International Journal of Management and Economics Invention ISSN (e): 2395-7220
||Volume||3||Issue||12||Pages-1480-1492||Dec.-2017||
Index Copernicus ICV: 72.76, DOI: 10.18535/ijmei/v3i12.01

The Water Challenge in Tunisia: Crisis or Détente Application of NIPALS Algorithm and Box and Jenkins Methodology

Imen Gam[1], Jaleleddine Ben Rejeb[2]

[1]Phd Student, LAMIDED laboratory, Sousse University.
[2]Quantitative Methods Professor. LAMIDED laboratory, Sousse University

Abstract: Two paradoxical analyses of the water sector in Tunisia can be advanced. According to the World Bank's report about the Tunisian water reserves published in 2009 "Tunisia is preparing to face, during the next decades, too many important water access problems, arising from growing demand and a decrease of supply. Institutions will also face more complex management problems" This orientation which highlights the scale of the water crisis in Tunisia is confronted with a totally contradictory statement of Abderrazek Souissi the General Director of the Office of Planning and the Hydraulic Equilibrium in the Agriculture Ministry, which proclaimed that "Water is scarce in Tunisia. But until now, we well managed our resources, so we cannot talk about water crisis. "

Thus, The main object of this paper is to analyze and dissect the evolution of the availability of water in Tunisia by forecasting it using the methodology of Box and Jenkins, thus, a manager can develop effective action plans to use these resources in a more efficient and optimal manner. Prior to the forecast step, we first apply the NIPALS algorithm to impute missing data in our series.

Keywords: Algorithm NIPALS, forecasting, Box and Jenkins.
JEL Classification: Q25, C53

Introduction

Faced with limited and unevenly distributed in space and time water resources, a water shortage problem is revealed in many countries having a situation of inequality, translated by excess water demand compared to available resources. More specifically, and in the Mediterranean basin, Tunisia is among the least endowed countries with water resources.

To be more rigorous, we try to analyze the current and future status of water resources in Tunisia, referring to Box and Jenkins technique.

The methodology of Box and Jenkins known for its simplicity and effectiveness has been applied in an important package of work such as that of Maidment and al (1985). Based on monthly data from January 1978 until December 1987, Al Dhowalia KH (1996) tried to determine the evolution of water demand for the city of Riadh referring to this methodology. Applying the same procedure, Lawgali FF (2008) tried to predict the demand for water in agriculture, industry and the domestic sector in Libya. The problem is that this forecasting technique is used only with a complete dataset however; our series suffer from missing data.

The management of missing data is a fundamental issue and a key problem in statistical practice. This type of data is often a nightmare for any statistician since don’t taken into account missing data can indeed significantly affect the calculated estimators.

This problem frequently occurs in a data table (due to non-response in the case of surveys, the non-data storage,…) and the most commonly used solutions to overcome it are the definitive removal of individuals having missing data or the imputation of missing values. Despite both techniques lead to obtaining a complete table, the second one (imputation of missing data) remains the most preferable because it keeps all the observed information. A literature review shows the existence of several works that seek to search the most reliable imputation technique. A multiple imputation method was developed by Little, RJA, Rubin, DB (1987). This technique consists in replacing a missing value by plausible values m (m> 1) .This method can be described in three steps. The first step consists in assigning values for each missing data, referring to a suitable random model. We pass, thereafter; to repeat m times the first step in order to get m completed tables. In a final step, we proceed to the analysis of these m tables based on statistical methods.

$$\boldsymbol{\beta}_{\mathbf{i}}^{*} = \frac{\mathbf{1}}{\mathbf{m}} \sum_{\mathbf{j=1}}^{\mathbf{m}} \boldsymbol{\beta}_{\mathbf{i,j}}^{*}$$

Rubin has shown that even with a small number m, we can have reliable results.

Schafer, JL (1999) from his part has used more simple techniques to impute missing data such as the mean, median or mode. Tenenhaus, M. (1998) has presented a more efficient imputation technique based on NIPALS algorithm proposed by WOLD in 1966. Wasito, I., Mirkin, B., (2006) has proposed the nearest neighbor approach for an imputation based on least squares. A comparison between several imputation techniques was developed by Preda, C., et al. (2005). Similarly Preda, C., et al (2005) has used the NIPALS algorithm to determine plausible values for missing data.

In this paper, we focus on the management of missing data in the first section. We pass thereafter to trace the evolution of water availability per capita in Tunisia (m^3/capita). This analysis is crucial because it allows us to clarify the situation of Tunisia against the risk of a fresh water shortage. Can we speak about a water crisis in Tunisia? and if so, is it a temporal or permanent crisis?

We finish our paper with a conclusion including recommendations.

1. Data

To examine the future evolution of water availability in Tunisia (m^3 per capita per year), we opt to a database collected from the National Institute of Statistics (NIS). Inappropriately, this database is incomplete. So we start by completing it using a stepwise-descending sequential approach. It is obvious that the availability m^3 per capita per year is the quotient of the total resources which can be mobilized per year divided by the total annual population.

$$\text{Water availability } (m^3/\text{capita/year}) = \frac{\text{Total resources which can be mobilized}}{\text{Total annual population}}$$

Then, the first step consists in collecting a complete database tracing the evolution of the Tunisian population per year over the period 1968 – 2012 from the NIS. The analysis of the graph below shows that the Tunisian population has steadily increased during the last four decades. In 1968, Tunisian population was about 4933 thousands; by 2012 it had grown nearly 2 times to over 10780 thousands.

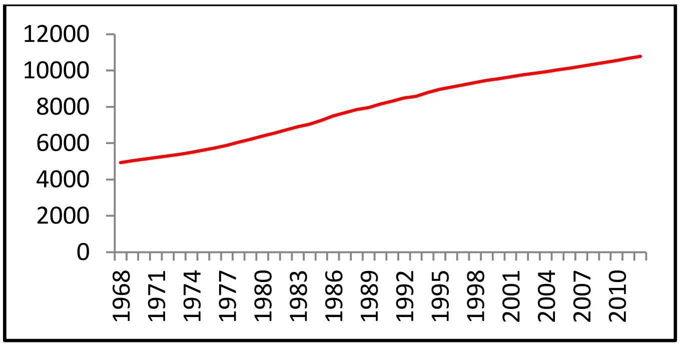


Figure 1: Evolution of the Tunisian population (thousands)

For the “Total resources which can be mobilized” series, once again, we have an incomplete database collected from the NIS and the "General Direction of Water Resources" as mentioned in the figure 2 below. An important part of this dataset is given as the percentage of "Global water potential". Thus, we must begin by collecting the series of "Global water potential".

