American Journal of Applied Mathematics
2016; 4(3): 124-131
http://www.sciencepublishinggroup.com/j/ajam
doi: 10.11648/j.ajam.20160403.12
ISSN: 2330-0043 (Print); ISSN: 2330-006X (Online)

Methodology Article

Application of Statistical Methods of Time-Series for Estimating and Forecasting the Wheat Series in Yemen (Production and Import)

Douaik Ahmed[1], Youssfi Elkettan[2], Abdulbakee Kasem[2]

[1]The National Institute of Agronomic Research (INRA), Rabat, Morocco
[2]Department of Mathematics Faculty of Sciences, University Ibn Tofail, Kenitra, Morocco

Email address:
ahmed-douaik@yahoo.com (D. Ahmed), elkettani@univ-ibntofail.ac.ma (Y. Elkettan), kasemabdulbakee@gmail.com (A. Kasem)

To cite this article:
Douaik Ahmed, Youssfi Elkettan, Abdulbakee Kasem. Application of Statistical Methods of Time-Series for Estimating and Forecasting the Wheat Series in Yemen (Production and Import). *American Journal of Applied Mathematics*. Vol. 4, No. 3, 2016, pp. 124-131.
doi: 10.11648/j.ajam.20160403.12

Received: April 4, 2016; **Accepted**: April 15, 2016; **Published**: May 4, 2016

Abstract: Due to the importance of the wheat crop which represents 90% of the grain consumed, In this papers, we compared between the following statistical methods : Box and Jenkins model, exponential smoothing models (with trend and without seasonal) and Simple regression for estimating and forecasting to two time series of wheat(production and import). We reached to the following results: 1. Brown exponential smoothing model for modeling the imported wheat series. 2. ARIMA (1, 1, 1) model for modeling the product wheat series. For the wheat crop, the ratio of production to consumption is expected to reach 6.3% in 2015 and continues to decline even up to 5.4% in 2020. This means that the problem of food security well be worse in Yemen.

Keywords: Time Series, Wheat Crop, Forecasting, Box and Jenkins, Exponential Smoothing

1. Introduction

The human needs to know the past in order to predict the future to find optimal solutions of many problems which face humanity in this century. Yemen is one of the Arab countries where local demand for food is growing exponentially. Therefore, it suffers from a huge lack to cover all the population needs of foodstuffs especially wheat which represents staple food of most the population. Although in recent years the amount production of wheat compared with imported wheat reach in 2010 to 92%. According to what has mention above we compered these statistical methods of time series: Box and Jenkins methodology ,exponential smoothing model nd Simple regression to estimate and forecast the two wheat time series (import,product) from 1961 to 2010 of the Organization's site of Food and Agriculture (FAO) and the Central Bureau of Statistics in Yemen. We used these programmes *SPSS* , *EVIEWS* and *EXCLE* .

2. Theoretical Formulation

2.1. Holt and Brown's Exponential Smoothing Method

In the case where the series has a trend, we can adopt the following prediction formula:

$$\hat{y}_{t+h} = a_t + b_t h$$

The values a_t and b_t are constantly updated by the following equations:

$$a_t = \alpha_1 y_t + (1-\alpha_1)(a_{t-1} + b_{t-1})$$

and

$$b_t = \alpha_2 (a_t - a_{t-1} + (1-\alpha_2) b_{t-1})$$

This forecast model is known as the model name of HOLT. A special case of model HOLT, called model BROWN or dual exponential smoothing is obtained when the smoothing constants α_1 and α_2 are related to the same parameter α, by the relations: $\alpha_1 = \alpha(2-\alpha)$ et $\alpha_2 = \frac{\alpha}{2-\alpha}$. For these two models, we need to give initial values a_0 and b_0 to produce forecasts. Thus we take b_0 which equals the coefficient simple linear regression calculated on the basis of the first five values of the series. Thereafter, a_0 is deduced by the relation: $a_0 = y_1 - b_0$, as the smoothing constants they are set by the user. In practice often gives a value to α between 0.01 and 0.30. [3]

2.2. Stationary Process

A second process is stationary if:

- $\forall t \in Z, E(X_t^2) < \infty$
- $\forall t \in Z, E(X_t) = m$
- $\forall t \in Z, \forall h \in Z, Cov(X_t, X_{t+h}) = \gamma(h)$

2.3. White Noise

White noise is a stationary process such that:

