International Journal of Innovation and Applied Studies
ISSN 2028-9324 Vol. 16 No. 1 May 2016, pp. 94-101

http://www.ijias.issr-journals.org/

Application of statistical methods of time-series for estimate and predict of the food gap in Yemen

***Douaik Ahmed*[1]*, Youssfi Elkettani*[2]*, and Abdulbakee Kasem*[2]**

[1]Institut National de Recherche Agronomique de Rabat, Morocco

[2]Department of Mathematics Faculty of Sciences, University Ibn Tofail, Kenitra, Morocco

ABSTRACT: In this paper we provided modeling for the food gap in Yemen. We have studied this model by descriptive and analytical studies and formulated a model for the food gap, we estimated its parameters and predicted for the coming ten years using the Box and Jenkins methodology of time series analysis. Then, we compared this methodology to the exponential smoothing and simple regression methods.
We found the following main results for the three time series regarding the food gap:
1. ARIMA (1, 1, 1) model to predict the price of food importation series.
2. Brown exponential smoothing model to predict the price of food exportation series.
3. ARIMA (1, 1, 1) model to predict the price of food production series.
Through the results, we concluded that food production will not meet the local demand for food, where of the equation: local demand consumption of food = food importation + food production - food exportation. The ratio of production to consumption is expected to reach 29.3 % in 2015 and to continue to decline to reach 28.8 % in 2020.

KEYWORDS: Time Series, Food security, Forecasting.

1 INTRODUCTION

Forecasting is a decision-making tool used by many Governments and businesses to help in budgeting, planning, and estimating future growth. In the simplest terms, forecasting is the attempt to predict future outcomes based on past events and management insight. The problem of food security is a very important subject which must be mobilized government officials and decision-makers to this fact.then.. Yemen is one of the the developing countries where local demand for food is growing exponentially, this means that it suffers from a huge lack to cover all the population needs of foodstuffs. Therefore, we applied statistical methods for time series by compared between the Box–Jenkins method which discovered in 1970 by Scientists George Box and Gwilym Jenkins (autoregressive moving average ARMA or ARIMA models), exponential smoothing (trend, seasonal) which is used a lot in the short-term forecast and simple linear regression, to estimate and of the food deficit for three annual data time series from 1961 to 2010 taken of the Organization's site of Food and Agriculture (FAO) and the Central Bureau of Statistics in Yemen.

2 THEORETICAL FORMULATION

2.1 SIMPLE REGRESSION ANALYSIS

A relation between the dependent variable Y and the independent variable X can be modeled as a function of type $Y = \alpha + \beta X$, Y : dependent variable (explained), X : independent variable (explicative).

2.2 MODEL OF TIME SERIES

2.2.1 DEFINITION

[1] A time series model for the observed data $\{x_t\}$ is a specification of the joint distributions (or possibly only the means and covariances) of a sequence of random variables $\{X_t\}$ of which $\{x_t\}$ is postulated to be a realization.

2.2.2 CHOOSE THE CORRECT MODEL FOR PREDICTION

For the comparison of time series models by using the methodology of Box and Jenkins, exponential smoothing and simple regression, we used as a measure of the quality of forecasts $MAPE$ (Mean Absolute Percentage Error). This is the relative error made by a predictive model. The following equation shows how we calculate this criterion:

$$MAPE = (\frac{100}{n}\sum_{t=1}^{n}|\frac{D_t - P_t}{D_t}|)$$

Where : y_t :initial observation, p_t :forecasting the initial observation and n :number of the initial observations [2]. Other criteria for comparison of models:

BIC (Bayesian Information Criterion, 1978), R^2, significance of the parameters, residue analysis:(Ljung - Box,Durbin-Watson). The BIC is formally defined as

$$BIC = -2 * ln\,\hat{L} + K * ln(n)$$

where:

- x = the observed data;
- θ = the parameters of the model;
- n = the number of data points in x, the number of observations, or equivalently, the sample size;
- k = the number of free parameters to be estimated. If the model under consideration is a linear regression, k is the number of regressors, including the intercept;
- $\hat{L}$ = the maximized value of the likelihood function of the model M , $\hat{L} = p(x\,|\,\hat{\theta}, M)$, where $\hat{\theta}$ are the parameter values that maximize the likelihood function.