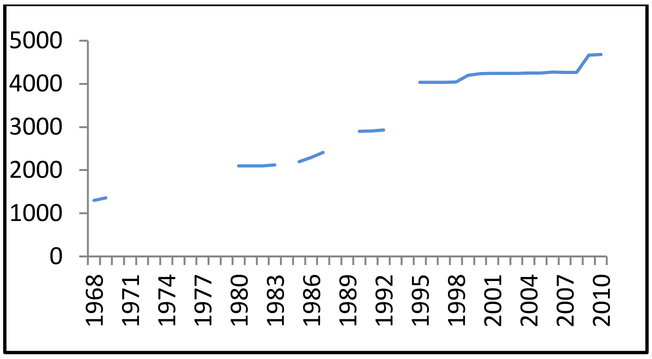


Figure 2: Evolution of Total resources which can be mobilized (million m^3)

Similarly, referring to the database provided by the "General Direction of Water Resources in Tunisia", we can collect annual data covering the period 1968 – 2010 retracing the evolution of "total water potential" and water resources by type (Deep aquifers, Shallow aquifers and Surface water). This base is required to study the future trend of "total resources which can be mobilized" and the "availability per capita". Again, we note that the database suffers from the presence of several missing data as marked by the following graph:

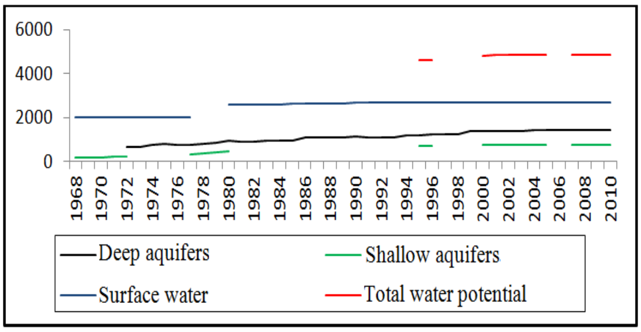


Figure 3: Evolution of Deep aquifers, Shallow aquifers, Surface water and Total water potential (million m^3)

Thus, before any statistical processing, we must start by completing the database and estimating missing values. Our strategy is organized in four steps:

- Step 1: Estimate the missing data in the data sets of: deep aquifers, Shallow aquifers and Surface water resources.
- Step 2: Complete the database of "Total Potential" knowing that:

$$Total\ Potential = Deep\ aquifers + Shallow\ aquifers + Surface\ water$$

- Step 3: Complete the database of «total resources which can be mobilized»
- Step 4: retrace the evolution of the "Availability m3 per capita per year"

- The NIPALS algorithm

One of the most serious problems that we confront in the statistical practice is the problem of missing data. Indeed, the majority of statistical methods must be applicable only with complete tables. Therefore, one of the most practiced solutions is to remove the individuals with missing data. However, this method is qualified by injudicious and estimates derived after applying this technique are unreliable especially for small-scale tables. Thus, to overcome this difficult and common problem in practice and best manage the missing data, the most affordable solution is to provide a plausible value for each missing data. The choice of this value is very important seeing that each imputed data, despite that it is uncertain, will play the same role as observed data. For this purpose, literature has developed and proposed an assortment of imputation methods.

To handle the problem of missing data in our database , we rely on NIPALS algorithm (**N**on Linear Estimation by **I**terative **P**artial **L**east **S**quares) proposed by Wold in 1966. According to Michel Tenenhaus (1998) the aim of this algorithm is "to perform principal component analysis in the presence of Missing Data."

NIPALS algorithm is based on the decomposition formula for principal component analysis as written follow:

$$X = \sum_{h=1}^{a} t_h \; P_h^{'}$$

with:

"a" is the rank of the variables-individuals matrix denoted X.

$t_h = (t_{h1}\,, t_{h2}\,, \dots, t_{hn})^{'}$ and

$P_h = \left(P_{h1}\,, P_{h2}\,, \dots, P_{hp}\right)^{'}$ represent respectively the principal factors and principal components.

Thus, individuals x_i can be expressed as follows:

$$x_i = \sum_{h=1}^{a} t_{hi} \; P_h \qquad i = 1, \dots, n$$

According to Tenenhaus Michel (1998), in the case of presence of missing data, NIPALS algorithm is written as follows:

Step 1: $X_0 = X$

Step 2: For h = 1, 2, …, a :

Step 2.1: t_h = first column of X $_{h-1}$

Step 2.2: Repeat until convergence P_h:

Step 2.2.1: For j = 1, 2, ...,

$$\text{p:} \quad P_{hj} = \frac{\sum_{\{i:x_{ji}\; and\; t_{hi}\; exist\}} x_{h-i,ji}\, t_{hi}}{\sum_{\{i:x_{ji}\; and\; t_{hi}\; exist\}} t_{hi}^2}$$

Step 2.2.2: Normalize P_h to 1

Step 2.2.3: For i = 1, 2, …,

$$\text{n :} \quad t_{hi} = \frac{\sum_{\{j:x_{ji}\; exist\}} x_{h-i,ji}\, P_{hj}}{\sum_{\{j:x_{ji}\; exist\}} P_{hj}^2}$$

Step 2.3 : $X_h = X_{h-1} - t_h P_h^{'}$

Therefore, the NIPALS algorithm estimates missing data by the mean of the reconstitution formula as described below:

$$x_i = \bar{X}_i + \sigma_i \sum_{h=1}^{a} t_{hi} \; P_{hj}$$

A necessary condition to obtain effective estimations is to not exceed 50% of missing data in the dataset.

We will begin by describing descriptive statistics for all relevant variables in our sample. The results are summarized in the table below:

Table 1: Statistics of the series "Total Potential"

	Number	**% Missing Data**	**Mean**	**Standard-deviation**
Total Potential (TP)	19	55.814	4462.33	607.45

Table 2: Statistics of the series "DA", "SA" and "SW"

	Number	**% Missing Data**	**Mean**	**Standard-deviation**
Deep aquifers (DA)	40	6.97674	1102.03	259.264
Shallow aquifers (SA)	24	44.186	558.647	228.028
Surface water (SW)	41	4.65116	2579.27	165.747
% of Total Missing data	19%			
R^2	0.954			

Table 1 shows that missing data in the "Total Potential" series far exceeded the threshold of 50%. Whereas, this threshold is respected by other series (Shallow aquifers series, Deep aquifers, Surface Water). In addition, the first principal component has the maximal explanatory power (about 0.954). Thus, and in order to obtain a consistent and efficient estimates we proceed as follows:

- We apply, in a first step, the NIPALS algorithm on a matrix composed of three series namely: deep aquifers, Shallow aquifers and Surface water resources in order to propose a plausible value for each missing data.

