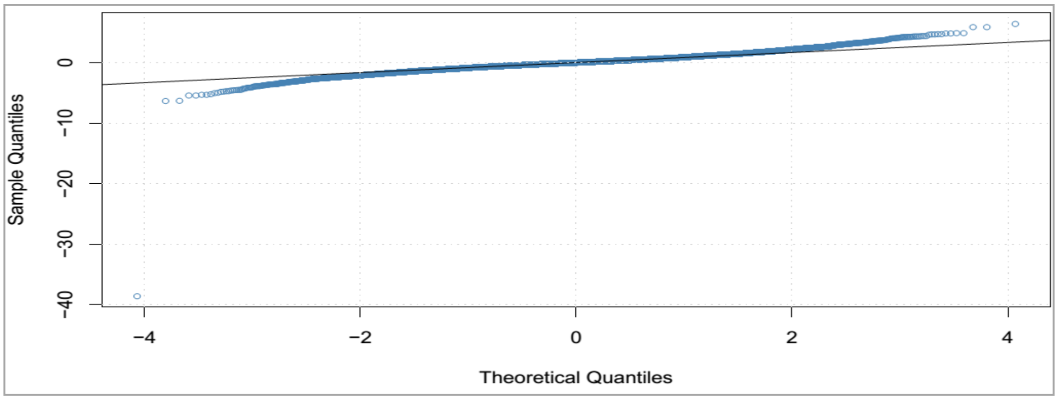
The coefficients of the equation of the mean (except the $\boldsymbol{MA(1)}$ coefficient) are relatively stable; they are therefore very close to those of the model estimated without GARCH compounds; as Bollerslev (1986) demonstrates with an example, the virtue of this approach is that a GARCH model with a small number of terms appears to perform as well as or better than an ARCH model with many; in our case, at least five (5) arch levels were statistically significant, but we prefer to work with a **GARCH(1,1)** which overcome the ARCH models.

3.3 Validation of ARIMA-GARCH fitted model

For the validation step, the following three aspects of the residuals from the fitted GARCH model should be tested:

The standardized residuals from the GARCH model should approach normal distribution (*if we assumed the conditional distribution of error terms as the normal distribution*). For this point, we can use a Shapiro-Wilk (S-W) test and the Jarque-Bera normality test. Histogram of the residuals and ***quantile-quantile*** (***Q-Q***) ***plots*** are also a good visual tool to check normality.

Figure4. *Quantile-Quantile* (*Q-Q*) *plots of fitted model residuals*

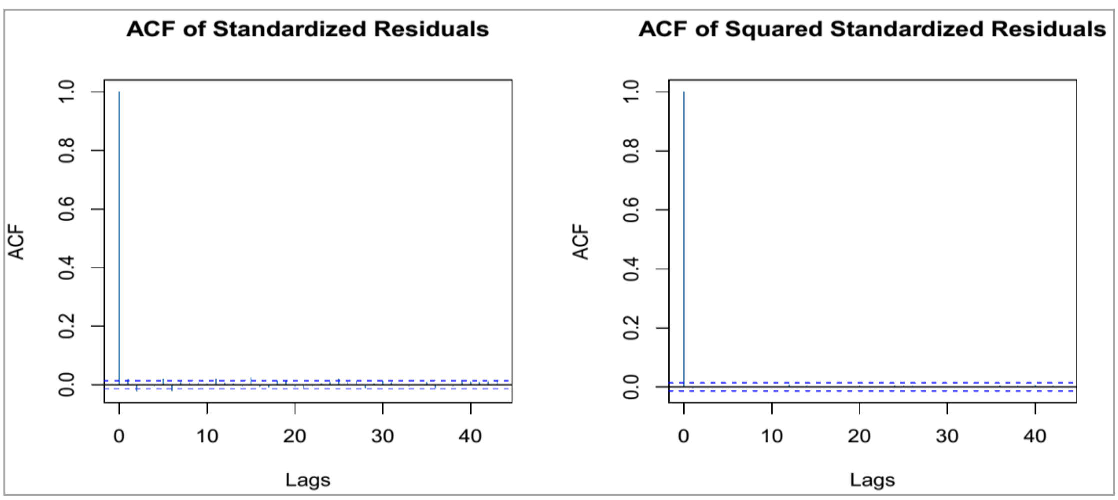


Source: Graphs using R program.

