

Time Series Modelling and Forecasting for a System of Credit and Debit in the Cameroon's NSIF

Jimbo Henri Claver^{1*}, Ngongo Isidore Séraphin², Alemoka Jean Jacques³, Andjiga Gabriel Nicolas³,
Etoua Remy Magloire³ and Takeru Suzuki³

¹Department of Applied Mathematics and Statistics, AUAF & Waseda University, Tokyo, Japan

²Department of Mathematics, ENS, University of Yaoundé I, Cameroon

³Department of Mathematics, University of Yaoundé I, Cameroon

Email: jimbo.maths@gmail.com

Abstract In this article, we develop a model of forecasting credit and debit of pension funds of the NSIF in Cameroon. By using time series tools and relying on the ARMA model (Auto - Regressive Moving Average), we appropriately analyze and predict the main existing credit and debit of funds. The aim is to elaborate a model that is able to provide reliable information on credit and debit, mainly on the financial balance of the regime in order to guarantee and also ensure the management and financial planning of pension funds managed by the National Social Insurance Fund (NSIF) in Cameroon.

Keywords: time series analysis, pension funds, credit, debit, ARMA, NSIF, forecasting and planning

1 Introduction

The policy of retirements is nowadays one of the major concerns of modern societies, and for any national economy. It appeared at the beginning of the 20th century in Western Europe and was essentially based on the wage earning. The retirement systems have been set in Sub-Saharan African countries during the colonial period ([5]). Equally some organizations were created to address the problem of social security.

In Cameroon for example, a pension fund is an autonomous entity in charge of providing services to its members and to all the administration that is involved. It is managed by the National Social Insurance Fund (NSIF) and contributes in supporting the workers of private and public sectors under a well-defined labor code. The financial situation of the NSIF of Cameroon pension fund authority shows a deficit so far since debit are more than credit. Alongside the accrued deficit can be added the problems of recovery which is estimated in 2011 for more than 600 billion FCFA [5]. The main objective of this research work is to develop a model and system of forecasting NSIF management in Cameroon. To address this problem we will use the data of credit and debit of statistical yearbooks (2010-2016) obtained from the General Directorate, Department of Studies, Communication and Translation (GDDSCCT). For this end, we shall provide ourselves with mathematical and statistical tools such as time series analysis, or chronological series, that correspond to statistical data analysis techniques of the observations regularly spaced in time. It is well established that the domain of application of these tools is vast and spreads from Astronomy to Economics, passing through Biology or the Theory of Signal. They have also aroused a very keen interest, which has led to the development of various models such as AR, MA, ARMA, ARIMA, SARIMA, GARCH, the list is not exhaustive, for further reading see for example [7]. In this work, we use specifically the ARMA model (Auto Regressive Moving Average) and techniques from Box and Jenkins book to set a model that will allow us to adequately analyze the evolution of credit and debit of pension funds and the NSIF in Cameroon. We finally use the R software, and the command "predict" involved in the algorithm to carry out forecasting for the year (2017-2018). This article is divided into three main parts, the first part describes the concept of time series analysis using the selected models; the second part focuses on the application of the model and the third part will analyze the obtained results.

2 Time Series Analysis

2.1 Univariate Process

White noise A white noise u_t by definition is a stationary process with mean zero and constant variance in the weak sense.

$$u_t \rightsquigarrow N(0; \sigma^2)$$

So, u_t is a white noise if for all $t \in T$: $E[u_t] = 0$; $E[u_t^2] = \sigma^2 < \infty$; $E[u_t; u_{t-h}] = 0$ with u_t and u_{t-h} independents if $h \neq 0$ et $(t-h) \in T$.

Linear process There exists a series $\{\varepsilon\}_{t \in \mathbb{R}}$ so that $\{y_t\}$ can be represented by:

$$y_t = A(L)\varepsilon_t = \left(\sum_{j=-\infty}^{+\infty} a_j L^j \right) \varepsilon_t = \sum_{j=-\infty}^{+\infty} a_j \varepsilon_{t-j} \quad (1)$$

Decomposition of Wold : Every stationary process in the weak sense with mean zero $\{y_t\}$, admits the following representation:

$$y_t = \sum_{j=0}^{+\infty} \varphi_j \varepsilon_{t-j} + k_t \quad (2)$$

where $\varphi_0 = 1$ and $\sum_{j=0}^{+\infty} \varphi_j^2 < \infty$

The term ε_t is a white noise that represents the error made in foreseeing y_t from a linear function of its historic: $Y_{t-1} = \{y_{t-j}\}_{j=1}^{\infty}$; $\varepsilon_t = y_t - E[y_t | Y_{t-1}]$. The variable k_t is non-correlated ε_{t-j} , for all $j \in \mathbb{Z}$, although k_t may be foreseen arbitrarily from a linear function of Y_{t-1} : $\sum_{j=0}^{+\infty} \varphi_j \varepsilon_{t-j}$ Is a stochastic linear component and k_t is a deterministic linear component of y_t .

Process of moving average

MA of order 1 We say that the following series y_t is a process of moving average of order 1 and we note $MA(1)$, if it is generated by a white noise ε_t under the form:

$$\begin{aligned} y_t &= \varepsilon_t - \theta \varepsilon_{t-1} \\ &= (1 - \theta L)\varepsilon_t \end{aligned}$$

The transfer function of the filter reduces to an only term. By a simple calculation, we can show that:

$$\text{Var}(y_t) = (1 + \theta^2)\sigma_\varepsilon^2; \quad \text{Var}(y_t) > \sigma_\varepsilon^2$$

Meaning that by modelling, we reduce the variance of the phenomenon which is by nature, the peculiarity of every modelling.