- We pass, thereafter, to complete the series "Total Potential" by using the following formula:

Total Potential = deep aquifers + Shallow aquifers + surface water

We follow the approach of the NIPALS algorithm already described above. We first determine the components t_{hi} and P_{hj}. We note that the first principal component captures the maximum of the total variance. Results are summarized in the following table:

Table 3: Values of Ph

	P_h
deep aquifers	0.631299
Shallow aquifers	0.521421
Surface water	0.574092

Values of t_h are given in Annex 1. Thus, to estimate missing values we need to solve the following equations:

Deep aquifers $_i$ = 1102.03 + (0.631299) (t_i) (259.264)

Shallow aquifers $_i$ = 558.647 + (0.521421) (t_i) (228.028)

Surface water $_i$ = 2579.27 + (0.574092) (t_i) (165.747)

Once missing values are imputed, we can deduce missing data in the vector "Total Potential" and replot another time the evolution of the different series "DA", "SA", "SW" and "TP" (see figure 4). The "Total Potential" series indicate that the total available volume of water in Tunisia has shown a small rise recording an average annual growth rate of around 1.5% per year.

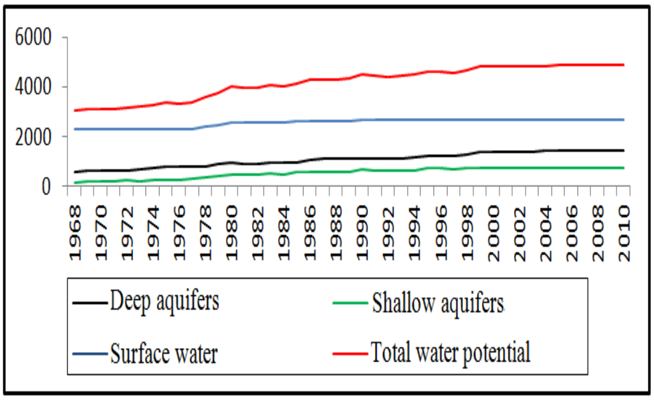


Figure 4: Evolution of deep aquifers, Shallow aquifers, Surface water and Total Potential (million m3)

After imputation of missing data in the "Total Potential" series, we pass in a second step to reapply the same procedure in order to complete the dataset of "total resources which can be mobilized per year". We proceed, then, as follows:

- We replace the percentages of "total potential" in the "total resources which can be mobilized per year" series by their exact values.

- We apply, thereafter, NIPALS algorithm on a matrix containing four series: "deep aquifers", "Shallow aquifers", "Surface water" and "total resources which can be mobilized per year" to determine the rest of missing data in "total resources which can be mobilized per year" series.

Descriptive statistics of "total resources which can be mobilized per year" series are found in the table below:

Table 4: Statistics of "Total resources which can be mobilized per year" series

	Number	% Missing data	Mean	Standard-error
Total resources which can be mobilized	30	30.2326	3272.3	1125.15
R^2	0.986			

International Journal of Management and Economics Invention ISSN (e): 2395-7220
||Volume||3||Issue||12||Pages-1480-1492||Dec.-2017||
Index Copernicus ICV: 72.76, DOI: 10.18535/ijmei/v3i12.01

As before, we first determine the components t_{hi} and P_{hj}. We note that the maximum of the total variance is captured by the first two principal components. After this step we found the results summarized in the following table:

Table 5: Values of P_h

	P_{h1}	P_{h2}
Total resources which can be mobilized	0.501487	0.740101

The values of t_h are shown in Appendix 2. Subsequently, to estimate missing values we need to solve the following equation:

Total resources which can be mobilized $_i$ = 3272.3 + (1125.15) (0.501487) (t_{i1}) + (1125.15) (0.740101) (t_{i2})

The following figure shows the evolution of available resources after imputation of missing data.

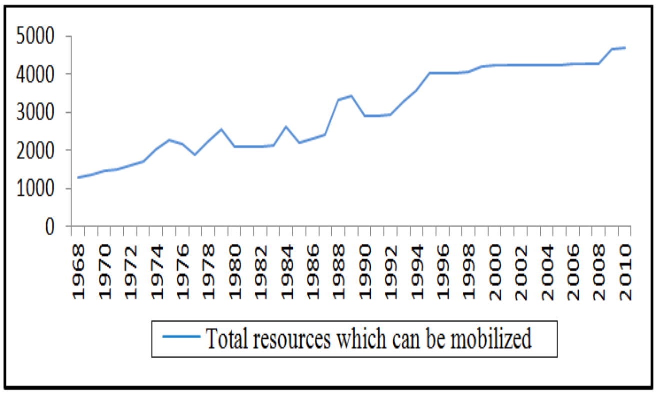


Figure 5: Evolution of Total resources which can be mobilized (million m^3)

During the same period 1968 – 2010, the evolution of the "Total resources which can be mobilized" is more important than this of "Total potential". This result can be explained by the fact that Tunisian State pays more and more attention in investment of water exploitation project and the amelioration of water quality to be available to the satisfaction of human needs. Another explanation can be also advanced. Due to the population growth and the dramatic rise of urbanization rate, the needs of healthy water increase more and more and subsequently exploitation of water grows.

A final step in this sequential approach is to determine the evolution of "per capita availability per m^3" using the following formula:

$$\text{per capita availability}(m^3)_t = \frac{\text{Total resources which can be mobilized}_t}{\text{Total population}_t}$$

It is undeniable that Tunisia is characterized by avarice of water resources. Indeed, during the last four decades the "per capita availability per m^3" is still below 500 m^3 / person / year, threshold indicator of the presence of a critical situation.