- $E(\varepsilon_t) = m, \forall t$
- $Var(\varepsilon_t) = \sigma^2, \forall t$
- $Cov(\varepsilon_t, \varepsilon_{t-1}) = \gamma(h), \forall t, \forall h > 0$

This notion of white noise corresponds to the usual assumptions on residues in multiple regression. Random variables ε_t are also called random shocks. we implicitly assumes that random shocks ε_t follow a normal distribution $N(0,\sigma^2)$

2.4. Autocorrelation

The autocorrelation function is the application ρ of Z in R defined by:

$$\rho(h) = \frac{\gamma(h)}{\gamma(0)}, h \in Z$$

$\rho(h)$ Measuring the correlation between X_i and X_{i+h} because:

$$\frac{Cov(X_t, X_{t+h})}{\sqrt{V(X_t)}\sqrt{V(X_{t+h})}} = \frac{\gamma(h)}{\sqrt{\gamma(0)}\sqrt{\gamma(0)}} = \frac{\gamma(h)}{\gamma(0)}, h \in Z$$

2.5. Autocorrelation Partial

The partial autocorrelation function with delay k is defined as the partial correlation coefficient between X_t et X_{t-k} the influence of other variables shifted by k periods $X_{t-1}, X_{t-2,\dots,X_{t-k+1}}$ have been withdrawn.

2.6. Autoregressive Process AR(p)

Let a process $(X_t, t \in Z)$, X_t is said autoregressive process of order p $(AR(p))$ if

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \dots + \phi_p X_{t-p} + \varepsilon_t$$

Where ε_t, white noise and $\phi_1, \phi_2, \dots, \phi_p$ are constants.

2.7. Moving Average Process (MA(Q))

Let a process $(\varepsilon_t, t \in Z)$, X_t is said autoregressive process of order q $(MA(q))$ if

$$X_t = \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \dots + \theta_q \varepsilon_{t-q}$$

Where ε_t, white noise and $\theta_1, \theta_2, \dots, \theta_p$ are constants.

2.8. Moving Average Processes Autoregressive

A stationary process X has an $ARMA$ representation (p,q) Minimum if it satisfies:

$$\Phi(L)X_t = \Theta(L)\varepsilon_t,$$

$$\Phi(L)X_t = X_t + \phi_1 Xt - 1 + \dots + \phi_p X_{t-p}$$

$$\Theta(X_t) = \varepsilon_t + \theta_1 \varepsilon t - 1 + \dots + \theta_q \varepsilon_{t-q}$$

where

$$\phi_p \neq 0, \theta_q \neq 0 \; and \; L(X_t) = X_{t-1}$$

the polynomials Θ and Φ have their upper modules strictly roots to 1.

Θ and Φ have not common roots.

$\varepsilon = (\varepsilon_t, t \in Z)$ is a white noise of variance $V(\varepsilon_t) = \sigma^2 \neq 0$

2.9. Process ARIMA

A process $X = (X_t, t \geq 0)$ is a process $ARIMA(p,d,q)$ [Autoregressive integrated moving average] if it satisfies an equation of type :

$$(1 - \sum_{i=1}^{p} \varphi_i L^i)(1 - L^d)X_t = \delta(1 + \sum_{i=1}^{q} \theta_i L^i)\varepsilon_t, t \geq 0$$

where δ constant $L(X_t) = X_{t-1}$ and ε_t is white noise.

2.10. Augmented Dickey–Fuller Test

The Augmented Dickey–Fuller test (ADF) is a unit root test of the null hypothesis of unit root (or non stationarity). The ADF test estimated three models:

$$\Delta X_t = \alpha_1 X_{t-1} + \sum_{j=1}^{p} \beta_j \Delta X_{t-j} + \varepsilon_t$$

$$\Delta X_t = \alpha_0 + \alpha_1 X_{t-1} + \sum_{j=1}^{p} \beta_j \Delta X_{t-j} + \varepsilon_t$$

$$\Delta X_t = \alpha_0 + \alpha_1 X_{t-1} + \sum_{j=1}^{p} \beta_j \Delta X_{t-j} + \delta_t + \varepsilon_t$$

The null hypothesis of ADF test is the unit root hypothesis of the variable X_t is the hypothesis $H_0 : \alpha_1 = 0$. The ADF test consists of comparing the estimated value Student t associated with the parameter α_1 to the tabulated values of this statistic. The values tabulated for different test however tabulated values of Student test. The critical values of this statistic, ADF denoted in the following, are given by MacKinnon (1996). The null hypothesis H_0 of non-stationary of the time series is rejected at the 5% level when the observed value of the Student's t-test is less than the critical value tabulated by MacKinnon (1996) or $t_{abs} < ADF_{0.05}$.