Auto-correlation of y_t . We can demonstrate that:

$$\rho(k) = \begin{cases} \frac{-\theta}{1 + \theta^2} & \text{si } k = 1 \\ 0 & \text{si } k \geq 2 \end{cases} \quad (3)$$

Now we can draw the Auto-Correlation Function **ACF**. This plot is called auto-correlogram. The partial auto-correlation function exists in order to define to which extent of shift, there is a direct relation between y_t and the preceding values.

MA of order q We say that the series y_t is a process of moving average of order q and we note $MA(q)$, if it is generated by a white noise ε_t under the form:

$$y_t = \varepsilon_t - \theta_1\varepsilon_{t-1} - \theta_q\varepsilon_{t-q}$$

$$= \theta(L)\varepsilon_t$$

We suppose obviously that q is inferior to the number of observations. Auto-correlation of y_t . We can demonstrate that:

$$\begin{cases} \rho(k) \neq 0 & \text{si } k \leq q \\ \rho(k) = 0 & \text{si } k > q \end{cases} \tag{4}$$

In summary: To recognise whether a series y_t is a process of Moving average of order q , we first draw the auto-correlogram of this series. For each values of correlation, we shall have an interval of confidence on that value. If we can decide that till a shift q , those values are different from 0, and that they are statistically different from 0, then we could say that the series follows a process $MA(q)$. We can also realise that such a series is stationary.

Auto-regressive processes

AR of order 1 We say that a series y_t follows an auto regressive process order 1 and we note $AR(1)$, if we can write:

$$y_t = \phi y_{t-1} + \varepsilon_t$$

$$\varepsilon_t = y_t - \phi y_{t-1}$$

$$\varepsilon_t = (1 - \phi L)y_t$$

where the series ε_t is a white noise. We can also notice that we do a regression of shifted series by 1 on the series itself and the residuals are white noise. The function of transfer can be written as:

$$y_t = \varepsilon_t + \phi y_{t-1}$$

$$y_t = \varepsilon_t + \phi y_{t-1} + \phi^2 y_{t-2}$$

$$y_t = \varepsilon_t + \phi y_{t-1} + \phi^2 y_{t-2} + \dots + \phi^k y_{t-k} \dots$$

The function of transfer has an infinite number of terms. If we come back to our formal writing under the form of a polynomial, we notice that we have calculated the inverse of the polynomial $(1 - \phi L)$ and that

$$(1 - \phi L)^{-1} = \sum_{i=0}^{\infty} \phi^i L^i \tag{5}$$

The variance of y_t is obtained by simple calculations as follows:

$$Var(y_t) = \frac{\sigma^2}{1 + \phi^2} \tag{6}$$

We perform the variance reduction through modelling under the condition that ϕ is taken in absolute value less than 1.

Also we can show that as far as the auto-correlation of y_t is concerned: Moreover, we can demonstrate that $\rho(k) = \phi^k$ for $k \geq 1$. We obtain two types of correlograms depending whether ϕ is positive or negative.

Stationarity: the process $AR(1)$ is stationary only if the absolute value of ϕ is less than 1, *i.e.* $|\phi| < 1$.

AR of order p A series y_t follows an auto regressive process order p and we note $AR(p)$, if we can write:

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

$$\varepsilon_t = (1 - \phi_1 L + \phi_2 L^2 + \dots + \phi_p L^p)y_t$$

$$\varepsilon_t = \Phi(L)y_t$$

where the series ε_t is a white noise. Here again, we do a regression of p series shifted on the series itself.

Auto-correlation: For a process $AR(p)$, we cannot say anything of its function ACF , if it is only for $\rho(k) \neq 0$ for all k . It is in order to recognise an auto-regressive process that we have introduced the partial auto-correlation function. The partial auto-correlation with shift k ($PACF(k)$) is defined as being the correlation between:

The residual of the regression of the series y_{t+k} by the series: $y_{t+1}, y_{t+2}, \dots, y_{t+k-1}$ and ;

The residual of the regression of the series y_t by the series: $y_{t+1}, y_{t+2}, \dots, y_{t+k-1}$. In other words:

$$y_{t+k} = \alpha_1 y_{t+1} + \alpha_2 y_{t+2} + \dots + \alpha_{k-1} y_{t+k-1} + u \tag{7}$$

$$y_{t+k} = \beta_1 y_{t+1} + \beta_2 y_{t+2} + \dots + \beta_{k-1} y_{t+k-1} + v \tag{8}$$

$$y_{t+k} = Corr(u, v) \tag{9}$$

We should understand that the partial auto-correlation is the correlation between y_t and y_{t+k} , once we have explained those by the values between the both $y_{t+1}, y_{t+2}, \dots, y_{t+k-1}$.

Duality: We introduce a table of duality between process MA and AR :

Table 1. Duality between MA and AR

	AR(p)	MA(q)
ACF		= 0 for all $k > q$
PACF	= 0 for all $k > p$	

We recognise that a series follows a process $MA(q)$ if its Auto-Correlation Function ACF cancels from a shift q or that it follows a process $AR(p)$ if its Partial Auto-Correlation Function $PACF$ cancels from a shift p .

Mixed processes

ARMA processes We say that a series y_t follows a process $ARMA$ order (p, q) and we note that $ARMA(p, q)$, if we can write:

$$y_t - \phi_1 y_{t-1} - \dots - \phi_p y_{t-p} = \varepsilon_t - \theta_1 y_{t-1} - \dots - \theta_q y_{t-q} \tag{10}$$

$$\Phi(L)y_t = \Theta(L)\varepsilon_t \tag{11}$$

where the series ε_t is a white noise.