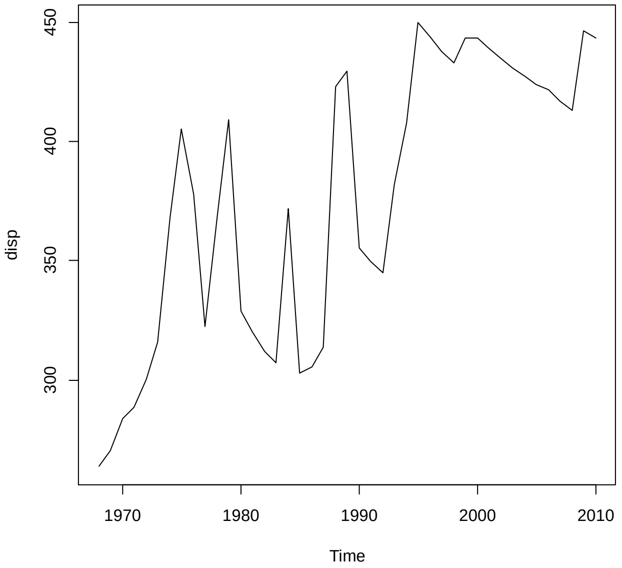


Figure 6: Evolution of the per capita availability m^3

2. Methodology and results

Referring to the methodology of Box and Jenkins, we adopt a systematic study for the "per capita availability m3" series. The main aim of this study is to choose the most appropriate ARMA representation of the studied phenomenon, eventually we end up with the closest prediction to the reality.

This approach is based on four basic steps namely:

- Parameters identification

- Estimation
- Validation
- Prediction

2.1. Stationary study

A time series X_t (t =1,…, T) is generally defined as a sequence of real numbers, indexed by integers, such as time. Before treatment of a time series, it is essential and vital to study their characteristics, ie to analyze its expectation and its variance. If these characteristics vary over time, the time series is called non-stationary; otherwise and in the presence of an invariant stochastic process, X_t is considered stationary.

One of the criticisms of a non-stationary series also called integrated series or unit root series is that we can study its behavior only in the concerned period and we cannot generalized to other periods. Thus, for a forecasting aim, non-stationary series have a very limited value.

In a formal way, the stochastic process X_t is considered stationary if these three conditions are met:

$$E(Xt) = E(Xt+m) = \mu \; ; \forall t \text{ et } \forall m \qquad (6)$$

$$Var\,(Xt) < \infty \; ; \forall t \qquad (7)$$

$$Cov\,(Xt\,,Xt+h) = E[(Xt-\mu)\,(Xt+h-\mu)] = \gamma_h \qquad (8)$$

To test the stationarity of a time series, we often opt for two types of tests: formal and informal tests.

Informal tests

Informal tests boil down simply to the visual analysis of graphics and correlograms.

- <u>Graphics analysis</u>

A first indication about the stationarity of a time series can be provided by graph analysis. In fact, visually, the series of per capita availability m^3 (figure 6) exhibited an overall upward trend and even uses a logarithmic transformation (Figure 7) does not solve this problem and no significant changes in its behavior are noted. This observation leads us to suspect the presumption of non stationarity of our time series. This intuition can be reinforced by the analysis of the correlogram.

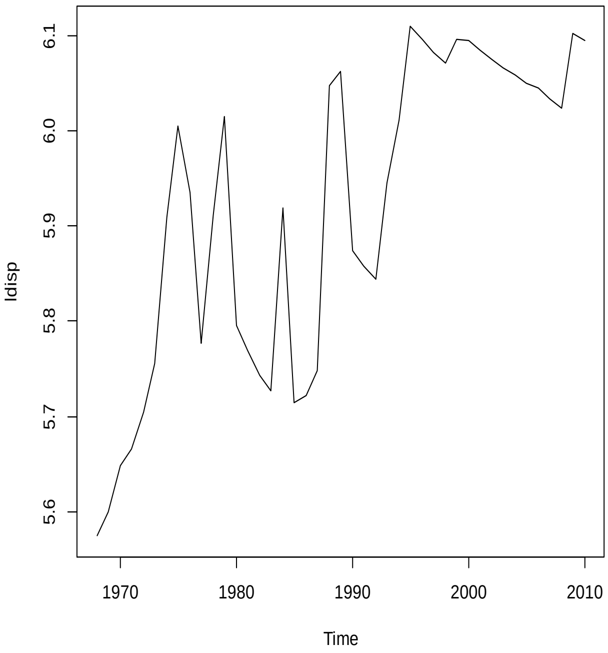


Figure 7: logarithmic transformation of the series of per capita availability m^3

- <u>correlogram analysis</u>

A brief overview of the correlogram of our series (Figure 8) confirms and consolidates the presumption of non-stationary. Indeed, the autocorrelations are significantly different from zero and decreases very slowly. In addition, only the first partial autocorrelation is significantly different from zero. So, it becomes necessary to verify this intuition by applying more efficient and more reliable statistical tests known as formal tests.