R^2 (the coefficient of multiple determination):The elevation of the linear fit of the regression equation between the dependent variable Y and the set of explanatory variable is determined by the coefficient of determination R_2 = SCR/SCT with $0 \le R^2 \le 1$ where $SCR = \sum(\hat{Y}_t - y)^2$, $SCT = \sum(Y_t - y)^2$, $\hat{Y}_t$ is the estimated values, Y_t is the real values, y is the average of real values.

2.2.3 FORECAST BY EXPONENTIAL SMOOTHING

SERIES WITH TREND COMPONENT WITHOUT SEASONAL

$$\hat{y}_{t+h} = a_t + b_t h, h = 1,2,3,...(According\ to\ the\ number\ required\ to\ provide) \qquad (1)$$

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

$a_t = \alpha_1 y_t + (1-\alpha_1)(a_{t-1} + b_{t-1})$ and $b_t = \alpha_2(a_t - a_{t-1} + (1-\alpha_2)b_{t-1})$ (2)

This prediction model is known as Holt's model name. A special case of Holt model, referred to as Brown model or dual exponential smoothing is obtained when the smoothing constant α_1 and α_2 are linked to the same parameter α, by the relations:

$\alpha_1 = \alpha(2-\alpha)$ (3) , $\alpha_2 = \alpha/(\alpha-2)$ (4)

For these two models, we need to give initial values a_0 and b_0 to produce forecasts. Thus we takes b_0 equal to simple linear regression coefficient calculated on the basis of the first five series of values, thereafter, a_0 is deduced by the relation: $a_0 = y_1 - b_0$.

2.2.4 FORECAST BY THE METHODOLOGY OF BOX AND JENKINS

Box and Jenkins based on the notion of ARIMA process. This technique has four steps: identification, Estimation, Diagnostic and Forecast. The model consists of two parts: the frist one is autoregression (AR) and the second is moving average (MA)and model is generally noted ARMA (p,q), where p is the order of the AR part and q the order of the party MA, a moving average and autoregressive model of order (p,q) (abbreviated as ARMA (p,q)) is a discrete time processes $(X_t, t \in N)$ verifying:

$$Y_t = \phi_p Y_{t-p} + ... + \phi_1 Y_{t-1} \quad + \alpha_t - \theta_1 Y_{t-1} - .... - \phi_q Y_{t-q}$$

where ϕ_i and θ_i parameters are constants and the terms of errors α_t are independent of the process.

AUTOREGRESSION MODEL AND MOVING AVERAGE MODEL

An autoregressive model of order p, abbreviated AR (p) is written:

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + ... + \phi_p Y_{t-p} + \alpha_t$$

and the notation MA (q) refers to the moving average model of order q is written:

$$Y_t = \alpha_t - \theta_1 Y_{t-1} - \theta_2 Y_{t-2} - ... - \theta_q Y - t - q$$

where $\phi_1,...,\phi_p$, $\theta_1,...,\theta_q$ are the model parameters, α_t an **white noise** that define is a series of random observation noncorrelation and sometimes assume that it is a series of independent random variables and have the same distribution, zero mean and constant variation such that 1. $E(\alpha_t) = m, \forall t$, 2. $Var(\alpha_t) = \sigma^2, \forall t$, 3. $cov(\alpha_t, \alpha_{t-h}) = \gamma(h), \forall t, h > 0$. Additional constraints on the parameters are necessary to ensure the stationarity. A process is stationary in the second if: 1. $\forall t \in \mathbb{Z}, E(X_t^2) < \infty$, 2. $\forall t \in \mathbb{Z}, E(X^2) = m$, 3. $\forall t \in \mathbb{Z}, \forall h \in \mathbb{Z}, Cov(X_t, X_{t+h}) = \gamma(h)$.

The model requires that a original series Y_t is stationary, but if it is not, it must be converted to a stationary taking the necessary differences, in this case, the ARMA model is written ARIMA (autoregressive integrated moving average). This model is characterized by three types of parameters: an autoregression (p), an integration (d) and a moving average (q) and is symbolized as follows: $ARIMA\ (p,d,q)$ [6]. The Box-Jenkins methodology is summarized in four steps to found the best model for estimating and forecasting time series. These steps are sometimes nested with one another.