The process ARMA(1,1) While taking $p = 1$ et $q = 1$, the series follows $ARMA(1, 1)$ an process if we can write:

$$(1 - \phi L)y_t = (1 - \theta L)\varepsilon_t \tag{12}$$

The properties of an $ARMA(1, 1)$ are:

- For the series y_t to be stationary and invertible (i.e., that we may calculate ε_t according to y_t), we must have: $|\phi| < 1$ and $|\theta| < 1$.
- The ACF of an $ARMA(1, 1)$ decreases exponentially from $k = 1$, whereas the ACF of an $AR(1)$ decreases exponentially from $k = 0$.
- The $PACF$ of an $ARMA(1, 1)$ resembles that of an $MA(1)$ from $k = 2$.

We can say, roughly, that if we are not obviously in front of a process AR or MA , it is very likely to find ourselves before an $ARMA$ process.

ARIMA process A series y_t follows an **ARIMA** process (Auto-Regressive Integrated Moving Average) order (p, d, q) order, if it follows an **ARMA** order $(p + d, q)$:

$$\Phi(L)y_t = \Theta(L)\varepsilon_t \tag{13}$$

where the value $L =$ is root of order d of the polynomial $\Phi(L)$. We model the series then under the form:

$$\Phi(L)(1 - L)^d y_t = \Theta(L)\varepsilon_t \quad (14)$$

where the polynomial $\Phi(L)$ is the degree p and the polynomial $\Theta(L)$ is the degree q . We shall say that the series y_t follows an $ARIMA(p, d, q)$ process.

Remarque 2.1 A y_t series following the $ARIMA$ process is not stationary since we must apply the operator of differentiation to have a white noise generating it.

2.2 Principle of Construction of an ARMA Model

Principle of construction of a model The Box- Jenkins models were introduced [3] in the economy in 1973 with an article having significant impacts. As its name implies those models contain two processes: AR for Auto-Regressive and MA for Moving Average. Two parameters (p, q) defining the model: p is the degree of auto-correlation and q for the degree of the moving average. The main problem is that of the strategy of modelling, i.e the choice of parameters p and q . A process in three steps must be adopted.

- **Identification:** It deals with the analysis of functions of auto-correlations and partial auto-correlation to see if the series is stationary or not; to see if the stationary process a white noise follows, a moving average or an autoregressive structure; and finally to identify the order structures.

- **Estimation:** It consists to use the packages already available to evaluate the parameters of the model. Those parameters must meet two conditions: they must be in the limits of invertibility and stationarity. They must be statistically significant. If they do not meet those conditions, we should restart the process. In this section the main interest is laid upon techniques of estimation and models of evaluation.

The techniques of estimation and models of evaluation are:

- Criterion of log-likelihood: We examine the log of probability of results from the estimated parameters.
- Akaike Information Criterion (AIC): statistics permitting to decide on the order of the model. The AIC criterion regards both the quality of adjustment of the model to the observed series and the number of parameters used in the adjustment. Seek the model describing the series adequately and whose AIC criterion is a minimum.

- Schwartz Bayesian Criterion: statistics helping to decide on the order of a model. The SBC regards both the quality of adjustment of the model to the observed series and the number of parameters used in the adjustment. Seek the model describing the series adequately with a minimum SBC. The SBC is based on Bayesian Criteria (Maximum of Likelihood)

- **Diagnostic:** The diagnostic is relating to the statistic adequation of the model. This one leans on an analysis of remainders which must follow a white noise. A Q statistic allows to test if the set of the auto-correlation process is different from the white noise.

$$Q(ddl) = n \sum_{i=1}^k ACF_i^2 \quad \text{o} \ddot{A} \acute{z} \quad ddl = k - p - q$$

This statistic distributes according to a distribution of χ^2 .

Validation of the parameters of the model The tests that are undergone by the models are of two types: the tests regarding the parameters ϕ_i and θ_i of the model and those regarding the hypothesis done on ε_t .

- Tests concerning the parameters: To compare a formulation $ARMA(p, q)$ with a formulation (p_1, q_1) ; it is practical to be placed in the case where one of the models is a particular case to the other. We suppose in the following that we start from an $ARMA(p, q)$ model and examine the tests corresponding to diverse values of p_1, q_1 .

- * $p_1 = p - 1, q_1 = q$. We want to check whether it is possible to reduce by one unit the number of delays intervening in the auto-regressive part. This to test the significance of the coefficient ϕ_p , which can be done with a test of Student type.

- * $p_1 = p + 1, q_1 = q$. We want to check whether it is necessary to introduce an extra delay in the

auto-regressive part. We must then estimate the model $ARMA(p_1, q_1) = ARMA(p + 1, q)$ and test in this second model the nullity of ϕ_{p+1} . This obviously lead us to the previous case.

• Tests concerning the residuals: The issue is to know whether the hypothesis on normality and white noise (independence of residuals are satisfied).

Choice of a model Several models might pass the verification phase et we must choose in that set. The first thing to be done is doubtlessly to identify the «twins» models, i.e those corresponding in fact to different representations of a model (or close models and that are equivalent for the forecasting having chosen a representative in each twin class, we can be found at the head of several models really different. The choice of a model will be then done according to some criteria.

Criteria of predictive power: The objective is to minimise the residual variance, to maximise the simple and modified coefficient of determination. The coefficient of determination being defined as follows:

$$R^2 = 1 - \frac{\hat{\sigma}^2}{v}, \quad v \text{ which is the empirical variance of the differentiated series } d \text{ times.}$$

$$\text{This is modified as follows: } \bar{R}^2 = 1 - \frac{\hat{\sigma}^2 / (N - p - q)}{v / (N - 1)}.$$

Criteria of information: Another approach, introduced by Akaike, 1969, consists in assuming that models $ARMA(p, q)$ provide approximations of the reality and the true unknown law of observations $\nabla_{y_t}^d$. Inevitably satisfy such a model. We can then base our choice of a model on a measurement of the gap between the unknown law and the proposed model.