2.11. Box Jenkins Methodology

This is the technique for select the most appropriate $ARMA$ or $ARIMA$ model for a given variable. It comprises four steps:

1. Identification of the model, this involves selecting the most appropriate lags for the AR and MA parts, as well as selecting if the variable requires first-differencing to become stationarity. The ACF and $PACF$ are used to identify the best model. (Information criteria can also be used)
2. Estimation, this usually involves the use of a least squares estimation process.
3. Diagnostic testing, which usually is the test for autocorrelation. If this part is failed then the process returns to the identification section and begins again, usually by the addition of extra variables.
4. Forecasting, the $ARIMA$ models are particularly useful for forecasting due to the use of lagged variables.

3. Application

3.1. Graph Series

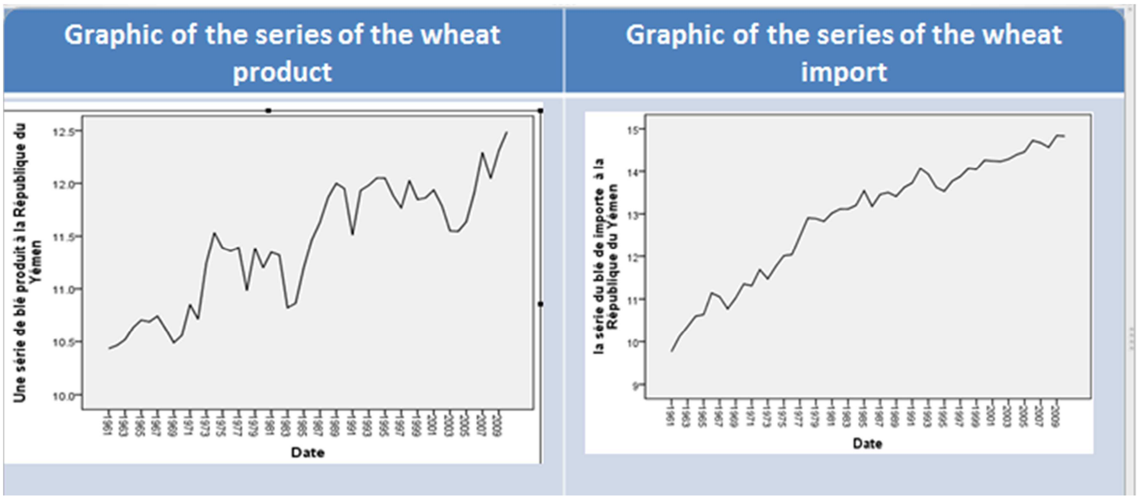

Figure 1. *Graph of time series of wheat(product and import)from 1961 to 2010.*

Through the graph figure, we observed a general upward trend over the period, this means that the series is not stationary.