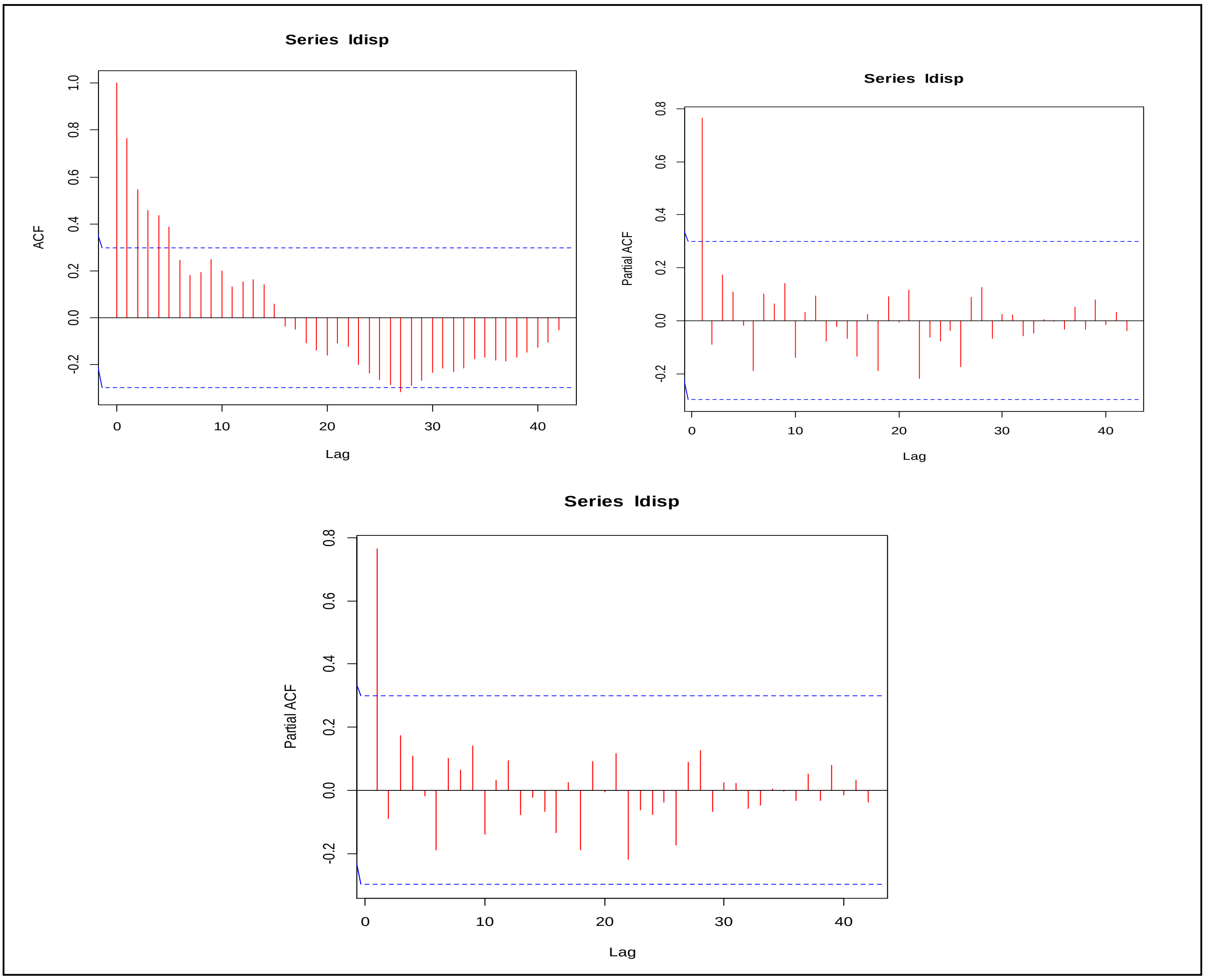


Figure 8: correlogram of the series log per capita availability m^3 (ldisp)

formal tests

A large body of stationary tests is developed in the literature such as: Dickey Fuller test (DF), Augmented Dickey Fuller test (ADF), Philips Perron test, KPSS test ... Generally these tests can be classified into two groups. The first group is based on the null hypothesis of no unit root in the chronic (KPSS test (1992)). The other group is based on the null hypothesis of presence of unit root in the time series (ADF test). In our analysis, we opt for three tests (ADF, PP, KPSS). Table 1 below summarized the different results.

Table 6: Stationarity tets

	ADF	PP	KPSS
Ldisp	-2.5026 (0.3752)	-17.2953 (0.08416)	1.5397 (0.01)
Dldisp	-4.2732* (0.01)	-33.4768* (0.01)	0.0501* (0.1)

(.) P value

* Stationary at 5%

Table 6 shows that our variable is not stationary in levels. Therefore, we proceed to the first differentiation. Thus, it appears that at 5% threshold the variable "Ldisp" is integrated of order 1. The graphic presentation of our series in the first difference confirms this result (figure 9)

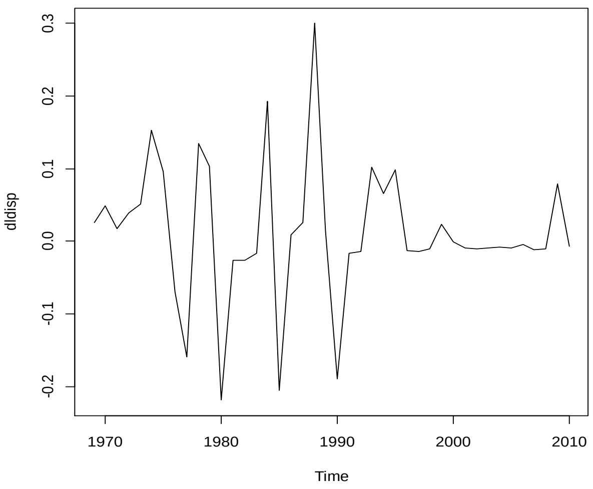


Figure 9: Evolution of the series dldisp

2.2. Parameters Identification

Once obtaining a stationary series we can pass to the identification of the orders of the different sought processes by examining the correlogram. In other words, we must determine the orders p and q of the ARMA process.

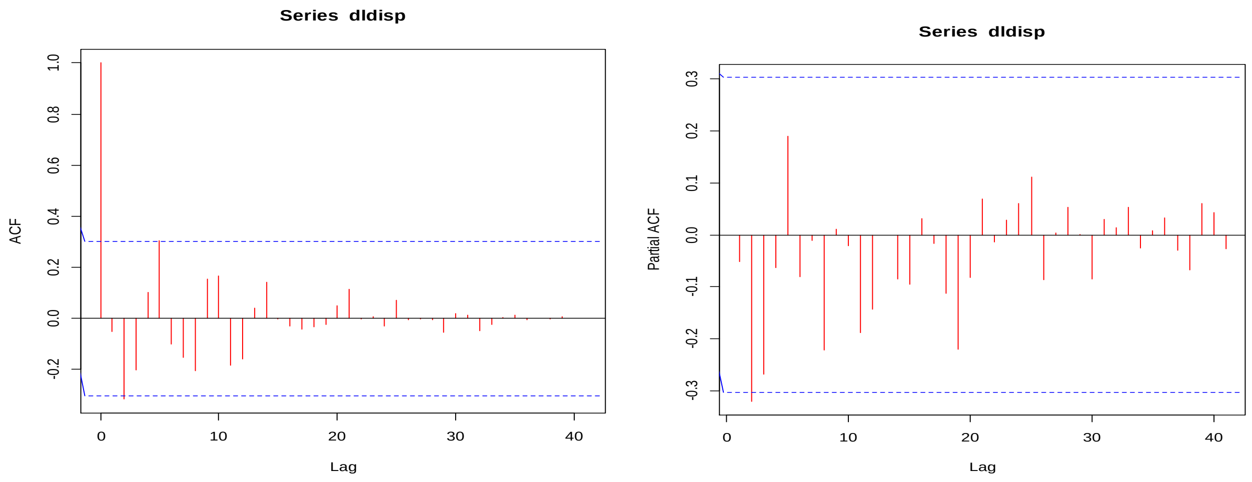


Figure 10 : correlogram of the series dldisp

The analysis of the correlogram shows that only the second partial autocorrelation is significantly different from zero. We can deduce, then, that p = 2. For the simple autocorrelations, for precaution, we take the order q = 3.

After this step, eleven processes namely: AR (1), MA (1), MA (2), MA (3), ARMA (2,3), ARMA (2,2), ARMA (2,1), ARMA (1,3), ARMA (1,2), ARMA (1,1), AR (2), should be estimated in the next step.