The measurement usually retained is the quantity of information of Kullback. Either $f_0(y)$ the unknown density of observations $\{f(y), f \in F_{p,q}\}$, the family of densities corresponding to the model $ARMA(p, q)$ the gap between the true law and the model is measured by:

$$I(f_0, F_{p,q}) = \min_{f \in F} \int \ln \frac{f_0(y)}{f(y)} f_0 dx$$

The aim is then to find the p and q which minimise the estimator of this quantity of information. And we propose some:

- $AIC(p, q) = \ln(\hat{\sigma}^2) + \frac{2(p+q)}{N}$; (Akaike, 1969)
- $BIC(p, q) = \ln(\hat{\sigma}^2) + (p+q) \frac{\ln N}{N}$; (Akaike, 1977)
- $\rho(p, q) = \ln(\hat{\sigma}^2) + (p+q)c \frac{\ln N}{N}$; (avec $c > 2$) (Hannan – Quinn, 1979)

Let's note that only the estimations p and q deduced of the last two criteria are convergent to a selection asymptotically correct of the model for further details, see [10].

Forecasting A large number of phenomena changing in time, in an underlying, cyclic or irregular manner, the statistic methods of analysis of chronological series permits to describe, decompose and foreseeing the evolution of those phenomena either economic or meteorological or linked to a disturbance of a production process. The **R** software besides its graphic possibilities of a large number of functions allowing the exploitation of chronological data.

The necessary conditions being sufficient (stationarity, white noise etc), we proceed to the forecasting which deals greatly with several economic structures. Having observed y_1, y_2, \dots, y_n , we can foresee that the future values $y_{n+1}, y_{n+2} \dots$. For that we shall essentially use the decomposition models to do the forecasting, and the methods of this work aim to do forecasting in the short term. We understand then by forecasting to the horizon h which provides an evaluation of the value of the series on the date $t = n + h$ (n =number of observations)

Error of forecasting Generally having carried out forecasting of the series at certain horizon, it is appropriate to estimate their errors in order to appreciate the performance of the model. Thus, we can then represent the errors of forecasting and calculate the amount of errors of forecasting squared called MSE (Mean Squared Error) define by:

$$MSE = \sum_{t=1}^n err^2(t)$$

where $err(t) = (y_t - \hat{y}_t)$ are called errors of forecasting which must be minimal for a good model.

3 Applications and Simulations

In this section, we shall proceed with direct applications on the real data of credit and debit of pension funds of the NSIF. We adopt the approach proposed in [3] by Box Jenkins to model a chronological series presenting quite a few factors with all the necessary conditions besides, with the help of an $ARMA(p, q)$ model. We apply then the packages available in the **R** software to forecast the series in a horizon h to be determined.

3.1 Presentation of the Data

The data submitted to our study are the collection of the credit and debit of the PVID branch contained in the statistical yearbook of the NSIF Cameroon. They are given in months and calculated in billion francs CFA. Also They represent a sample of data collected over seven years taken from 2010 to 2016.

The credits The following table gives in billion francs CFA and by months, the credits of the pension funds of the NSIF (from 2010 to 2016).

Table 2. The credits

	2010	2011	2012	2013	2014	2015	2016
January	2,168	4,438	4,642	2,843	2,843	2,843	5,781
February	3,181	1,256	1,954	4,842	1,642	1,845	7,507
March	5,181	4,025	5,425	2,745	1,225	3,545	4,854
April	3,931	5,152	5,602	3,845	3,845	5,543	3,516
May	2,381	3,548	4,245	6,452	2,954	2,605	7,845
June	3,161	2,685	1,856	4,985	5,845	6,842	6,145
July	1,505	4,454	6,514	2,895	6,851	4,315	8,648
August	2,109	1,964	3,335	3,654	2,364	3,854	6,845
September	5,523	5,543	4,985	5,845	2,845	5,654	5,945
October	8,337	7,284	4,985	7,235	8,554	6,464	9,654
November	4,085	2,841	1,287	1,264	3,684	1,541	6,257
December	2,495	3,685	3,508	3,254	4,657	4,34	4,608
Total	44,057	46,875	48,338	49,859	47,309	49,391	77,605

(Source : *statisticals yearbook – NSIF – 2010 – 2016*)

Table 3. Characteristics of credits

Average	variance	standard deviation	max	min	extent
4.055	3.85447	1.96328	9.654	1.225	8.429

The debits The following table gives in billion francs CFA and by months, the debits of the pension funds of the NSIF (from 2010 to 2016).

Table 4. The debits

	2010	2011	2012	2013	2014	2015	2016
January	4,543	5,158	6,452	3,365	6,158	3,158	7,025
February	2,654	2,645	2,354	5,215	2,344	2,154	6,154
March	5,688	4,985	5,152	3,254	4,325	3,284	5,125
April	3,915	3,654	4,268	3,854	6,216	5,168	6,125
May	4,385	5,454	3,354	7,368	3,215	3,258	4,258
June	2,034	3,154	3,294	5,295	6,802	6,348	7,154
July	3,818	5,921	5,345	3,158	7,254	5,258	4,318
August	3,615	3,542	4,365	5,645	3,158	3,154	6,215
September	5,524	5,963	5,235	5,134	3,645	5,157	4,258
October	7,524	2,452	5,845	6,259	7,166	7,365	8,155
November	3,145	4,845	2,366	3,294	4,215	3,125	5,214
December	5,254	6,658	5,365	6,158	7,658	6,315	10,612
Total	52,099	54,431	53,395	57,999	62,156	53,744	74,613

(Source : *statistics yearbook – NSIF – 2010 – 2016*)

Table 5. Characteristics of debits

Average	variance	standard deviation	max	min	extent
4,862	2,742844	1,656153	10,612	2,034	5,578

3.2 Study of the Normality

We present in this paragraph the normality tests of real data by the test of Shapiro, so the package is available in the software **R** and explained in [1]. The table 6 contains values of tests and critical probabilities of the normality test of Shapiro. Let's remind that the null hypothesis H_0 is the normality and the alternative hypothesis H_1 is the non normality. We reject the null hypothesis if the critical probability is lower to a risk α generally fixed at 5%. At the threshold of $\alpha = 5\%$, only the credits pretend to be normal, due to their critical probability $P_c = 0.0527$, which are higher than α . The expenses, having a critical probability $P_c = 0.01713$, will in practice need other transformations that can normalise them. We can also check this test using histograms or by using box plot, thus the concentration of observations around the average testifies of the normality of the data. For example, in the box plot diagram, we notice that the blue band of credits is quasi-central and the near the average value (4,055), which shows their normal character.