For our case, the residuals don't follow a normal distribution, we had the p-values of the Shapiro-Wilk (S-W) test and Jarque-Bera are: $p - value = 0.000$; both below the risque $\alpha = 0.05$; furthermore, we see this normality rejection from the *Q-Q* plot (***Figure 4*** *above*).

A *second step* for validation of GARCH modeling is we have to check that the standardized squared residuals should not be auto-correlated. We can use Box and Pierce (1970) and Ljung and Box (1978) statistics tests for this purpose. We see from the ACF of squared standardized residuals (Figure 5 below), that are not auto-correlated because all auto-correlation terms are inside the confidence intervals; the same result was found for ACF of standardized residuals.

Figure5. Autocorrelation functions of standardized and squared standardized residuals



Source: plotting results by R program

A third step for validating the GARCH model is to run the ARCH-LM test on the residuals can also be conducted to check for remaining ARCH effects in the residuals; for our estimating results, there are no ARCH effects (the p-value of L6M test is: 0.99), so we reject the presence of heteroskedasticity.

3.4 Forecasting results

Table 3: Sugar prices Forecasting over the period (2020M03D01-2020M03D30)

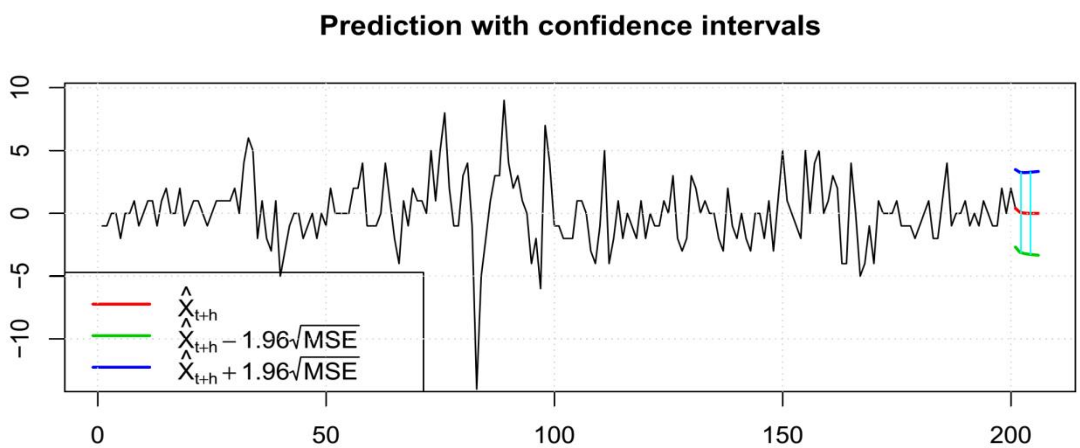
Forecasting For The ARIMA(4,1,0) component models					
Time	Point Forecast	Lo 8 0	Hi 8 0	Lo 95	Hi 95
1	0.1207294	0.11574909	0.1257098	0.11311265	0.1283462
2	0.1204272	0.11315139	0.1277030	0.10929981	0.1315546
3	0.1203169	0.11125424	0.1293795	0.10645677	0.1341770
4	0.1204910	0.10984094	0.1311411	0.10420313	0.1367789
5	0.1203165	0.10820141	0.1324315	0.10178808	0.1388449
6	0.1204403	0.10706294	0.1338176	0.09998140	0.1408992
7	0.1203336	0.10576082	0.1349064	0.09804645	0.1426208
8	0.1204273	0.10477575	0.1360789	0.09649031	0.1443643
9	0.1203424	0.10365822	0.1370265	0.09482616	0.1458586
10	0.1204173	0.10278006	0.1380546	0.09344347	0.1473912
11	0.1203512	0.10179309	0.1389093	0.09196904	0.1487333
12	0.1204094	0.10098796	0.1398309	0.09070686	0.1501120
13	0.1203580	0.10009816	0.1406178	0.08937326	0.1513427
14	0.1204034	0.09934851	0.1414583	0.08820273	0.1526040
15	0.1203633	0.09853380	0.1421928	0.08697797	0.1537486
16	0.1203987	0.09782840	0.1429690	0.08588041	0.1549170
17	0.1203674	0.09707371	0.1436612	0.08474275	0.1559921
18	0.1203950	0.09640496	0.1443851	0.08370539	0.1570847
19	0.1203707	0.09569937	0.1450420	0.08263917	0.1581022
20	0.1203922	0.09506185	0.1457225	0.08165280	0.1591315
21	0.1203732	0.09439721	0.1463492	0.08064636	0.1601000
22	0.1203899	0.09378687	0.1469930	0.07970405	0.1610758
23	0.1203752	0.09315692	0.1475934	0.07874845	0.1620019
24	0.1203882	0.09257058	0.1482058	0.07784482	0.1629316
25	0.1203767	0.09197044	0.1487829	0.07693308	0.1638203
26	0.1203869	0.09140556	0.1493682	0.07606379	0.1647099
27	0.1203779	0.09083132	0.1499244	0.07519032	0.1655654
28	0.1203858	0.09028582	0.1504858	0.07435185	0.1664198
29	0.1203788	0.08973433	0.1510233	0.07351212	0.1672455
30	0.1203850	0.08920643	0.1515635	0.07270151	0.1680685
Forecasting For The GARCH(1,1) component models					