1. Identification: This is the most important step is to select the rank which model consists of it, first Know (d) through determine the degree of integration with study the stability of original time series. If the series is not stationary, for example having a general trend, we take the first difference, the series can become stationary after a number of differences, we can use several methods to know stationary of the series as the basic unit root test of Dickey-Fuller test and the expanded Dickey-Fuller (1979, 1981, 1986).[7]. Then we know (p) and (q), we test the significance of the autocorrelation function, through $SE(r_k) = \frac{1}{\sqrt{n}}$ and test the hypothesis $H_o : \rho(k) = 0$

against $H_1 : \rho(k) \neq 0 \ for \ k = 1,2,....using \ |Z| = \frac{r(k)}{SE(r(k))}$, H_o is rejected if $|Z| > 2$ and we deduce that $\rho(k)$ is significantly different from zero. The standard error for each of the autocorrelation coefficients must be less than $2 * SE(r_k)$ $, k > q$ where $SE(r_k) = \sqrt{\frac{1}{n} + 2\sum_{j=1}^{q} r^2(j)}, k > q$. To test the significance of the partial autocorrelation function is calculated : $SE(\phi_k k) = \frac{1}{\sqrt{n}}$

and test the hypothesis $H_o : \phi(kk) = 0$ against $H_1 : \phi(kk) \neq 0 \ for\ k = 1,2,....using\ |Z| = \frac{\phi(kk)}{SE(\phi(kk))}$, H_o is rejected if $|Z| > 2$ and we deduce that $\phi(kk)$ is significantly different from zero. The standard error for each of the autocorrelation coefficients must be less than $2*SE(\phi(kk)), k > p$ where $SE(\phi(kk)) = \frac{1}{\sqrt{n}}$. It also can be identified through graph $ACF, PACF$

2. Estimation : This usually involves the use of a least squares estimation process.

3. Diagnostic : The diagnostic of model is generally to do many tests and checks. The most important are: the stationary, study of residues and statistics.

4. Forecast : After identifying the model (determination of p, d and q), estimated parameters and diagnosed, it is used for the preparation of forecasts.

3 APPLICATION

3.1 THE PRICE SERIES OF VALUES (IMPORTS, EXPORTS AND PRODUCTION) FOOD

3.1.1 GRAPH SERIES

Fig. 1. Graph series of imports food 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.

3.1.2 AUTOCORRELATION AND AUTOCORRELATION PARTIAL

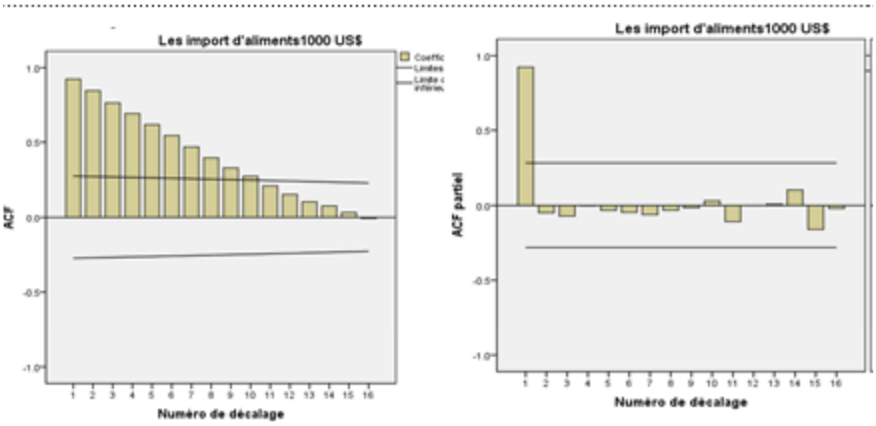

Fig. 2. Autocorrelation and partial autocorrelation series of imports food

We examine the autocorrelation and partial autocorrelation function in figures (2) and (2-1) 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 (3) that the series is not stationary.

Null Hypothesis: SERIES02 has a unit root
Exogenous: Constant
Lag Length: 0 (Automatic - based on SIC, maxlag=10)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		1.171568	0.9976
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(SERIES02)
Method: Least Squares
Date: 08/23/13 Time: 23:40
Sample (adjusted): 1962 2010
Included observations: 49 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
SERIES02(-1)	0.050066	0.042734	1.171568	0.2473
C	16893.65	33392.83	0.505906	0.6153

R-squared	0.028375	Mean dependent var	45989.92
Adjusted R-squared	0.007702	S.D. dependent var	156861.8
S.E. of regression	156256.6	Akaike info criterion	26.79635
Sum squared resid	1.15E+12	Schwarz criterion	26.87356
Log likelihood	-654.5105	Hannan-Quinn criter.	26.82564
F-statistic	1.372572	Durbin-Watson stat	2.067120
Prob(F-statistic)	0.247276		