3.2. Autocorrelation and Autocorrelation Partial

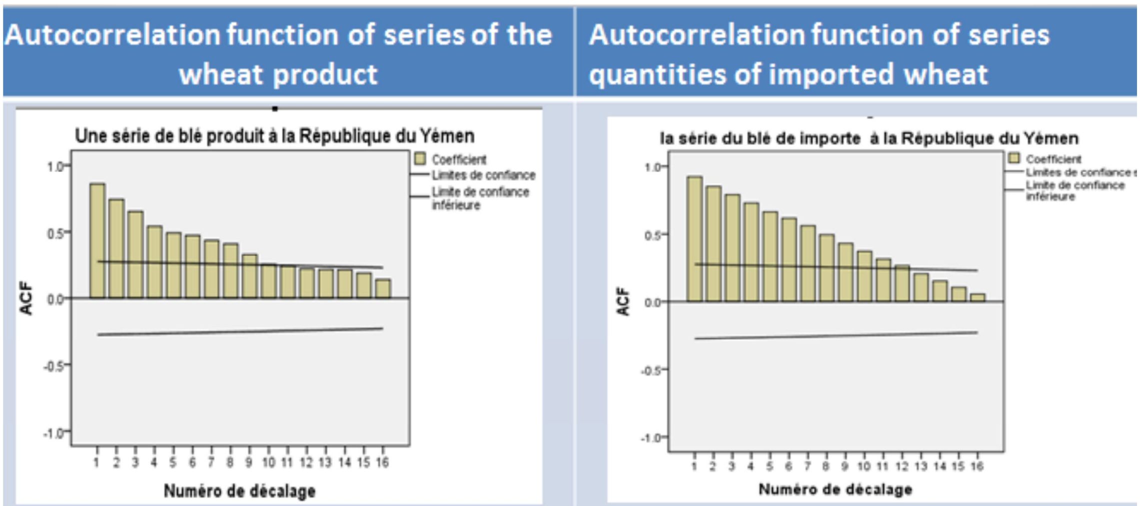

Figure 2. *Autocorrelation of time series of wheat(product and import).*

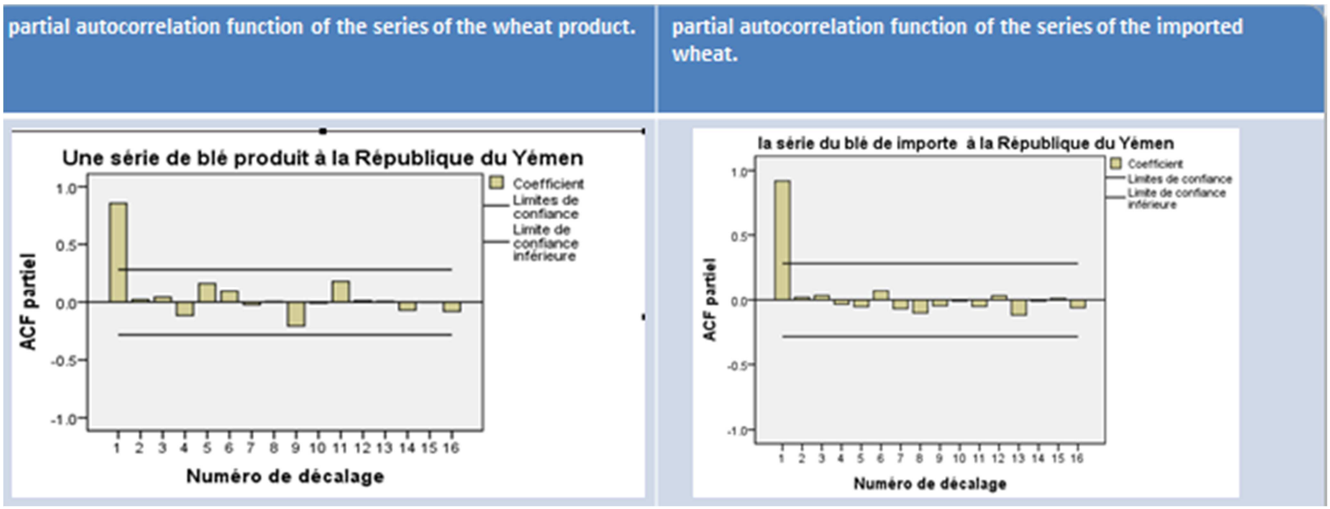

Figure 3. *Autocorrelation partial of time series of wheat(product and import).*

We examine the autocorrelation and partial autocorrelation function in figures 2 and 3 we observed that the estimated autocorrelation parameter decreases exponentially towards zero while that only the first partial autocorrelation parameter is not significant. To confirm the previous results we execute the Dickey-Fuller test and observed in Figure 4 and 5 that the series is not stationary.

Null Hypothesis: LA_SERIE_DU_BLE_PRODUIT_ has a unit root
Exogenous: Constant
Lag Length: 0 (Automatic - based on SIC, maxlag=10)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-0.516042	0.8790
Test critical values:	1% level	-3.571310	
	5% level	-2.922449	
	10% level	-2.599224	

*MacKinnon (1996) one-sided p-values.