Table 6. Normality test of Shapiro

Series	Statistic value of the test	Critical probability
Credits	0.97075	0.0527
Debits	0.96332	0.01713

3.3 Stationarity

This paragraph we check whether our data are stationary or not. To this end, we use the Augmented Dickey-Fuller test. Let's remind that with the Dickey-Fuller test, the null hypothesis is the non-stationarity by default. We reject it then if the critical probability is lower than the threshold α , which is generally fixed at 5%. The following table 7 summarize the main characteristics of this test.

At the threshold $\alpha = 5\%$, every critical probability of the stationarity test of Dickey Fuller presented in the table below are lower to this value. Which means that all our series are stationary. Let's remind that

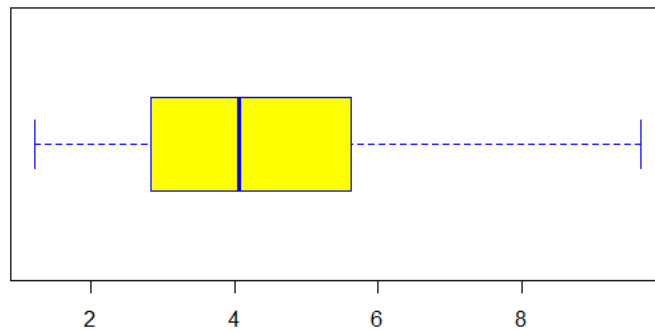


Figure 1. Box of credits

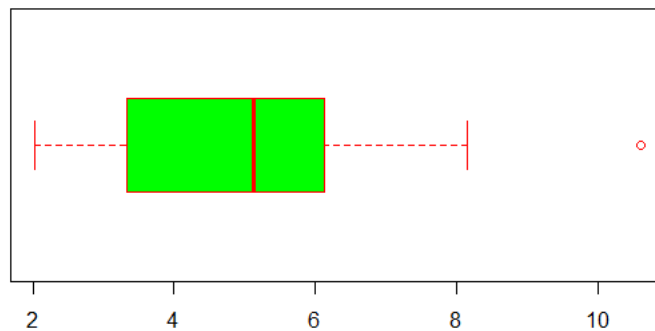


Figure 2. Box of debits

Table 7. Critical probability of the Dickey-Fuller test

Series	Statistic value of the test	Critical probability	Decision
Credits	-3.4776	0.04907	Stationary
Debits	-4.4867	0.01	Stationary

a stationary series fluctuate around an average value and its auto-correlation function declines quickly towards zero.

3.4 Seeking the parameters p, q

We present the correlograms issued in the studied series. The adequate analysis of correlograms, allows us to determine those maximal orders of p and q of Auto-Regressive and the Moving Average parts of the model respectively. Actually those are, the ranks to be counted from 0 to the ACF (which determines q_{max}) and from 1 to the $PACF$ (which determines p_{max}) after which threshold of estimated auto-correlations

of the process are non-significant to the threshold of 5%. Graphically that is illustrated by two dotted line bands in blue determining the confidence interval of estimated coefficients of those parameters. The reading of the stationary series correlograms give the table 8 which recaps the orders maximal p and q read on these latter.