Null Hypothesis: D(SERIES02) has a unit root
Exogenous: Constant
Lag Length: 0 (Automatic - based on SIC, maxlag=10)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-6.493626	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(SERIES02,2)
Method: Least Squares
Date: 08/23/13 Time: 23:42
Sample (adjusted): 1963 2010
Included observations: 48 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
D(SERIES02(-1))	-0.956073	0.147233	-6.493626	0.0000
C	44811.74	24083.40	1.860690	0.0692

R-squared	0.478264	Mean dependent var	324.3542
Adjusted R-squared	0.466922	S.D. dependent var	219088.4
S.E. of regression	159961.3	Akaike info criterion	26.84402
Sum squared resid	1.18E+12	Schwarz criterion	26.92199
Log likelihood	-642.2566	Hannan-Quinn criter.	26.87349
F-statistic	42.16718	Durbin-Watson stat	2.009967
Prob(F-statistic)	0.000000		

Fig. 3. Dickey-Fuller test of differentiated first series of price values on food imports

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

IDENTIFICATION 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: $H_0 : \phi_{11} = 0$; $H_0 : \phi_{11} \neq 0$, $SE(\phi_{11}) = \frac{1}{\sqrt{n}} = \frac{1}{\sqrt{50}} = 0{,}141$, $Z = \phi_{11}/(SE(\phi_{11})) = \frac{0{,}924}{0{,}141} = 6{,}4 > 2$, 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 $\phi_{kk} < 0.282$ *for each* $k = 2,3,....$. Thus, can say that there is no reason for the partial autocorrelation function does not stop after the first parameter that supports the possibility of using the AR (1) and therefore $ARIMA(1,1,0)$.

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

Table 1. Comparison of model

Comparison of ARIMA models with the exponential smoothing and simple regression for the series of price values of food importation.						Comparison of ARIMA models with the exponential smoothing and simple regression for the series of price values of food exportation.					
Model	R-deux	Bic	MAPE	Luing Box	Sig du Model	Model	R-deux	Bic	MAPE	Luing Box	Sig du Model
ARIMA (1,1,0)	0.92	24.1	16.9	0.20	const=0.01 ϕ_1 =0.94	ARIMA (1,1,0)	0.91	19	25.6	0.17	const=0.42 ϕ_1 =0.16
ARIMA (0,1,0)	0.92	24	16.8	0.25	const=0.008	ARIMA (0,1,0)	0.92	18.8	26.7	0.03	const=0.31
ARIMA (1,1,1)	0.93	24.2	17	0.71	const=0.08 ϕ_1 =0.00 θ_1 =0.08	ARIMA (2,1,0)	0.91	19.1	24.7	0.65	const=0.08 ϕ_1 =0.11 ϕ_2 =0.25
Brown	0.89	24.4	17	0.12	Alpha =0	Brown	0.80	19.4	29.8	0.0	Alpha =0
Holt	0.93	24.1	16.4	0.15	Alpha =0 Gama =1	Holt	0.92	18.9	26.5	0.20	Alpha =0 Gama =0.97
Regression	0.70	----	106	0.25	const=0.02 Year=0.0	Regression	0.92	----	93	0.25	const=0. Year=0.

Table 2. Comparison of model

Comparison of ARIMA models with the exponential smoothing and simple regression for the series of price values of food production.					
Model	R-deux	Bic	MAPE	Luing Box	Sig du Model
ARIMA (1,1,0)	0.99	21.7	5.7	0.31	const=0.001 ϕ_1 =0.22
ARIMA (0,1,0)	0.98	21.8	4.8	0.15	const=0.00
ARIMA (1,1,1)	0.99	21.9	5.3	0.31	const=0.4 ϕ_1 =0.00 θ_1 =0.00
Brown	0.99	21.5	4.9	0.06	Alpha =0
Holt	0.99	21.6	4.6	0.14	Alpha =0 Gama =0.17
Regression	0.87	----	15.7	0.12	const=0.001 Year=0.0

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. The $ARIMA(1,1,1)$ for modeling the series of price values of food imports.
2. Brown's exponential smoothing model for modeling the series of price value of food exports.
3. The $ARIMA(1,1,1)$ for modeling the series of price values of food products.