International Journal of Management and Economics Invention ISSN (e): 2395-7220
||Volume||3||Issue||12||Pages-1480-1492||Dec.-2017||
Index Copernicus ICV: 72.76, DOI: 10.18535/ijmei/v3i12.01

2.3.Estimation and model validation

We proceed, in a first step, to test parameters significance for all possible combinations of ARMA processes proposed to model the "dldisp" series.

We retain, thereafter, the process that minimizes the information criterion (AIC, SC ...) among all the ARMA models validated in the first step.

Thus, for the series (ldisp), the analysis of all possible combinations shows that only the ARIMA (2,1,2), and ARIMA (1,1,1) have significant coefficients. So, for the rest of our study, we check only those two processes.

- For the ARIMA (2,1,2): The coefficient of the moving average of order 2 is significantly different from zero since the calculated t-statistics is well beyond the tabulated value (≈ 2.0167) at the significance threshold of 5 %. The same conclusion is adopted for the coefficient of the moving average of order 1. Therefore, the ARIMA (2,1,2) remains a candidate for modeling the series (ldisp).
- Regarding the ARIMA (1,1,1): for the part AR, the autoregressive coefficient of order 1 is significantly different from zero since the calculated t-statistics is well beyond its tabulated value at the significance level of 5%. Moreover, the moving average coefficient of order 1 is significantly different from zero at the threshold of 5%.

Following this first phase of the validation step, the two processes ARIMA (2,1,2) and ARIMA (1,1,1) remain competitors for the modeling of the series (ldisp). Thus, to strengthen the good orders p and q, we retain the model that minimizes one of the information criteria (in our case we opt to the Akaike criterion "AIC"). The following table shows the different values of AIC for each validated ARMA process.

Table 7: AIC criterion

Model	$\boldsymbol{ARIMA}(2,1,2)$	$\boldsymbol{ARIMA}(1,1,1)$
AIC	**-75.33**	**-74.22**

To conclude, we can say that the “per capita availability m^{3}” series (ldisp) can be modeled by an ARIMA (2,1,2) process following the next equation:

$$ldisp_t = 0.5209\, ldisp_{t-1} - 0.8199\, ldisp_{t-2} - 0.5911\, \varepsilon_{t-1} + 0.6221\, \varepsilon_{t-2} + \varepsilon_t$$

Or, otherwise:

$$(1 + 0.5209\ L - 0.8199\, L^2) ldisp_t = (1 - 0.5911\, L + 0.6221\, L^2)\varepsilon_t$$

To build a clearer idea about the quality of this adjustment, we use the superposition of the two series (ldisp and ldisp adjusted) on the same graph as traced by the figure below:

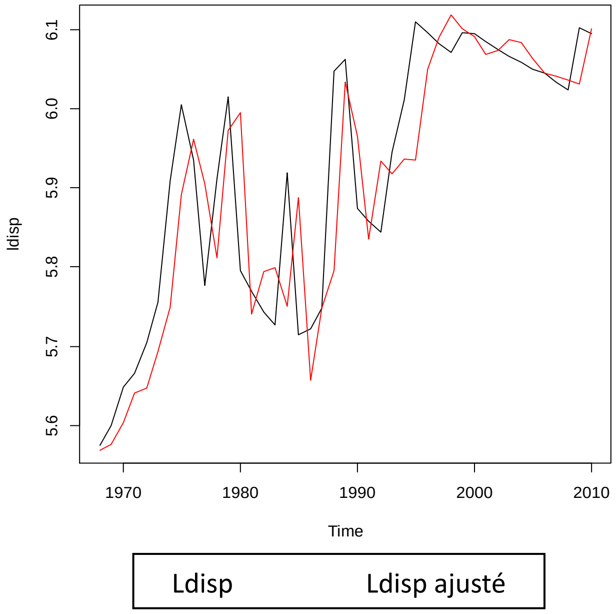


Figure 11: Diagram of the chronic and its adjustment by the ARIMA (2,1,2) process

Visual analysis of this figure shows a certain harmony between the two graphs of both observed and adjusted series. In this case we can speak about a good fit and pass later to residual tests to ensure the adequacy of the accepted model.

Residuals tests

To judge whether the estimate is of good quality, residuals tests are often used.

To diagnose the various residuals anomalies, we conduct tests of absence of autocorrelation and normality tests. Residues of a good process have several qualities such as independence and normality.

- Test of absence of Autocorrelation

By definition, residuals autocorrelation problem arises when the covariance of the error is different from zero. Thus, in the presence of this problem the variance-covariance matrix of residuals will not be diagonal, error terms of different observations are not independent and estimators obtained using the Ordinary Least Square method, despite being unbiased, they don't have the minimum variance.

Generally, a process is considered well estimated if the estimated residuals $e_t = \frac{\hat{\Phi}(L)}{\hat{\Theta}(L)}$ behave as a white noise. In other words, by the absence of autocorrelation test, we seek to check the two hypotheses below:

$$H_0: e_t \textit{ is a white noise}$$

$$H_1: e_t \textit{ is not a white noise}$$

The Ljung Box test with its statistic:

$$Q(k) = n(n+2)\sum_{j=1}^{k}\frac{\hat{\rho}^2(j)}{(n-j)}$$

(where n is the total number of periods) is often applied to test the hypothesis of serial independence in a time series. This statistic is distributed according to a chi-squared distribution with k degree of freedom. The application of this test allows us to accept the null hypothesis (H_0) at the 5% threshold if $Q(k)$ is less than the quantile 0.95 of the corresponding chi-squared law (the critical probability "p-value" must be important and exceeds the fixed threshold). The result of this test is summarized in the following table:

Table 8: Ljung-Box test

	$\boldsymbol{ARIMA(2,1,2)}$ **(ldisp)**
p-value	0.5647

The examination of the above table reveals that residuals are not correlated (p-value is greater than the 5% level). So, the selected process is of good quality. Following the test of absence of autocorrelation, we must test the normality of residuals.

- normality test

Residues of an ideal model must always possess the properties of a normal distribution.

The literature proposes an important package of normality tests. We endeavor in this paper three tests:

The Shapiro-Wilk test: This test proposed in 1965 by Samuel Shapiro and Martin Wilk sits on the following statistic:

$$W = \frac{\left(\sum_{i=1}^{n} a_i x_i\right)^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

We reject the normality hypothesis if the statistic is too small, Otherwise, if the p-value is less than the fixed threshold.

Anderson Darling test: Theodore Anderson and Donald Darling suggested this test in 1952.

The test Lilliefors: proposed by Hubert Lilliefors, this test is an adaptation of the Kolmogorov-Smirnov test.