	meanForecast	meanError	standardDeviation	lowerInterval	upperInterval
1	-2.860580e-05	0.003902622	0.002466723	-0.007677604	0.007620392
2	-1.122438e-05	0.003910188	0.002481655	-0.007675052	0.007652603
3	-8.634333e-06	0.003911068	0.002494852	-0.007674186	0.007656917
4	-3.709668e-05	0.003913695	0.002506524	-0.007707798	0.007633604
5	3.793547e-07	0.003915807	0.002516853	-0.007674461	0.007675220
6	2.797809e-06	0.003915859	0.002525999	-0.007672145	0.007677740
7	2.695314e-06	0.003915874	0.002534100	-0.007672276	0.007677667
8	3.216357e-06	0.003915885	0.002541280	-0.007671777	0.007678210
9	4.385095e-06	0.003915888	0.002547644	-0.007670615	0.007679385
10	4.531525e-06	0.003915889	0.002553288	-0.007670469	0.007679532
11	4.575957e-06	0.003915889	0.002558294	-0.007670425	0.007679577
12	4.636065e-06	0.003915889	0.002562735	-0.007670365	0.007679637
13	4.678511e-06	0.003915889	0.002566676	-0.007670323	0.007679680
14	4.687848e-06	0.003915889	0.002570175	-0.007670313	0.007679689
15	4.692481e-06	0.003915889	0.002573280	-0.007670309	0.007679694
16	4.696082e-06	0.003915889	0.002576037	-0.007670305	0.007679697
17	4.697902e-06	0.003915889	0.002578485	-0.007670303	0.007679699
18	4.698500e-06	0.003915889	0.002580659	-0.007670303	0.007679700
19	4.698823e-06	0.003915889	0.002582590	-0.007670302	0.007679700
20	4.699017e-06	0.003915889	0.002584305	-0.007670302	0.007679700
21	4.699107e-06	0.003915889	0.002585829	-0.007670302	0.007679700
22	4.699144e-06	0.003915889	0.002587182	-0.007670302	0.007679700
23	4.699163e-06	0.003915889	0.002588385	-0.007670302	0.007679700
24	4.699174e-06	0.003915889	0.002589453	-0.007670302	0.007679700
25	4.699179e-06	0.003915889	0.002590402	-0.007670302	0.007679700
26	4.699181e-06	0.003915889	0.002591246	-0.007670302	0.007679700
27	4.699182e-06	0.003915889	0.002591995	-0.007670302	0.007679700
28	4.699182e-06	0.003915889	0.002592661	-0.007670302	0.007679700
29	4.699183e-06	0.003915889	0.002593253	-0.007670302	0.007679700
30	4.699183e-06	0.003915889	0.002593779	-0.007670302	0.007679700

Source: forecasting results using R program.

For future dynamic of sugar prices, we predict that stability trend would be maintained; we expected the prices on March will record 0.12 $/kg, with a 95% confidence to reach the 0.16$/kg and decrease to the level of 0.07$/kg over this month.

Figure2. Forecast results of sugar prices over the period (2020M2-2020M3).