Augmented Dickey-Fuller Test Equation
Dependent Variable: D(LA_SERIE_DU_BLE_PRODUIT_)
Method: Least Squares
Date: 08/30/13 Time: 02:38
Sample (adjusted): 1962 2010
Included observations: 49 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
LA_SERIE_DU_BLE_PRODUIT_(-1)	-0.036553	0.070832	-0.516042	0.6082
C	8457.491	8088.526	1.045616	0.3011

R-squared	0.005634	Mean dependent var	4723.102
Adjusted R-squared	-0.015523	S.D. dependent var	25098.99
S.E. of regression	25293.04	Akaike info criterion	23.15441
Sum squared resid	3.01E+10	Schwarz criterion	23.23162
Log likelihood	-565.2829	Hannan-Quinn criter.	23.18370
F-statistic	0.266299	Durbin-Watson stat	2.147802
Prob(F-statistic)	0.608244		

Figure 4. *Dickey-Fuller test of time series of wheat product.*

Null Hypothesis: LA_SERIE_DU_BLE_IMPORT_A has a unit root
Exogenous: Constant
Lag Length: 2 (Automatic - based on SIC, maxlag=10)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		2.452214	1.0000
Test critical values:	1% level	-3.577723	
	5% level	-2.925169	
	10% level	-2.600658	

*MacKinnon (1996) one-sided p-values.

Augmented Dickey-Fuller Test Equation
Dependent Variable: D(LA_SERIE_DU_BLE_IMPORT_A)
Method: Least Squares
Date: 08/30/13 Time: 19:50
Sample (adjusted): 1964 2010
Included observations: 47 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
LA_SERIE_DU_BLE_IMPORT_A(-1)	0.091719	0.037402	2.452214	0.0183
D(LA_SERIE_DU_BLE_IMPORT_A(-1))	-0.430251	0.156036	-2.757384	0.0085
D(LA_SERIE_DU_BLE_IMPORT_A(-2))	-0.426762	0.169553	-2.516977	0.0156
C	28236.68	35458.47	0.796331	0.4302

R-squared	0.213954	Mean dependent var	57976.47
Adjusted R-squared	0.159114	S.D. dependent var	176588.8
S.E. of regression	161931.7	Akaike info criterion	26.90900
Sum squared resid	1.13E+12	Schwarz criterion	27.06646
Log likelihood	-628.3615	Hannan-Quinn criter.	26.96826
F-statistic	3.901396	Durbin-Watson stat	2.028321

Figure 5. *Dickey-Fuller test of time series of wheat import.*

When we execute the first differences, we note of figure 6 and 7 that a stationary series.

Null Hypothesis: D(LA_SERIE_DU_BLE_PRODUIT_) has a unit root
Exogenous: Constant
Lag Length: 0 (Automatic - based on SIC, maxlag=10)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-7.594177	0.0000
Test critical values:	1% level	-3.574446	
	5% level	-2.923780	
	10% level	-2.599925	

*MacKinnon (1996) one-sided p-values.

Augmented Dickey-Fuller Test Equation
Dependent Variable: D(LA_SERIE_DU_BLE_PRODUIT_,2)
Method: Least Squares
Date: 08/30/13 Time: 02:39
Sample (adjusted): 1963 2010
Included observations: 48 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
D(LA_SERIE_DU_BLE_PRODUIT_(-1))	-1.138199	0.149878	-7.594177	0.0000
C	5342.318	3712.817	1.438885	0.1570

R-squared	0.556291	Mean dependent var	881.3125
Adjusted R-squared	0.546645	S.D. dependent var	37722.47
S.E. of regression	25399.16	Akaike info criterion	23.16359
Sum squared resid	2.97E+10	Schwarz criterion	23.24156
Log likelihood	-553.9262	Hannan-Quinn criter.	23.19306
F-statistic	57.67153	Durbin-Watson stat	1.963053
Prob(F-statistic)	0.000000		

Figure 6. *Dickey-Fuller test of first differences of wheat product.*

Null Hypothesis: D(LA_SERIE_DU_BLE_IMPORT_A) has a unit root
Exogenous: Constant
Lag Length: 0 (Automatic - based on SIC, maxlag=10)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-8.300790	0.0000
Test critical values:	1% level	-3.574446	
	5% level	-2.923780	
	10% level	-2.599925	

*MacKinnon (1996) one-sided p-values.