Table 9: Test of Normality

	Anderson-Darling	**Shapiro-Wilk**	**lillie**
$\boldsymbol{ARMA(2,1,2)}$ **(ldisp)**	0.02911	0.1188	0.05632

The Anderson-Darling test has shown that residues of the "per capita availability (m^3)" follow a normal distribution in the statistical threshold of 10%. While Shapiro-Wilk test and Lillie test highlighted that these residues follow a normal distribution in the statistical threshold of 5%.

Thus, we can deduce that our estimated process is perfect and can be validated. We can, thereafter, pass to the forecasting step.

2.4. Forecast

The forecasting step is very important since it allows us to reveal the changing in "water per capita availability (m^3)" in Tunisia in the next years.

The forecast of the series studied at a horizon of nineteen years (2011 - 2025) is presented by the following figure:

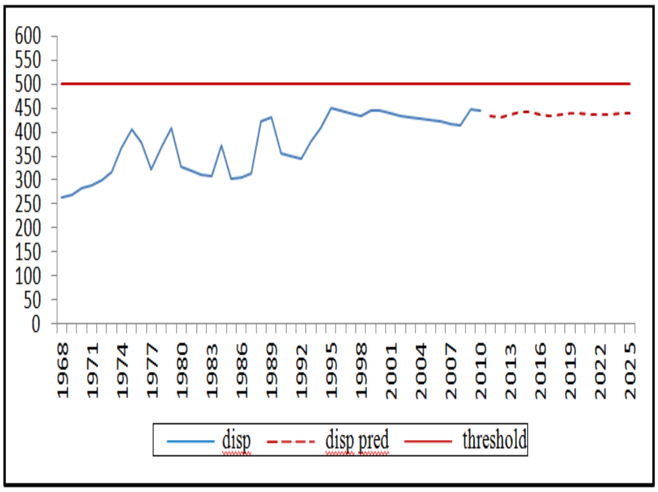


Figure 12: Forecast of the per capita availability (m^3)

Visual analysis of the above figure does not reflect any cheering sign; in contrast, it notes a sort of uneasiness and even an alarming and critical situation. Indeed, it reveals that the availability of water in Tunisia is stabilizing at a dangerous and appalling level in the coming years. Although the threshold indicator of the presence of water drought problem is set at 1000 m^3 per capita per year, Tunisian citizen fails to exceed the threshold of 500 m^3 of water per year, threshold of absolute scarcity according to Falkenmark. Thus, we affirm that Tunisia is still among the countries suffering from a shortage in water resources and this worrying situation will not be improved in the next years, according to our forecasts, rather it will spread from year to another.

By this predictive analysis, we reach the water management problem in Tunisia and a crucial enigma of making decision. In fact, the establishment of an effective and sustainable patrimonial management in order to plan projects and investments in water sector conserve and streamline water consumption, improve the efficiency of infrastructure mainly through reducing the volume of water waste and rehabilitation of water networks ... becomes a decisive issue

Conclusion and Recommendations

Without claiming completeness, the problem of water situation, although it is very interesting and treated by many researchers, we treat it a little differently using NIPALS algorithm to impute missing data in our series and prediction using Box and Jenkins methodology reveals a deterioration in water per capita availability (m^3) in Tunisia in the coming years.

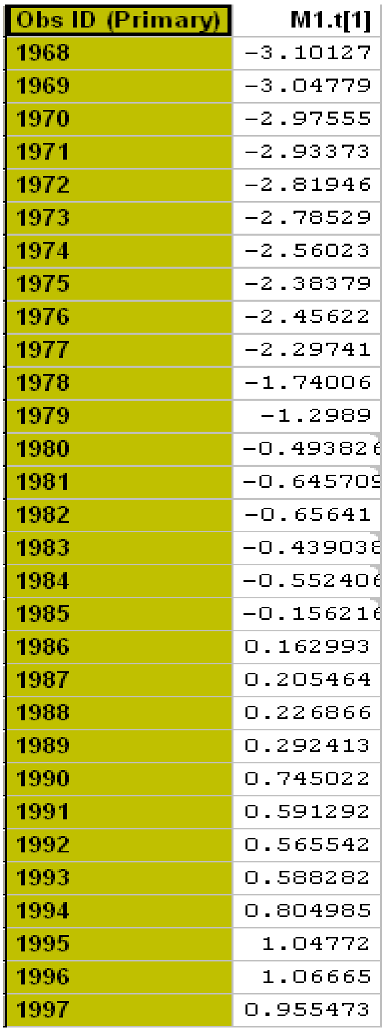
At this stage, and in a context of structural shortage ((less than 500 m^3 / year per capita) and facing to growing uncertainties related to climate change, population growth, the evolution of tourism sector, industry and irrigated land, an appropriate and effective management is essential and urgent decisions should be taken to ensure the sustainability of human, economic and ecological development. Hence, Tunisian State must develop action plans for a more efficient use of resources, a suitable management of the different uses of water and encourage more and more subscribers to rationalize their use.

Brief, Tunisia needs to develop a more sustainable development scenario that can only be achieved gradually through necessary reforms on two areas: water conservation and integrated management of networks and water resources.