Source: Our plotting form R program**.**

4. Conclusion

The estimation of the volatility of agricultural products is a big challenge for producers, consumers, and the government's strategies and decision making. Under this context, we tried to analyze the pattern (or the behavior) of Sugar prices over the period (1962-2020) using ARIMA-GARCH models. We found that the prices exhibited high volatility over this period, where we detected a clustering of hyper-volatility in the (2007-2012) sub-periods, (a period of the world financial crisis).

For modeling results, the fitted ARIMA (4,1,1)-GARCH (1,1) was the optimal model among the other candidate models. On average, the variation of sugar prices depends positively on their previous variations, where the errors follow a GARCH (1,1) model. We gave the forecasting results for 30 periods, it seems to keep the same trend of prices and variations levels.

We highly suggested using such statistical models for other agricultural products, and we think that an econometrics study that includes factors affecting sugar prices will be great future work in this field.

References

[1] Acharya, S. S., & Agarwal, N. L. (1994). Agricultural prices-analysis and policy. *Agricultural prices-analysis and policy.*

[2] Baron, P. 2008. Sugar - is the situation hopeless?. 4th Dubai Sugar Conference ; Thinking Outside the Barrel', 3-5 February, Dubai,United Arab Emirates.

[3] Bollerslev, T "Generalized Autoregressive Conditional Heteroscedasticity," *Journal of Econometrics*, 31 (1986), 307-327.

[4] Box, G. E., & Pierce, D. A. (1970). Distribution of residual autocorrelations in autoregressive-integrated moving average time series models. *Journal of the American statistical Association*, *65*(332), 1509-1526.

[5] Chan, J. C. (2013). Moving average stochastic volatility models with application to inflation forecast. *Journal of Econometrics*, *176*(2), 162-172.

[6] Engle, R. F.(1982), Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation, *Econometrica*, 50, 987–1008.

[7] Gideon E. Schwarz." *Estimating the dimension of a model* ", Annals of Statistics, vol. 6, nº 2, 1978, p. 461-464. DOI 10.1214/aos/1176344136.

[8] Heady, E. O., & Kaldor, D. R. (1954). Expectations and errors in forecasting agricultural prices. *Journal of Political Economy*, *62*(1), 34-47.

[9] Holt, C. C. (1957). Forecasting seasonal and trends by exponentially weighted moving averages. Office of Naval Research, Research Memorandum, No. 52.

[10] Hyndman, R. J., Koehler, A. B., Ord, J. K., & Snyder, R. D. (2008). Forecasting with exponential smoothing: the state space approach. Berlin: Springer-Verlag.

[11] Hyndman, R.J. and Athanasopoulos, G. (2018). Forecasting: principles and practice", 2nd ed., OTexts, Melbourne, Australia. Section 3.4 "Evaluating forecast accuracy. https://otexts.org/fpp2/accuracy.html

[12] Hyndman, R.J. and Koehler, A.B. (2006) "Another look at measures of forecast accuracy". *International Journal of Forecasting*, 22(4), 679-688.

[13] Intriligator.M.D, Bodkin.R.G, Hsio.C.(1996),"Econometric Models, Techniques, and Applications", Prentice Hall edition.

[14] Kwiatkowski, D., P.C.B. Phillips, P. Schmidt and Y. Shin (1992), "Testing the Null Hypothesis of Stationarity against the Alternative of a Unit Root", *Journal* of Econometrics*, 54, 159-178.*

[15] Leybourne, S.J. and B.P.M. McCabe. (1994). A Consistent Test for a Unit Root", *Journal of Business and Economic Statistics*, 12, 157-166

[16] Ljung, G. M., & Box, G. E. (1978). On a measure of lack of fit in time series models. *Biometrika*, *65*(2), 297-303.

[17] N. Apergis, A. Rezitis. (2011). Food Price Volatility and Macroeconomic Factors: Evidence from GARCH and GARCH-X Estimates**.** Journal of gricultural and Applied Economics, 43 (1). pp. 95-110.

[18] Pablo M. PINCHEIRA, Carlos A. MEDEL. (2016). Forecasting with a Random Walk. Finance a úvěr-*Czech Journal of Economics and Finance*, 66, No. 6.

[19] Xiong, T., Li, C., Bao, Y., Hu, Z., & Zhang, L. (2015). A combination method for interval forecasting of agricultural commodity futures prices. *Knowledge-Based Systems*, *77*, 92-102.

Appendix

For data used in this work, you can freely download it through this link:
: https://www.macrotrends.net/2537/sugar-prices-historical-chart-data