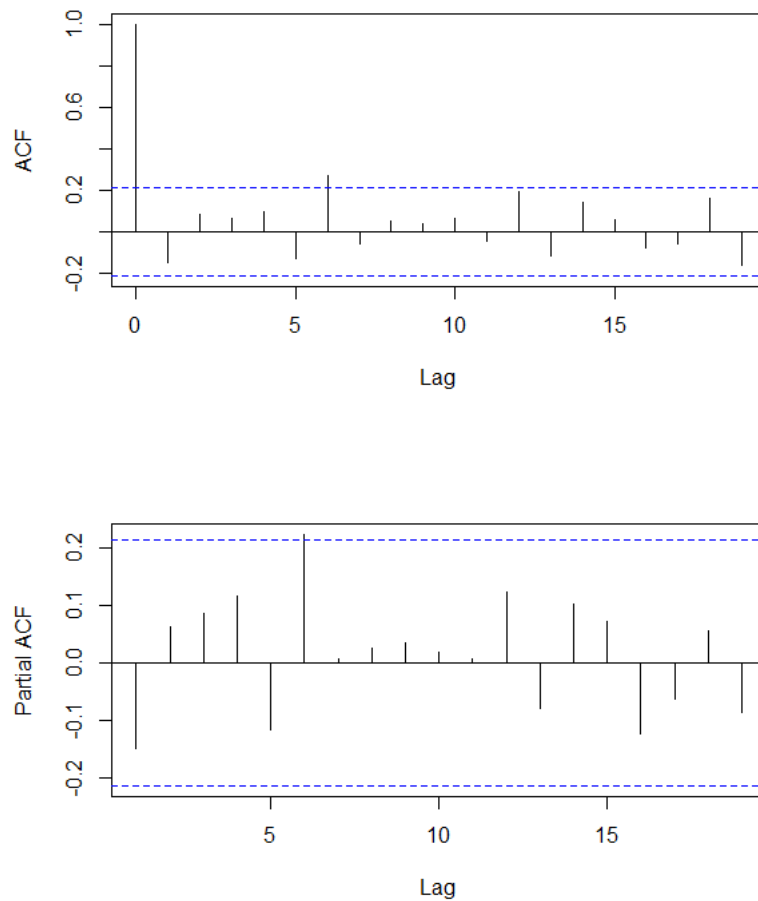


Figure 3. Correlogram of credits

Table 8. Maximal orders of p and q

Orders	p	q
Credits	6	7
Debits	6	7

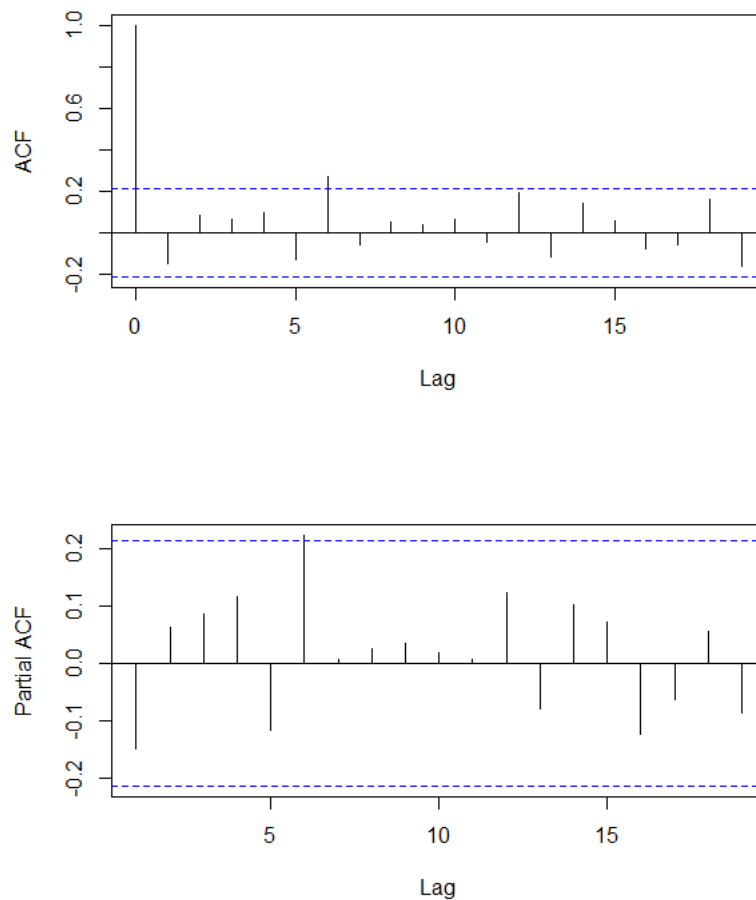


Figure 4. Correlogram of debits

3.5 Estimation of the Model

On second thoughts, every series being stationary, it is our task to find the “good” couple (p, q) compatible with the data and to be estimated with an $ARMA(p, q)$ the coefficients of the model then chosen. The model could then be written:

$$\Phi(L)y_t = \Theta(L)\varepsilon_t$$

where ε_t is a white noise of variance σ^2 and $\Phi(L) = 1 - \phi_1L - \phi_2L^2 - \dots - \phi_pL^p$, $\Theta(L) = 1 + \theta_1L + \theta_2L^2 + \dots + \theta_qL^q$ with $\phi_1, \phi_2, \dots, \phi_p$; $\theta_1, \theta_2, \dots, \theta_q$ and σ^2 are the parameters to be estimated. For the methods of estimation of those parameters (cf [3]). Very delicate and expensive exercise since the significance of the parameters p et q must be tested, and to verify the hypothesis of noises of residuals. The table 9 summarizes the characteristics of the estimated models at the end of this work.

Table 9. Estimated models

Series	Estimated models	Box-Ljung	AIC
Credits	ARMA(6,7)	0,1197	358,5
Debits	ARMA(6,1)	0.1741	328,16

3.6 Validation of the Model

The analysis of the residuals shows that they are compatible with the hypotheses of white noise. Let's keep in mind that the hypotheses of this proposed test by Box-Pierce and improved by Ljung-Box [3] is: H_0 : $\hat{\varepsilon}_t$ is a white noise against the alternative hypothesis H_1 : $\hat{\varepsilon}_t$ is not a white noise. At the threshold $\alpha = 5\%$, every critical probability of table 9 above are higher than this value. Likewise, the totality of residuals are from Gauss point of view white noise, all the models above are validated, but the question that one asks is to know if the estimated coefficients of those models are significant? Which leads us to analyse the following tables containing estimated coefficients.

Table 10. Coefficients of credits

orders	ar1	ar2	ar3	ar4	ar5	ar6
coefficients	0.2314	0.2278	0.2545	-0.2344	-0.3242	0.6835

orders	ma1	ma2	ma3	ma4	ma5	ma6	ma7
coefficients	-0.0807	-0.2631	-0.1468	0.4514	0.5379	-0.6213	-0.2891

Table 11. Coefficients of debits

orders	ar1	ar2	ar3	ar4	ar5	ar6	ma1
coefficients	0.0256	0.1020	0.0724	0.1243	-0.088	0.3100	-0.1559

The exploration of estimated coefficients generally presents their significance. That's why it is very important to note that with the program of AIC criterion set up, we also obtain a standard deviation σ generally minimal which is one the most powerful criteria of choice of parameters p and q too. Our model can generally be written as:

$$\Phi(L)y_t = \Theta(L)\varepsilon_t \quad (15)$$

with $\Phi(L) = 1 - \phi_1L - \phi_2L^2 - \dots - \phi_pL^p$; $\Theta(L) = 1 + \theta_1L + \theta_2L^2 + \dots + \theta_qL^q$; p and q chosen in the table 8 of the corresponding value.