Augmented Dickey-Fuller Test Equation
Dependent Variable: D(LA_SERIE_DU_BLE_IMPORT_A,2)
Method: Least Squares
Date: 08/30/13 Time: 19:51
Sample (adjusted): 1963 2010
Included observations: 48 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
D(LA_SERIE_DU_BLE_IMPORT_A(-1))	-1.201944	0.144799	-8.300790	0.0000
C	68608.30	26359.65	2.602778	0.0124

R-squared	0.599663	Mean dependent var	-1041.604
Adjusted R-squared	0.590960	S.D. dependent var	270693.7
S.E. of regression	173125.5	Akaike info criterion	27.00220
Sum squared resid	1.38E+12	Schwarz criterion	27.08016
Log likelihood	-646.0527	Hannan-Quinn criter.	27.03166
F-statistic	68.90311	Durbin-Watson stat	2.089467
Prob(F-statistic)	0.000000		

Figure 7. *Dickey-Fuller test of first differences of wheat import.*

we note that the series is stationary. We deduce that d = 1 in the ARIMA model (p, d, q).

3.3. Identification and Selction of Model for Wheat Production Series

Although it appears that each partial autocorrelation parameter after the second parameter is not significantly different from zero at a = 0.05 but the autocorrelation function is gradually decreasing towards zero, this may be sufficient evidence that the random process is AR (1). For ensure we test the following statistical hypothesis: $H_0:\varphi_{11}=0$; $H_0:\varphi_{11}\neq 0$,

$$SE(\varphi_{11})=\frac{1}{\sqrt{n}}=\frac{1}{\sqrt{50}}=0{,}141\ ,\quad Z=\varphi_{11}/(SE(\varphi_{11}))=\frac{0{,}929}{0{,}141}=6{,}5>2\ ,$$

we deduce that the first partial autocorrelation parameter is not significantly different from zero at $\alpha=0.05$. We examining the autocorrelation partial parameters, we find that $\varphi_{kk}<0.282$ *for each* $k=2,3,....$, that supports the possibility of using the AR (1) and therefore $ARIMA(1,1,0)$.

For import wheat series we get the same results. And then we compared between ARIMA models with the exponential smoothing(Holt,Brown)and simple regression. We get the following results.

Comparison of ARIMA models with the exponential smoothing and simple regression for time series of import wheat

Model	R²-	Bic	MAPE	Lujng Box	Sig du Model
ARIMA (1,1,0)	0.79	20.5	18.2	0.21	const=0.15 ϕ1 =0.44
ARIMA (0,1,0)	0.79	20.3	18	0.45	const=0.20
ARIMA (2,1,0)	0.80	20.5	18.3	0.47	const=0.15 ϕ1 =0.30 ϕ2 =0.43
Brown	0.78	20.4	18.4	0.17	sig =0
Holt	0.80	20.4	17.4	0.39	sig =0 Gama =1
Regression	0.73	----	23.3	0.74	const=0.015 Year =0.0

Comparison of ARIMA models with the exponential smoothing and simple regression for time series of product wheat

Model	R²-	Bic	MAPE	Lujng Box	Sig du Model
ARIMA (1,1,0)	0.95	24.3	38.6	0.67	const=0.15 ϕ1 =0.44
ARIMA (0,1,0)	0.95	24.4	34.7	0.70	const=0.19
ARIMA (2,1,0)	0.96	24.3	42.6	0.82	const=0.12 ϕ1 =0.30 ϕ2 =0.43
Brown	0.92	24.1	21.4	0.21	sig =0
Holt	0.96	24.12	16.8	0.21	sig=0 Gama =1
Regression	0.87	----	166	0.36	const=0.01 Year=0.0

Figure 8. *Cooparison of model.*

We take 40 observation of the original series and forecast for the next ten years, then compare between models by $MAPE$ and choose the best model. The results were as follows:

1. Brown's exponential smoothing model for predict the series of wheat production.

2. The $ARIMA(1,1,1)$ model for predict the series of wheat exports.

3.4. Tests of Residues

We test the best model:

- Graphic residues confidence limits, $ACF, PACF$
- Graphic dispersion of points in parallel form residuals around zero $ACF, PACF$
- Ljung-Box value is significant
- If the model realizes the previous tests, we use it to forecast.