3.2 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.

3.3 FORECASTING

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

Table 3. Forecasting prix value of the time series(exports, imports and production)

Forecasting

Year	value price series of food imports	value price series of food exports	value price series of food products
2011	2,388,977	162105	1,868,713
2012	2,523,527	185227	1,938,459
2013	2,628,949	204683	2,006,854
2014	2,768,258	222056	2,073,951
2015	2,880,080	241559	2,139,800
2016	3,024,192	264090	2,204,448
2017	3,142,373	288899	2,267,943
2018	3,291,327	315571	2,330,327
2019	3,415,827	344878	2,391,645
2020	3,569,662	376789	2,451,936

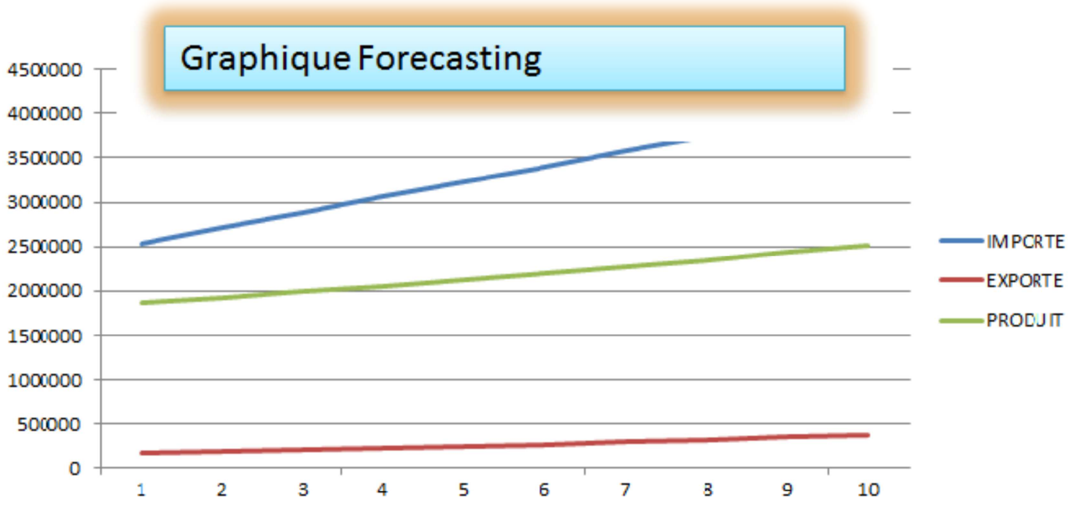

Fig. 4. Graphic Forecast prix value of time series(exports,imports and production)

4 CONCLUSION

- Through forecasting results we observed that food imports will continue to increase up to 92 % compared to food exports.
- According to forecasting results of the price value of food exports over the next 10 years, it seems that the value of export prices will increase slowly, which means that the Yemeni economy will remain as an importing economy at least during the current decade .
- We predicted to continue to decrease in the second decade of the current century (2011-2020) to 39 %.

REFERENCES

[1] Peter J. Brockwell et Richard A Davis,Introduction to Time Series and Forecasting, Springer,(2002).
[2] Philippe Marier,Prévision de la demande,Consortium de Recherche Université Laval.
[3] Lahoussaine Baamal,Cours d'analyse des Séries Chronologiques.Université Ibn Tofail,Kénitra,(2012).
[4] E. Ostertagova,O. Ostertag,The Simple Exponential Smoothing Model, http://www.researchgate.net/publication/256088917
[5] Rodolphe Palm, Etude des séries chronologiques par méthodes de lissage, Faculté Universitaire des Sciences agronomiques, Unité de Statistique, Informatique et Mathématique appliquées, Belgique,(2007)
[6] Bary Adnan Majed,Statistical forecasting methods-1, Université du Roi Saoud, Arabie Saoudite(2002).
[7] Ruey S. Tsay, Analysis of financial time series, Wiley,Hoboken, New Jersey,USA.(2005).
[8] Lahoussaine Baamal,Cours de Prévision par la Méthodologie de Box et Jenkins.Université Ibn Tofail,Kénitra,(2012).
[9] IBM, IBM SPSS Forecasting,(2011).
[10] A.Douaik,Prévision des rendements agricoles par les méthodes de lissage exponentiel. Mémoire ingénieur, Faculté des Sciences Agronomique de Gembloux, Belgique,(1991).