3.7 Forecasting

The main objective of the approach taken in this chapter was to forecast credits and debits of the NSIF pension funds. A large number of criteria being followed by our data, we think carrying out forecasting in a two-year horizon is possible. Either from 2017 to 2018. For this end, by using the **R** software, we obtain the following graphs and tables, with their forecast values and the forecast intervals.

According to the analysis of these tables, we found that the financial equilibrium observed in 2016 will be maintained until 2017 with a settlement of 892,906 millions FCFA. However, the imbalance will reappear in 2018 again, with a deficit of approximately 2,375 billion FCFA. On the account of these results, it is extremely important, to find management strategies and financing of funds, in order to anticipate on the future situation of the funds.

The figure 4 represents the temporal dynamic of forecasting. The curve in blue represents the evolution of credits. While that in red represents the evolution of debits. We can then from those two curves, see at which precise moment a deficit can occur or not.

3.8 Qualitative Analysis of the Forecasting

The objective of this section is to study the quality of model forecasting.

Table 12. Foreseen values of credits

Months	Lower bound	Forecasting	Upper bound
Jan 2017	5,707779	6,568132	7,428484
Feb 2017	5,232856	6,121807	7,010759
Mar 2017	5,705094	6,622364	7,539633
Apr 2017	6,422133	7,367448	8,312764
May 2017	6,155569	7,128666	8,101763
Jun 2017	6,332070	7,332691	8,333312
Jul 2017	5,756699	6,784594	7,812489
Aug 2017	4,080789	5,135714	6,190640
Sep 2017	5,842692	6,924411	7,006130
Oct 2017	7,176252	9,284534	10,392816
Nov 2017	4,417646	4,552302	5,686958
Dec 2017	4,251726	5,412500	6,573274
Total	67,158648	79,235163	91.054985
Jan 2018	6,381454	7,534842	8,754809
Feb 2018	5,909435	7,584211	8,334180
Mar 2018	4,384501	5,854741	6,860227
Apr 2018	6,104293	7,321562	8,630603
May 2018	6,840414	7,896512	8,416918
Jun 2018	7,019532	8,254136	9,645850
Jul 2018	6,446714	7,251384	8,122475
Aug 2018	4,773294	5,452305	6,498134
Sep 2018	6,537629	7,054842	8,311193
Oct 2018	7,873565	9,856145	10,695503
Nov 2018	2,117281	4,895426	5,987322
Dec 2018	2,953629	6,254865	7,871371
Total	75,950456	85,210971	95,751268

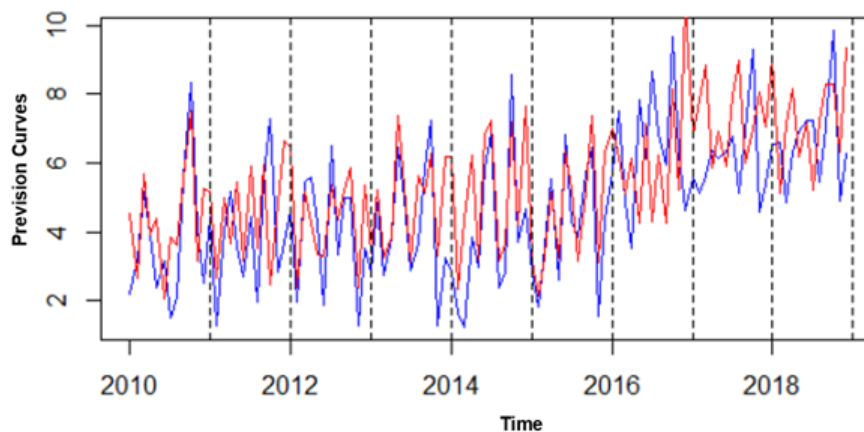


Figure 5. Time evolution of forecasting

Method

The study of this quality is generally done through a good analysis of the theoretical model used to fit the data.

The idea consists to suppose that we have n consecutive periods.

- We suppose that the theoretical model can be applied with k periods, $k < n$.

Table 13. Foreseen values of debits

Months	Lower bound	Forecasting	Upper bound
Jan 2017	5,116231	6,803078	7,489925
Feb 2017	6,124826	7,826647	8,528468
Mar 2017	7,133421	8,850216	9,567011
Apr 2017	4,42015	5,873784	6,605554
May 2017	5,150609	6,897353	7,644097
Jun 2017	4,159203	5,920922	6,682640
Jul 2017	6,167797	7,944490	8,721184
Aug 2017	7,176390	8,968059	9,759728
Sep 2017	4,184983	5,991628	6,798273
Oct 2017	6,193575	7,015197	8,836818
Nov 2017	7,202167	8,038765	9,875363
Dec 2017	6,210759	7,062334	8,913909
Total	67,156482	78,342257	89,851564
Jan 2018	8,219350	9,085903	10,952456
Feb 2018	4,227940	5,109471	6,991003
Mar 2018	6,236530	7,133040	8,029550
Apr 2018	7,245119	8,156609	9,068098
May 2018	5,253708	6,180178	7,106647
Jun 2018	6,262296	7,203746	8,145197
Jul 2018	4,270883	5,227315	6,183747
Aug 2018	6,279469	7,250884	8,222298
Sep 2018	7,288055	8,274453	9,260850
Oct 2018	7,296639	8,298021	9,299403
Nov 2018	5,305223	6,321590	7,337957
Dec 2018	8,313806	9,345159	10,376511
Total	76,453619	87,586369	98,564208

- We compare the theoretical period of forecasting for the $k + 1$ to n with the real data of $k + 1$ to n .
- We measure concretely the error margin between the theoretical model and the real data.