3.5. Forecasting

Then we use the previous models to calculate the forecast from 2011 to 2020 and the results were as follows:

Année	Forecast of series of the wheat product	Forecast of Series of the wheat import.
2011	251,887	2,872,202
2012	242,580	3,005,136
2013	236,454	3,138,069
2014	232,777	3,271,003
2015	230,184	3,403,936
2016	230,184	3,536,869
2017	230,603	3,669,803
2018	231,783	3,802,736
2019	233,533	3,935,760
2020	235712	4,068,603

Figure 9. *Forecasting of series of the wheat (product and import).*

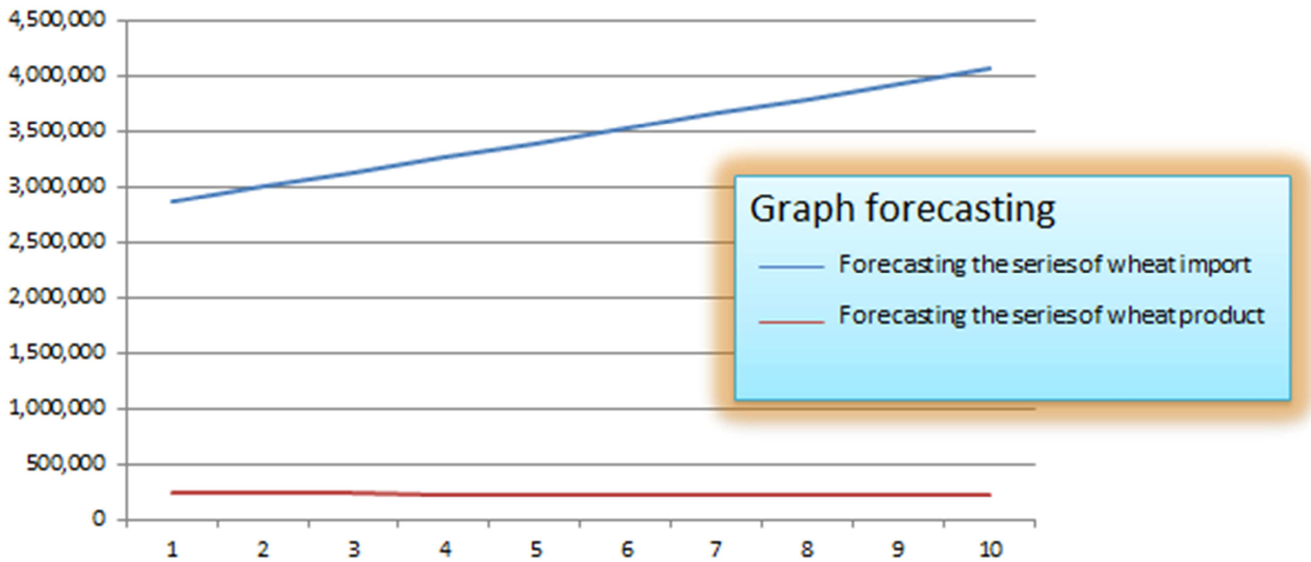

Figure 10. *Graph forecasting of series of the wheat (product and import).*

4. Conclusion

Wheat imports will increase from 2.9 million tonnes in 2011 to 4 million tonnes in 2020, where the proportion of imports was 92% in 2010 and it is expect that the wheat import proportion will increase to 94% in 2020. whereas, wheat production will drop by 6% during this decade.

References

[1] Peter J. Brockwell et Richard A Davis (2002), Introduction to Time Series and Forecasting, Springer.

[2] Philippe Marier, cours Prévision de la demande, Consortium de Recherche Université Laval.

[3] Lahoussaine Baamal, (2012), Cours d'analyse des Séries Chronologiques. Université Ibn Tofail,Kénitra.

[4] E. Ostertagova, O. Ostertag, The Simple Exponential Smoothing Model, http://www.researchgate.net/publication/256088917

[5] Rodolphe Palm, (2007), Etude des séries chronologiques par méthodes de lissage, Faculté Universitaire des Sciences agronomiques, Unité de Statistique, Informatique et Mathématique appliquées, Belgique.

[6] Bary Adnan Majed, (2002), Statistical forecasting methods-1, Université du Roi Saoud.

[7] Ruey S. Tsay, (2005), Analysis of financial time series, Wiley,Hoboken, New Jersey.

[8] Lahoussaine Baamal, (2012), Cours de Prévision par la Méthodologie de Box et Jenkins. Université Ibn Tofail.

[9] IBM, (2011), IBM SPSS Forecasting.

[10] A.Douaik, (1991), Prévision des rendements agricoles par les méthodes de lissage exponentiel. Mémoire ingénieur, Faculté des Sciences Agronomique de Gembloux, Belgique.