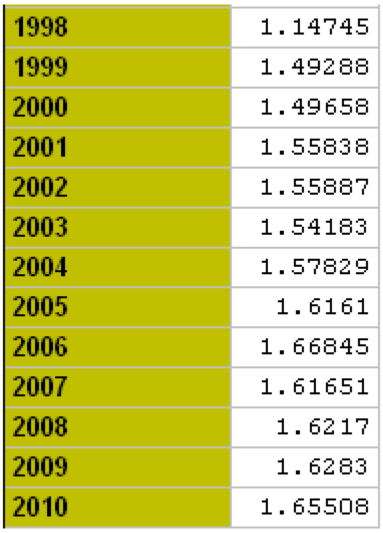
Bibliography

1. AL-DHOWALIA k.h.,(1996). Modeling Municipal Water Demand Using Box-Jenkins Technique. Jkau:eng. sct., VOLUME 8, 61-71
2. Preda C., Duhamel A., Picavet M., KechadiT., (2005). Gestion des données manquantes dans les grandes bases de données en Santé. Journées Francophones d'Informatique Médicale, Lille
3. Preda C., Saporta G., Hadj Mbarek M.,(2010). The NIPALS Algorithm for Missing Functional Data. Revue Roumaine de Mathématiques Pures et Appliquées, vol. 55(4), pp. 315-326
4. Josse J., Husson F., Pagés J., (2009). Gestion des données manquantes en Analyse en Composantes Principales. Journal de la Société Française de Statistique. Volume 150
5. Lawgali F.F., (2008). Forecasting water demand for agricultural, industrial and domestic use in Libya. International Review of Business Research Papers, VOLUME 4? 231-248
6. Little R.J.A., Rubin D.B. (1987). Statistical Analysis with Missing Data. New York: John Wiley.
7. Little RJA, Rubin DB. Statistical analysis with missing data. Wiley series in Probability and Statistics. 2nd ed. New York: Wiley, 2002.
8. Maidment d.r., et al (1985). Transfer Function Models of Daily Urban Water Use. Water Resources Research Volume 21, 425–432,
9. Schafer J.L., Imputation Procedures For Missing Data. USA Université de Pennsylvania 1999: http://www.stat.psu.edu/~jls/session2.pdf
10. Tenenhaus M., La régression PLS Théorie et pratique. Editions Technip, 1998.
11. Wasito, I., Mirkin, B., (2006), Nearest neighbours in least-squares data imputation algorithms with different missing patterns, Computational Statistics and Data Analysis, 50, 926-949.

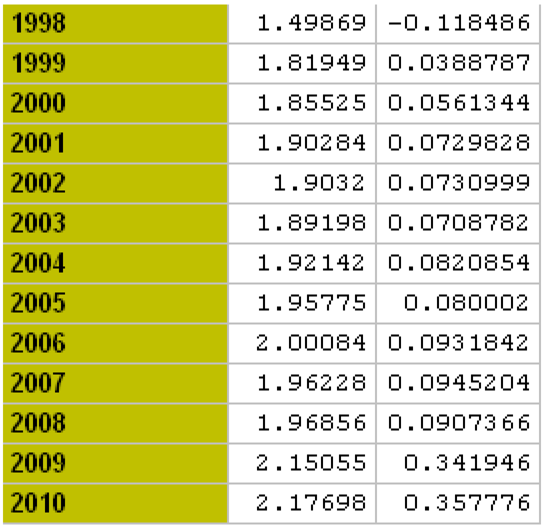
Annex 1



Obs ID (Primary)	M1.t[1]
1968	-3.10127
1969	-3.04779
1970	-2.97555
1971	-2.93373
1972	-2.81946
1973	-2.78529
1974	-2.56023
1975	-2.38379
1976	-2.45622
1977	-2.29741
1978	-1.74006
1979	-1.2989
1980	-0.493826
1981	-0.645709
1982	-0.65641
1983	-0.439038
1984	-0.552406
1985	-0.156216
1986	0.162993
1987	0.205464
1988	0.226866
1989	0.292413
1990	0.745022
1991	0.591292
1992	0.565542
1993	0.588282
1994	0.804985
1995	1.04772
1996	1.06665
1997	0.955473

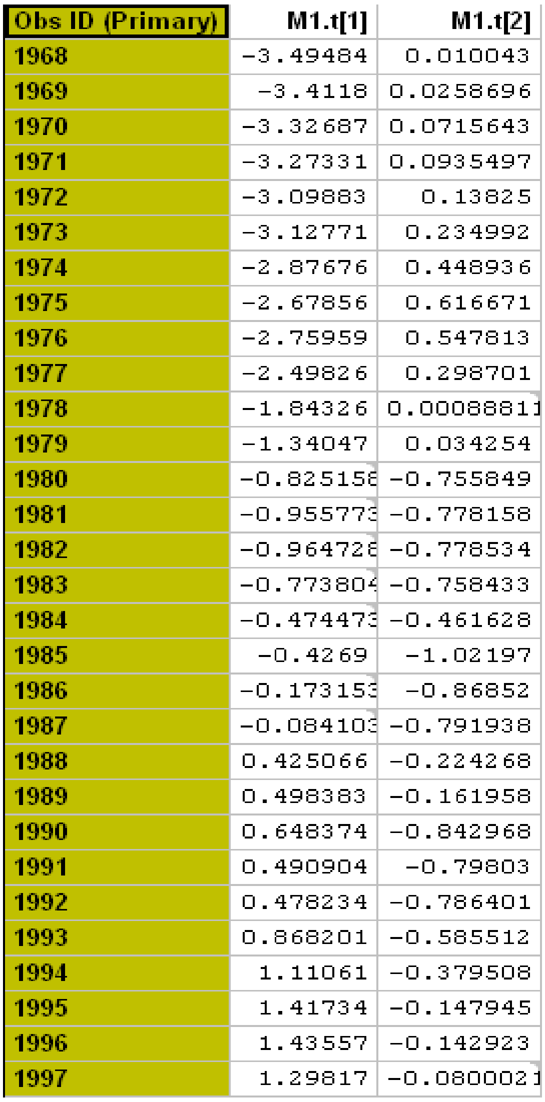


1998	1.14745
1999	1.49288
2000	1.49658
2001	1.55838
2002	1.55887
2003	1.54183
2004	1.57829
2005	1.6161
2006	1.66845
2007	1.61651
2008	1.6217
2009	1.6283
2010	1.65508



1998	1.49869	-0.118486
1999	1.81949	0.0388787
2000	1.85525	0.0561344
2001	1.90284	0.0729828
2002	1.9032	0.0730999
2003	1.89198	0.0708782
2004	1.92142	0.0820854
2005	1.95775	0.080002
2006	2.00084	0.0931842
2007	1.96228	0.0945204
2008	1.96856	0.0907366
2009	2.15055	0.341946
2010	2.17698	0.357776

Annex 2: Available resources



Obs ID (Primary)	M1.t[1]	M1.t[2]
1968	-3.49484	0.010043
1969	-3.4118	0.0258696
1970	-3.32687	0.0715643
1971	-3.27331	0.0935497
1972	-3.09883	0.13825
1973	-3.12771	0.234992
1974	-2.87676	0.448936
1975	-2.67856	0.616671
1976	-2.75959	0.547813
1977	-2.49826	0.298701
1978	-1.84326	0.00088811
1979	-1.34047	0.034254
1980	-0.825158	-0.755849
1981	-0.955773	-0.778158
1982	-0.964728	-0.778534
1983	-0.773804	-0.758433
1984	-0.474473	-0.461628
1985	-0.4269	-1.02197
1986	-0.173153	-0.86852
1987	-0.084103	-0.791938
1988	0.425066	-0.224268
1989	0.498383	-0.161958
1990	0.648374	-0.842968
1991	0.490904	-0.79803
1992	0.478234	-0.786401
1993	0.868201	-0.585512
1994	1.11061	-0.379508
1995	1.41734	-0.147945
1996	1.43557	-0.142923
1997	1.29817	-0.0800021