Application

We apply here, the theoretical model of forecasting on the data of credits and debits from 2010 to 2015. We can then check the margin error between the theoretical model and the real data.

The results obtained at the end of this work, are deposited in the following figures. These are the screen shots of the scripts in the console of the software **R**.

The obtained results from the theoretical model are very much closed to those of the real model (data). We can confirm that the margin error between the theoretical model and the real data is very small.

4 Conclusion

In this article, we deal with modelling a system of forecasting credits and debits of the pension funds of the NSIF Cameroon. This research has helped us to get involved with the chronological series in their whole set and especially in their prediction aspect. The ultimate goal in our context is to obtain a reliable prediction fitting the arrangement between the knowledge of the problem of a given system, the understanding of the internal and external influences and the adequate responses to the concern related to the approach being used. We were able to perform a prediction in a two-year horizon starting from 2017 to 2018. We wish to extend our investigation by associating in this approach some additional internal as well as external factors, that can influence the evolution of our variables or chronological series. In this respect, we have to consider the number of retired people, the number of subscribers, the rate of

```

> rp<-arima(tab6,order=c(6,0,2))
> p1<-predict(rp,n.ahead=6)
> p2<-predict(tab6)
> p2

```

	Point	Forecast	Lo 80	Hi 80	Lo 95	Hi 95
Jan 2016	4.838657	3.230588	6.446726	2.379327	7.297987	
Feb 2016	3.116492	2.080764	4.152220	1.532482	4.700502	
Mar 2016	4.299362	2.870521	5.728203	2.114138	6.484586	
Apr 2016	4.565386	3.048135	6.082637	2.244951	6.885821	
May 2016	4.579000	3.057225	6.100775	2.251646	6.906355	
Jun 2016	4.864018	3.247520	6.480516	2.391798	7.336238	
Jul 2016	5.160333	3.445359	6.875308	2.537506	7.783160	
Aug 2016	3.959727	2.643759	5.275694	1.947128	5.972325	
Sep 2016	4.940338	3.298476	6.582200	2.429327	7.451349	
Oct 2016	6.219357	4.152429	8.286285	3.058263	9.380450	
Nov 2016	3.648494	2.435961	4.861027	1.794085	5.502903	
Dec 2016	6.076695	4.057179	8.096212	2.988112	9.165279	

Figure 6. 2016 credits

```

> rp<-arima(tab6,order=c(6,0,3))
> p1<-predict(rp,n.ahead=6)
> p2<-predict(tab6)
> p2

```

	Point	Forecast	Lo 80	Hi 80	Lo 95	Hi 95
Jan 2016	6.575236	6.103841	7.046632	5.854299	7.296173	
Feb 2016	5.767880	5.347889	6.187871	5.125560	6.410201	
Mar 2016	4.969616	4.602255	5.336977	4.407785	5.531446	
Apr 2016	5.947864	5.501707	6.394021	5.265526	6.630202	
May 2016	3.943025	3.643018	4.243032	3.484204	4.401847	
Jun 2016	6.668967	6.154490	7.183444	5.882142	7.455792	
Jul 2016	4.085978	3.766495	4.405461	3.597370	4.574585	
Aug 2016	5.859256	5.395076	6.323436	5.149354	6.569158	
Sep 2016	4.073830	3.746946	4.400714	3.573904	4.573755	
Oct 2016	7.752814	7.122931	8.382696	6.789492	8.716136	
Nov 2016	4.917014	4.512643	5.321385	4.298582	5.535446	
Dec 2016	8.707948	7.983265	9.432631	7.599642	9.816254	

Figure 7. 2016 debits

the subscription, the age of the departure to retirement, the evolution of the life expectancy and many others. This will allow us to put in place a more sophisticated model of forecasting credits and debits of the pension funds of the NSIF in Cameroon and in developing countries as a whole.

References

1. Pierre Lafaye de Micheaux, Remy Drouilhet, Benoit Liqueur (2011): *Le logiciel R*. Collection Statistiques et probabilités appliquées.
2. Ljung G.M., Box G.E.P.(1978). *On a Measure of the Lack of Fit in Time Series Models* . Biometrika, vol. 65, pp. 297-303.
3. Box G.E.P., Pierce D.A.(1970): *Distribution of Residual Autocorrelations in Autoregressive Moving Average Time Series Models*. Journal of the American Statistical Association, vol. 65 .
4. CNPS.(2015): *Cahier de charge de la CNPS*. Yaoundé.
5. Ilo (2011): *Rapport à la Caisse nationale de Prévoyance sociale : évaluation actuarielle du régime de sécurité sociale géré par la Caisse nationale de Prévoyance sociale du Cameroun au 31 décembre 2008 / Bureau international du Travail, Département de la Sécurité sociale*. Genève: BIT, 2011 xi. 122 p. Boulier, J-F , Dupré.D (2002) *Gestion financière des fonds de retraite*. 2e édition Economica.
6. Serban, M., Brockwell A., Lehoczky J., Srivastava S. (2006). *Modelling the dynamic dependence structure in multivariate financial time series* . Journal of time series analysis, vol28, No5, 763-782.
7. Emmanuel César, Bruno Richard.(2006): *Les Séries Temporelles*. Université de Versailles Saint-Quentin-en-Yvelines.
8. Yves Aragon (2011) : *Séries temporelles avec R, méthodes et cas Pratique* . Springer, Paris, With a preface by Dominique Haughton.
9. Wilkie (1995): *More On a stochastic asset model for actuarial use*. British Actuarial Journal, Volume 1, Number 5, pp. 777-964(188).

10. Guy Mélard (1990): *Méthodes de prévision à court terme*. Éditions de l'Université de Bruxelles, 1990.